storage & search

up in the clouds

History of Information
April 15, 2010
For solo artist to earn US monthly min. wage $1,160

- Self-pressed CD: $1.48
- CD album: $1.55
- Retail album CD (high-end retail deal): $1.16
- Album download via iTunes: $1.23
- MP3 download via iTunes: $1.36
- MP3 download via Amazon: $1.63
- Retail album CD (low-end retail deal): $3.87
- Track download via iTunes: $12.39
- Spotify Stream: 849,817 plays per month, $0.004 per play

on behalf of the artist?

informationisbeautiful.net
Court to hear suit over "Tea Party" name

Jane Sutton
MIAMI
Wed Apr 14, 2010 2:37pm EDT

(Reuters) - A U.S. judge has agreed to referee a dispute among Florida political activists over who can use the phrase "Tea Party" in their name.

ODDLY ENOUGH

A trial has been scheduled to begin on December 6 in U.S. District Court in West Palm Beach, Florida, in a lawsuit that questions whether anyone has a trademark or intellectual property right to the "Tea Party" name.
where we left off

1988, WAIS

1990, Archie

1992, Veronica (Gopher)

1994, Lycos

1995, Alta Vista, Yahoo

1996, Inktomi

1997, Ask Jeeves
looking back to look forward
"men more frequently need to be reminded than informed"
S. Johnson, *Rambler 2*

search and storage across time

 assignments
looking back or forward?

perennial issues?
looking back or forward?

perennial issues?

too little or too much information?
looking back or forward?

perennial issues?
too little or too much information?
information wants to be free?
looking back or forward?

perennial issues?

too little or too much information?

information wants to be free?

information wants to be constrained?
old and new stories

"The second paid link pointed to the epic Gilgamesh ...[which], the site instructed me, recounts mankind's 'longing stretch toward the infinite.' In my search for immortality, I had found the oldest known author in the history of Western civilization.

"to become immortal through words pressed in clay"

John Battelle, *The Search*, 2005
stored writings

**Ebla** (Syria) c 2250 bce

**Babylon** (Iraq) 15,000 tablets

**Nineveh** (Iraq) c. 650 bce

*what was stored?*

- Ebla: accounts, lists, etc.
- Nineveh: *Gilgamesh*
He came a far road, was weary, found peace, and set all his labours on tablets of stone ...

See the tablet box of cedar, Release its clasp of bronze
Lift the lid of its secret
Pick up the tablet of lapis lazuli and read out
the travails of Gilgamesh, all that he went through

--Gilgamesh, 3000 BCE
[trans Andrew George]
new technologies

Lyceum [Greece]
(Aristotle)

Pergamum
[Turkey]
(c. 197 BCE)
200,000 scrolls
parchment

Alexandria
430,000 volumes
papyrus
to print & paper

overload?

972-83: Tripitaka, 5,048 volumes
memory to written record

questions of durability

1145, Roger of Sicily ordered all charters on paper to be copied to parchment then destroyed.

1248, paper accepted for us by the notaries of Languedoc (France).

"The written word on parchment will last a thousand years. The printed word is on paper. How long will it last? The most you can expect of a book of paper to survive is two hundred years. Only time will tell."

12
abundance & order

Alexandrian strategies

Zenodotus
alphabetical ordering

Callimachus
subject categories

Aristophanes & Aristarchus
scholarly versions
editorial commentary

in print

Conrad Gessner
Bibliotheca Universalis [Latin, Greek, Hebrew]
free &

constrained
free & constrained
serendipity & organization
Smirke, British Museum Reading Room, 1851

Pelz/Casey Reading Room, LOC, ca 1898

Labrouste, Bibliothèque Nationale 1868

Asplund, Stockholm City Library, 1928
more organization

Edward Gibbon
1737–1794
Dewey Decimal

Library of Congress

Thomas Jefferson -- Denis Diderot -- Francis Bacon
ment, irrespective of the time of day or night, or state of the weather. This single point establishes its superiority to all other modes of telegraphic communication now known.

Third. The whole apparatus will occupy but little space, (scarcely si-cubic feet, probably not more than four;) and it may therefore be placed, without inconvenience, in any house.

Fourth. The record of intelligence is made in a permanent manner, and in such a form that it can be at once bound up in volumes conven-
nient for reference, if desired.

Fifth. Communications are secret to all but the persons for whom they are intended.

These are the chief advantages of the electro-magnetic telegraph over other kinds of telegraphs, and which must give it the preference, provided the expense and other circumstances are reasonably favorable.

The newness of the whole plan makes it not so easy to estimate the expense, but an approach to a correct estimate can be made.
Photography applied to Catalogues of Books.—May not photography be usefully applied to the making of catalogues of large libraries? It would seem no difficult matter to obtain any number of photographs, of any required size, of the title-page of any book. Suppose the plan adopted, that five photographs of each were taken; they may be arranged in five catalogues, as follows: —Era, subject, country, author, title. These being arranged alphabetically, would form five catalogues of a library probably sufficient to meet the wants of all. Any number of additional divisions may be added. By adopting a fixed breadth—say three inches—for the photographs, to be pasted in double columns in folio, interchanges may take place of unerring slips, and thus librarians aid each other. I throw out this crude idea, in the hope that photographers and librarians may combine to carry it out. Albert Bloe, LL.D.

Photographic Copies of Ancient Manuscripts.—Might not photography be well employed in making fac-similes of valuable, rare, and especially of unique ancient manuscripts? If copies of such manuscripts could be multiplied at a moderate price, there are many proprietors of libraries would be glad to enrich them by what, for all purposes of reference, would answer equally well with the originals.

A.

[This subject, which has already been touched upon in our columns, has not yet received the attention it deserves. We have now before us a photographic copy of a folio page of a MS. of the fourteenth or fifteenth century, on which are inscribed a number of charters; and, although the copy is reduced so as to be but about 2 inches high and 1 ½ broad, it is perfectly...]

Notes & Queries, 1854, 1855
19th century public sphere

the penny press

&

the "paper of record"
"all the news that's fit to print"
An act concerning...

1. public archive
2. state printer
3. pilots for SF
4. comptroller
5. treasurer
6. sec. of state
8. translator
11. AG
14. Supreme Court
30. incorporation of cities
36. commissioner of deeds
41. notaries
49. lawful fences

48. incorporation of towns
53. weights & measures
55. limited partners
59. recorder's office
64. officers of health
67. surveyors
69. librarian
72. register of wills
89. marks & brands
90. reporter
93. conveyances
95. common law
117. incorp. of colleges
123. assayer

Statutes of California, 1849-50
There is a growing mountain of research. But there is increased evidence that we are being bogged down today as specialization extends. ... A record if it is to be useful to science, must be continuously extended, it must be stored, and above all it must be consulted"

--As we may think, 1945
A memex is a device in which an individual stores all his books, records and communications... mechanized... associative indexing... his trails do not fade.
Hi, Robot
The history of the vocoder, cryptography's top-secret funk machine.

By Sam Anderson  Published Apr 4, 2010

1 Comment | Add Yours

& the vocoder
"Our ineptitude in getting at the record is largely caused by the artificiality of systems of indexing.

"The applications of science have built man a well-supplied house, and are teaching him to live healthily therein. ... They may yet allow him truly to encompass the great record."
solving or creating problems?

1993, NCSA Mosaic
Mark Andressen

CERN releases W3 technology

1994, 200+ HTTP servers; traffic up x 1,000

1994, Netscape

1995, Internet Explorer

2009, Google Chrome
old problem, new tools?

1988, WAIS

1990, Archie

1992, Veronica (Gopher)

1994, Lycos

1995, Alta Vista, Yahoo

1996, Inktomi

1997, Ask Jeeves

"the web is a global library produced by millions of people"

-- Yochai Benkler

"the Internet ... is a library"

-- Frances Cairncross

"the web is a global library produced by millions of people"

-- Yochai Benkler

"the Internet ... is a library"

-- Frances Cairncross
real change?
real change?

yet another hierarchical officious oracle
changing ideas of organization?

"to organize the world's information"
"Up until now most search engine development has gone on at companies with little publication of technical details. This causes search engine technology to remain largely a black art and to be advertising oriented (see Appendix A). .. we have a strong goal to push more development and understanding into the academic realm."
For example, in our prototype search engine one of the top results for cellular phone is "The Effect of Cellular Phone Use Upon Driver Attention", a study which explains in great detail the distractions and risk associated with conversing on a cell phone while driving. This search result came up first because of its high importance as judged by the PageRank algorithm, an approximation of citation importance on the web [Page, 98]. It is clear that a search engine which was taking money for showing cellular phone ads would have difficulty justifying the page that our system returned to its paying advertisers. For this type of reason and historical experience with other media [Bagdikian 83], we expect that advertising funded search engines will be inherently biased towards the advertisers and away from the needs of the consumers.
For example, in our prototype search engine one of the top results for cellular phone is "The Effect of Cellular Phone Use Upon Driver Attention," a study which explains in great detail the distractions and risk associated with conversing on a cell phone while driving. This search result came up first because of its high importance as judged by the PageRank algorithm, an approximation of citation importance on the web [Page, 98].

It is clear that a search engine which was taking money for showing cellular phone ads would have difficulty justifying the page that our system returned to its paying advertisers. For this type of reason and historical experience with other media [Bagdikian 83], we expect that advertising funded search engines will be inherently biased towards the advertisers and away from the needs of the consumers.
“We lose money on our circulation by itself considered but with 20,000 subscribers we can command such Advertising and such prices for it as will render our enterprise a remunerating one.”

Horace Greeley
1811 – 1872
new solutions, new problems?

goodhart again?

Kicking Cash to Nefarious Clickers

By ALEX MANCE
Published: April 11, 2010

The click fraud rate in the first quarter of 2010 was one of the highest on record, according to Click Forensics, a firm that analyzes traffic on behalf of advertisers and ad networks. The rate was even higher than the previous quarter’s, which was unusual since the holiday shopping period, with its high ad rates, is usually the peak season for click fraud.

17.4%
Click fraud rate in Q1 2010

15.3%
Click fraud rate in Q4 2009

Web sites practice click fraud by showing ads, soliciting fraudulent clicks on those ads, receiving payments from advertisers for the clicks, and then kicking some money back to the clickers. Paul Pellman, the chief executive of Click Forensics, attributed the rising fraud rate to the emergence of sophisticated, widely distributed schemes wherein millions of hijacked personal computers, harnessed together into “botnets,” are used to fraudulently click on ads across thousands of Web sites.

Each computer is used only a few times and each ad is clicked infrequently, so the scheme is hard to detect. “The clicks are coming from many millions of IP addresses,” Mr. Pellman said. “It’s low frequency from that standpoint, and also low frequency on each site, so it can look like valid traffic.”
favoritism?

whose interest

---

Using site speed in web search ranking

(Friday, April 09, 2010 at 11:00 AM.
Webmaster Level: All)

You may have heard that here at Google we’re obsessed with speed, in our products and on the web. As part of that effort, today we’re including a new signal in our search ranking algorithms: site speed. Site speed reflects how quickly a website responds to web requests.

Speeding up websites is important — not just to site owners, but to all internet users. Faster sites create happy users and we’ve seen in our internal studies that when a site responds slowly, visitors spend less time there. But faster sites don’t just improve user experience; recent data shows that improving site speed also reduces operating costs. Like us, our users place a lot of value in speed — that’s why we’ve decided to take site speed into account in our search rankings. We use a variety of sources to determine the speed of a site relative to other sites.

If you are a site owner, webmaster or a web author, here are some free tools that you can use to evaluate the speed of your site:

* Page Speed, an open source Firefox/Firebug add-on that evaluates the performance of web pages and gives suggestions for improvement.
* YSlow, a tool from Yahoo! that suggests ways to improve website speed.
* WebPageTest shows a waterfall view of your pages’ load performance plus an optimization checklist.
* In Webmaster Tools, Labs > Site Performance shows the speed of your website as experienced by users around the world as in the chart below. We’ve also blogged about site performance.
SAN FRANCISCO — Bill Gross, the serial entrepreneur who pioneered search advertising, is unveiling a venture on Monday that aims to make money by allowing people using Twitter to bid on key words to give their posts top ranking.

Called TweetUp, the service will also organize the posts according to their popularity as measured by how often readers repost them and click on links they contain.

Mr. Gross said he had signed deals with other outside Twitter services like Seesmic, TwitterFeed and Twidroid to display TweetUp’s rankings. A TweetUp search bar will appear on Web sites like Answers.com and BusinessInsider.com. TweetUp will split revenue evenly with each partner, he said.

The goal is to cut through the clutter of thousands of irrelevant posts on topics of interest and keep the useful ones from disappearing into a torrent of messages.
and who is searching for what?

and should we worry?

"You have zero privacy.... Get over it!"
Scott McNealy, Sun Microsystems

"If you have something you don't want anyone to know, maybe you shouldn't be doing it in the first place."
Eric Schmidt, Google
paranoia or trustanoia

"shock"

"scary"

"benefits outweigh"

"I will pay attention"

"less worrying than facebook"

"google is throwing away its money"

"how difficult it was to escape the grasp of Google"
some thoughts

Robert Ang
Jody Leung
Kasey Chiu, Anne Sokolich
Holly Fine
Apin Sicoravit
Marco Santo Danesi
Daniel Wang
Justin Riddle
### Ambiguous Terms

<table>
<thead>
<tr>
<th>Search Term:</th>
<th>Configurations:</th>
<th>Google Search Results:</th>
<th>Bing Search Results:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perseverance logged in to personal Gmail account</td>
<td>Definitions, quotes, synonyms, NHL Hockey</td>
<td>Definitions, posters</td>
<td></td>
</tr>
<tr>
<td>Turned off cookies</td>
<td>Definitions, quotes, theatre company</td>
<td>Definitions, quotes</td>
<td></td>
</tr>
<tr>
<td>Logged in to father’s account</td>
<td>Quotes, definitions</td>
<td>Definitions, quotes</td>
<td></td>
</tr>
<tr>
<td>Balance</td>
<td>Logged in to personal Gmail account</td>
<td>Definitions, finance program, yoga</td>
<td>Selling onlineshoes, definitions, games</td>
</tr>
<tr>
<td>Turned off cookies</td>
<td>Disorders, software</td>
<td>Balance bar, definitions</td>
<td></td>
</tr>
<tr>
<td>Logged in to father’s account</td>
<td>Images of balance scale, disorders</td>
<td>New Balance shoe shopping, environment</td>
<td></td>
</tr>
<tr>
<td>Social Media</td>
<td>Logged in to personal Gmail account</td>
<td>How to use social media, business, blog posts, books (The Social Media Bible)</td>
<td>Youtube, marketing, advertising</td>
</tr>
<tr>
<td>Turned off cookies</td>
<td>Youtube, social media graphic diagrams</td>
<td>Business strategies, marketing</td>
<td></td>
</tr>
<tr>
<td>Logged in to father’s account</td>
<td>Youtube, social media graphic diagrams</td>
<td>Business strategies, marketing</td>
<td></td>
</tr>
</tbody>
</table>

### Specific Terms

<table>
<thead>
<tr>
<th>Search Term:</th>
<th>Configurations:</th>
<th>Google Search Results:</th>
<th>Bing Search Results:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soccer</td>
<td>Logged in to personal Gmail account</td>
<td>ESPN, FIFA, News results: college soccer scores, Youtube</td>
<td>Nike shopping, ESPN, MLS, Listings for soccer near Las Vegas</td>
</tr>
<tr>
<td>Turned off cookies</td>
<td>Selling jerseys, ESPN, image of soccer balls</td>
<td>ESPN, MLS, selling jerseys</td>
<td></td>
</tr>
<tr>
<td>Logged in to father’s account</td>
<td>FOX, ESPN</td>
<td>Nike soccer</td>
<td></td>
</tr>
<tr>
<td>Marathons</td>
<td>Logged in to personal Gmail account</td>
<td>Running directory, race directory, Boston Marathon</td>
<td>Oil company, Electric motors, marathon running</td>
</tr>
<tr>
<td>Turned off cookies</td>
<td>Running directory, race directory, Boston Marathon</td>
<td>Travel Expedia, oil company, running</td>
<td></td>
</tr>
<tr>
<td>Logged in to father’s account</td>
<td>Marine corps marathon, run a marathon</td>
<td>Travel, Paypal</td>
<td></td>
</tr>
<tr>
<td>Giants</td>
<td>Logged in to personal Gmail account</td>
<td>MLB scores, Twitter posts, SF Giants, New York Giants</td>
<td>Giant food, bicycles, SF Giants, Tickets</td>
</tr>
<tr>
<td>Turned off cookies</td>
<td>MLB scores, SF Giants, New York Giants</td>
<td>Giant food, bicycles, SF Giants, Tickets</td>
<td></td>
</tr>
<tr>
<td>Logged in to father’s account</td>
<td>MLB scores, SF Giants, New York Giants</td>
<td>Giant food, bicycles</td>
<td></td>
</tr>
</tbody>
</table>

### Student Related Terms

<table>
<thead>
<tr>
<th>Search Term:</th>
<th>Configurations:</th>
<th>Google Search Results:</th>
<th>Bing Search Results:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Law School</td>
<td>Logged in to personal Gmail account</td>
<td>Online law school, LSAT, Harvard Law School</td>
<td>Wikipedia, online law school</td>
</tr>
<tr>
<td>Turned off cookies</td>
<td>Twitter, Georgetown Law School, Yale Law School</td>
<td>Online law schools</td>
<td></td>
</tr>
<tr>
<td>Logged in to father’s account</td>
<td>Stanford Law School, Westlaw</td>
<td>Online law schools</td>
<td></td>
</tr>
</tbody>
</table>
Kasey Chiu & Anne Sokolich
Summary:

Google definitely keeps tabs on a user, and is a master of knowing location. They are very smart in utilizing their Gmail system because if you are honest when filling out the information about Gmail, then they pretty much know your demographic to a tee. Other search engines that aren’t privy to all of that knowledge aren’t able to tailor suit your search results to fit what they think is what you want to see (regardless if it’s what you want to see or not). And using previous search history and all of the aforementioned details, ads can pop up to hopefully “cater” to your needs/wants and therefore you’ll have a desire to CLICK. It’s a little creepy how much Google really knows about you (gmail, gchat, Google docs and of course your search information) but perhaps there is a positive side in that you can get search and/or ad results that are more relevant to you.
## Apin Sicoravit

<table>
<thead>
<tr>
<th>Term</th>
<th>Media Studies</th>
<th>FRIEND 1</th>
<th>FRIEND 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Media Studies at UC Berkeley</td>
<td>In Michigan, on university IP address</td>
<td>Same as FRIEND 2</td>
</tr>
<tr>
<td></td>
<td>Media Theory and Media Studies</td>
<td>New School</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MIT Comparative Media</td>
<td></td>
<td></td>
</tr>
<tr>
<td>News</td>
<td>CNN</td>
<td></td>
<td>CNN</td>
</tr>
<tr>
<td></td>
<td>FOX</td>
<td></td>
<td>FOX</td>
</tr>
<tr>
<td></td>
<td>GoogleNews</td>
<td></td>
<td>GoogleNews</td>
</tr>
<tr>
<td>Forever</td>
<td>Video results for forever</td>
<td></td>
<td>Video results for forever</td>
</tr>
<tr>
<td></td>
<td>Forever 21</td>
<td></td>
<td>Forever 21</td>
</tr>
<tr>
<td></td>
<td>See results for: forever lyrics</td>
<td></td>
<td>Google Books: Forever 21</td>
</tr>
<tr>
<td></td>
<td>Also see forever 21</td>
<td></td>
<td>Local Business results for: forever near Albany, CA</td>
</tr>
<tr>
<td>Steak</td>
<td>Wikipedia</td>
<td></td>
<td>Images results for steak</td>
</tr>
<tr>
<td></td>
<td>Omaha Steak.com</td>
<td></td>
<td>Wikipedia</td>
</tr>
<tr>
<td></td>
<td>Steak: How to turn cheap 'choice' steak into Gucci Prime Steak</td>
<td></td>
<td>Omaha steaks</td>
</tr>
<tr>
<td></td>
<td>See results for: Steak Recipes</td>
<td></td>
<td>Steakdigital.com</td>
</tr>
<tr>
<td></td>
<td>Local business results for steak near Berkeley, CA</td>
<td></td>
<td>Local Business results for Steak in Albany</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Video results for steaks</td>
</tr>
</tbody>
</table>
Marco Santo Danesi
Daniel Wang
Justin Riddle
coming up

April 20 readings to follow

---

Week 14

20 Apr: Books, newspapers & the future of publications
22 Apr: Social implications of the internet I

Required reading:


In the readings, Marshall gives several reasons why industries tended to 'localize' while Cairncross seems to suggest that, with the 'death of distance,' 'fate of location,' and 'increased mobility,' localization will be no more. Are there any arguments in Marshall that Cairncross has overlooked and that might yet prolong localization? If you think there are, discuss why you think Marshall is right. If you think there are not, suggest how Cairncross might account, for example, for Marshall's arguments about invention and skill (Marshall, §3 and beyond).