

World-Wide Web

Παγκόσμιο Πληροφοριακό Περιβάλλον ή Παγκόσμιος Ιστός

Ορισμός του WWW



The **World Wide Web (WWW)** is an **information space** where documents and other **web resources** are identified by **URLs**, interlinked by **hypertext** links, and can be accessed via the **Internet**.

M. Λικαΐδης, ΕΠΙ425

2

Παγκόσμιος Ιστός: Βασικές Όψεις

- Διαδικτυακό Σύστημα Ανάκτησης Υπερμεσικών Πληροφοριών
 - A wide-area **hypertext, multimedia information retrieval system** that provides access to a large universe of documents
- Ενιαίος τρόπος δημοσίευσης πληροφοριών (περιεχομένου) στο Διαδίκτυο
 - A **uniform way of accessing and viewing** some **information** on the Internet
 - subsumes the capabilities of ftp, gopher, wais and news
- Ένας πληροφοριακός χώρος μέσα στον οποίο κάθε πληροφοριακό στοιχείο διαθέτει **αναγνωριστικό / όνομα / ταυτότητα**, μέσω της οποίας μπορούμε να αποκτήσουμε πρόσβαση επάνω του
 - The WWW creates a world in which **information has a reference** by which it can be **accessed**

M. Λικαΐδης, ΕΠΙ425

3

Βασικά Συστατικά Στοιχεία

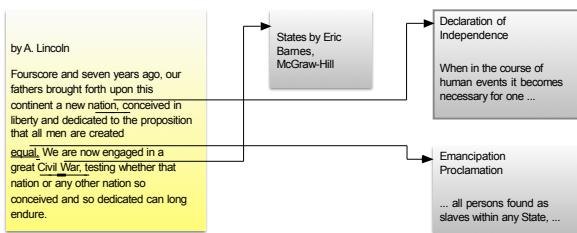
- **Αρχιτεκτονική Πελάτη/Εξυπηρετητή (Client/server):** οι «πελάτες» αλληλεπιδρούν με τους «εξυπηρετητές» (web servers)
- **Ενιαίο Σύστημα Διευθυνσιοδότησης Πόρων** (Addressing system
 - Uniform Resource Locators
 - <http://domain/directory/file.html>
- **Πρωτόκολλο αλληλεπίδρασης**, διαστρωμάτωσης εφαρμογών (application-level network protocol)
 - HTTP, Hypertext Transfer Protocol, is the language understood by browsers and web servers designed to move quickly from document to document
- **Γλώσσες Σήμανσης** (Markup Language)
 - every web server understands and every browser displays
 - includes support for HyperText and multimedia

M. Λικαΐδης, ΕΠΙ425

4

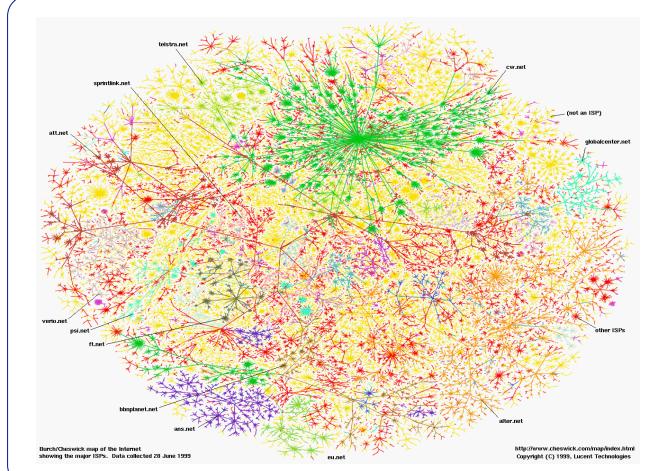
Hypertext - Υπερκείμενο

- Κείμενο με την προσθήκη συνδέσεων σε άλλα σχετικά έγγραφα
- Καθώς διαβάζουμε ένα έγγραφο και ακολουθούμε συνδέσμους, διασχίζουμε έναν «ιστό» από διασυνδέσεις



M. Λικαΐδης, ΕΠΙ425

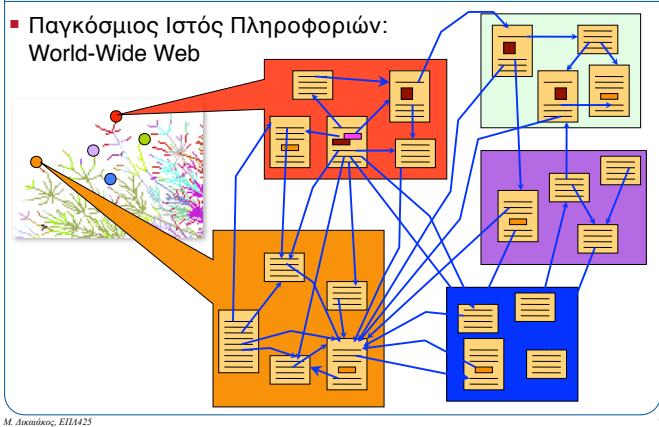
5



Barcheckwick map of the Internet
showing the major ISPs. Data collected 29 June 1999.
Copyright (C) 1999, Lucent Technologies

Λίγη Ορολογία: Παγκόσμιος Ιστός

- Παγκόσμιος Ιστός Πληροφοριών: World-Wide Web



7

Παγκόσμιος Ιστός: Αντίκτυπος

- Η ευρύτερα χρησιμοποιούμενη και ριζοσπαστικότερη εφαρμογή στην ιστορία του Υπολογισμού και της ανθρώπινης επικοινωνίας:
 - Δημιουργησε νέους βιομηχανικούς τομείς αλλά και άλλαξε την επικοινωνία και την παροχή υπηρεσιών σε ολόκληρους κλάδους της βιομηχανίας.
 - Στις επιστήμες, τροποποίησε τον τρόπο με τον οποίο διδάσκουμε, επικοινωνούμε, δημοσιεύουμε, κάνουμε έρευνα.
 - Στη διακυβέρνηση, επηρέασε τον τρόπο με τον οποίο οι κυβερνήσεις επικοινωνούν με τους πολίτες τους αλλά και το πώς οι πολίτες επικοινωνούν και επιλέγουν κυβερνήσεις.

M. Λικείδης, ΕΠΙ425

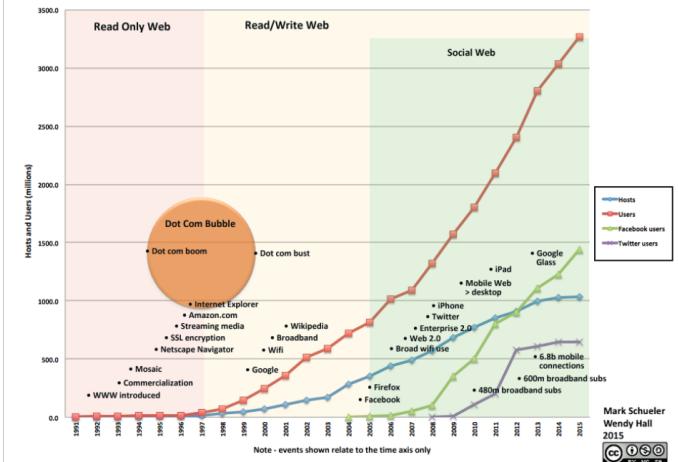
8

"Calm, focused, undistracted, the linear mind is being pushed aside by a new kind of mind that wants and needs to take in and dole out information in short, disjointed, often overlapping bursts—the faster, the better."

[Patricia Greenfield] concluded that "every medium develops some cognitive skills at the expense of others." Our growing use of the Net and other screen-based technologies has led to the widespread and sophisticated development of visual-spatial skills. We can, for example, rotate objects in our minds better than we used to be able to. But our "new strengths in visual-spatial intelligence" go hand in hand with a weakening of our capacities for the kind of "deep processing" that underpins "mindful knowledge acquisition, inductive analysis, critical thinking, imagination, and reflection."

— Nicholas Carr, *The Shallows: What the Internet is Doing to Our Brains*

Internet Growth - Usage Phases - Tech Events



Λίγη Ορολογία: Παγκόσμιο Πληροφοριακό Περιβάλλον

- "Web 2.0 (O'Reilly 2004): a supposed second-generation of Internet-based services such as **social networking sites**, **wikis**, communication tools, and **folksonomies** that let people **collaborate** and **share information** online in previously unavailable ways."

- Wikipedia

- **Mobile Internet**

- **Cloud Computing**

- **Internet of Things**



Key elements of the Web (revisited)

- **HTTP, URI and...**
- **Hypertext Markup Language (HTML)**: used for writing web pages
- **Cascading Style Sheets (CSS)**: stylistic info for web pages
- **PHP Hypertext Processor (PHP)**: dynamically create pages on a web server
- **JavaScript**: interactive and programmable web pages
- **Asynchronous JavaScript and XML (Ajax)**: accessing data for web applications
- **eXtensible Markup Language (XML)**: metalanguage for organizing data
- **Structured Query Language (SQL)**: interaction with databases

M. Λικείδης, ΕΠΙ425

12

Ιστορία και Εξέλιξη του Παγκόσμιου Ιστού

Ιστορική Εξέλιξη Υπερκειμένων

- Ο πρώτος ορισμός των Υπερκειμένων έγινε από τον Vannevar Bush (1890-1974), ο οποίος εισήγαγε τις θεμελιώδεις έννοιες των Υπερκειμένων για την περιγραφή του συστήματος **Memex**.
- Οι ιδέες γύρω από το Memex άρχισαν να αναπτύσσονται το 1932 - η πρώτη περιγραφή δημοσιεύθηκε το 1945 στο περιοδικό Atlantic Monthly, με τίτλο "As we May Think."
- <http://www.theatlantic.com/doc/194507/bush>



M. Λικαΐδης, ΕΠΙ425

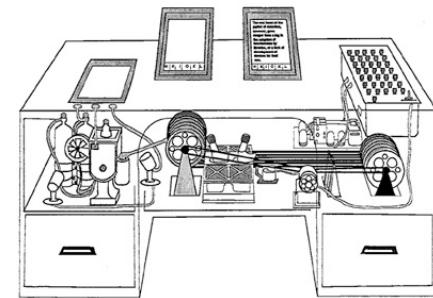
14

Η μηχανή MEMEX

- Ένα γραφείο που 'αποθηκεύει' 1 δισ. βιβλία, εφημερίδες και άλλα έγγραφα, και τις μεταξύ τους παραπομπές.
- Ενσωματωμένη κάμερα για λήψη και αποθήκευση εικόνων.
- Ένα κασετόφωνο για λήψη και αποθήκευση ήχων.
- Δυνατότητα "σχολιασμού" εγγράφων, με πρόσθεση παραπομπών ανάμεσα σε σχετικά έγγραφα.
- Δυνατότητα αναζήτησης κειμένων, είτε ακολουθώντας κάποια παραπομπή είτε με αναζήτηση λέξεων κλειδιών.

M. Λικαΐδης, ΕΠΙ425

15



M. Λικαΐδης, ΕΠΙ425



TED NELSON: Πρότεινε τους όρους «hypertext» και «hypermedia» το 1963, και τους πρωτοδημοσίευσε το 1965

"As early as 1960, Ted Nelson began to envision a world in which all of our media—documents, films, etc.—would be connected and interacting with one another on a vast system of computers. Nelson coined the term *intertwingled* to express ...

Ted Nelson and Doug Engelbart

DOUG ENGELBART: Εφηύρε το ποντίκι και ανέπτυξε τα πρώτα συστήματα υπερκειμένου.



Xanadu



WIKIPEDIA
The Free Encyclopedia

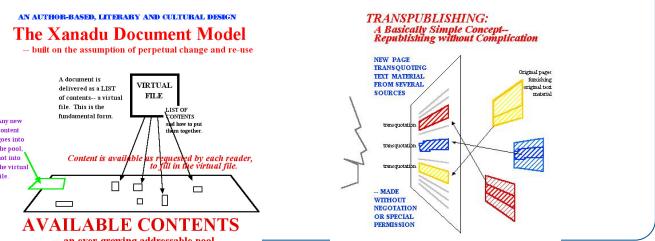
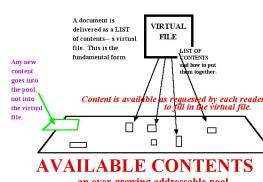


Original XANADU Rules:

- 1 Every Xanadu server is uniquely and securely identified.
- 2 Every document is uniquely and securely identified.
- 3 Every user is uniquely and securely identified.
- 4 Every document can consist of any number of objects.
- 5 Every document can consist of any number of parts each of which may be of any data type.
- 6 Every document can contain links of any type including virtual copies ("transclusions") to any other document in the system accessible to its owner.
- 7 Links are made by the document owner to any desired degree of granularity to ensure payment on any portion accessed, including virtual copies ("transclusions") of all or part of the document.
- 8 Every document is uniquely and securely identified.
- 9 Every document can have secure access controls.
- 10 Every document is uniquely and securely identified.
- 11 Every document can have secure access controls.
- 12 Every document can have secure access controls.
- 13 Every document is automatically moved to physical storage appropriate to its frequency of access from any given location.
- 14 Every document can be deleted.
- 15 Every Xanadu service provider can charge their users at any rate they choose for the storage, retrieval and publishing of documents.
- 16 Every transaction is secure and auditable only by the parties to that transaction.
- 17 The Xanadu client-server communication protocol is an openly published standard. Third-party software development and integration is encouraged.

AN AUTHOR-BASED, LITERARY AND CULTURAL DESIGN The Xanadu Document Model

— built on the assumption of perpetual change and re-use

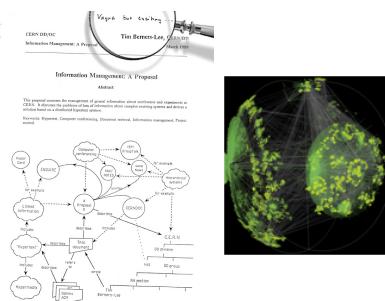
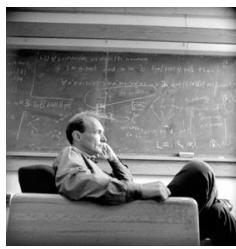


TRANSPUBLISHING: A Basically Simple Concept— Republishing without Complication

— MADE WITHOUT INTEGRATION OR SPECIAL PERMISSION

18

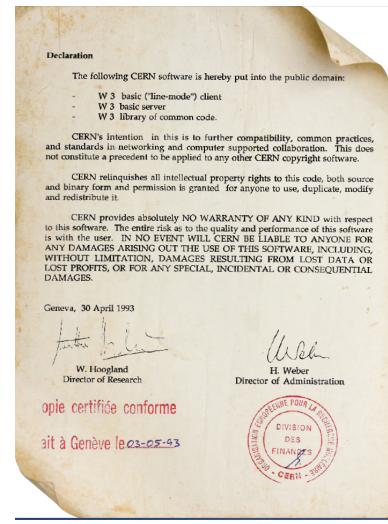
The Web – most successful information architecture in history



<http://www.w3.org/History.html> and

Tim Berners-Lee's presentation at the anniversary, <http://www.w3.org/2004/Talks/w3c10-HowItAllStarted/?n=1>

19



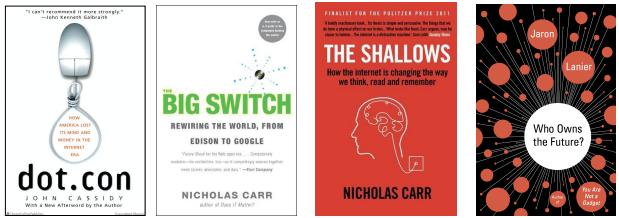
20

Προϊστορία του Παγκόσμιου Ιστού

1989-1990	O Tim Berners-Lee επινοεί το Web στο CERN της Γενεύης
11/1990	O Tim Berners-Lee κυκλοφορεί το πρώτο πρωτότυπο του WWW σε υπολογιστή NeXT
1/1992	Κυκλοφορεί ο κώδικας του πρώτου κειμενικού πλοηγού (textual browser), lynx, και του πρωτοκόλλου HTTP
3/93	Κυκλοφορεί από την NCSC ο Mosaic, ο πρώτος πλοηγός Ιστού
9/93	Το HTTP συμβάλλει στο 1% της διακίνησης δεδομένων στο Διαδίκτυο
12/94	Κυκλοφορεί ο πλοηγός Netscape 1.0 και ιδρύεται το W3C
8/95	Κυκλοφορούν τα Windows 95 και ο Internet Explorer 1.0
12/1995	Κυκλοφορεί η Java
12/04	Κυκλοφορεί ο πλοηγός Firefox 1.0
9/08	Κυκλοφορεί ο πλοηγός Chrome 1.0

M. Λαζαρίδης, ΕΠΙ425

21



WWW Consortium

- Ιδρύθηκε από τον Tim Berners-Lee τον 10/1994
- <http://www.w3.org>
- Στόχος: «να οδηγήσει τον Παγκόσμιο Ιστό στην επίτευξη της μέγιστης δυναμικότητάς του, αναπτύσσοντας κοινά πρωτόκολλα τα οποία θα προωθήσουν την εξέλιξή του και θα εγγυθούν την διαλειτουργικότητά του».
- Στο μάθημα θα μελετήσουμε πολλές από τις τεχνολογίες που καθοδηγούνται από το W3C
- W3C European Office - ERCIM



M. Λαζαρίδης, ΕΠΙ425

23

Διδάγματα

- **Big is beautiful:** the network is everything
- **Scruffy works:** let the links fail to make it scale
- **Democracy rules:** open, free and universal
- **But we lost (for a time) conceptual and contextual linking,** and **link descriptions** – the Web is a strangely linkless world
- **Missing links** – search engines fill the gap

M. Λαζαρίδης, ΕΠΙ425

Source: Wendy Hall, ACM Europe Workshop, Jan. 2012

24

ΣΤΑΤΙΣΤΙΚΕΣ για το Διαδίκτυο και το ΠΠΠ

ΣΤΑΤΙΣΤΙΚΕΣ για το Διαδίκτυο / ΠΠΠ

- CAIDA - www.caida.org
- Pew Research Center
- Kleiner, Perkins, Caufield, Byers (KPCB) Internet trends
- domaintools.com
- isc.com
- RIPE NCC (Regional Internet Registry) www.ripe.net
- Internet traffic measurements
 - <http://www.internettrafficreport.com>
 - <http://netflow.internet2.edu>

M. Λικαΐδης, EΠΙ425



26

Άλλες πηγές

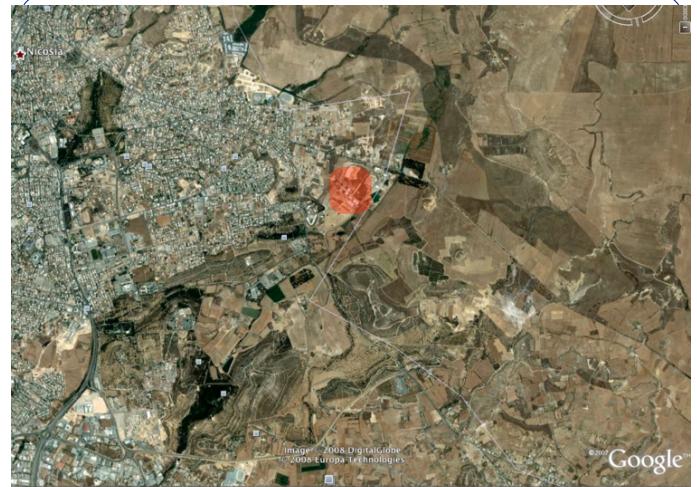
- Internet hosts Jul-15
 - <http://ftp.isc.org/www/survey/reports/current/>
- Internet users Jun-15
 - <http://www.internetworldstats.com/emarketing.htm>
- Facebook users Sep-15
 - <http://www.statisticbrain.com/facebook-statistics/>
- Twitter users Sep-15
 - <http://www.statisticbrain.com/twitter-statistics/>

Πηγή: <http://growthchart.weebly.com/>

M. Λικαΐδης, EΠΙ425

27

Μια επιτυχημένη ιστορία του Web..



SEPTEMBER 1ST-7TH 2007

www.economist.com

Sarkozy's 100 days
Why are mosques so hard to build?
America turns against the death penalty
A critical look at animal testing
The credit crunch and the economy

Who's afraid of Google?

Web Privacy China Antitrust News! Diversification Copyright more »
Don't be evil

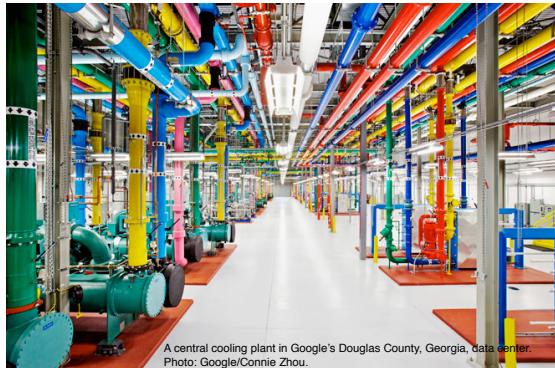
Google Search I'm Feeling Wary
Search: the web me

SEPTEMBER 28TH-OCTOBER 4TH 2016

Cameron's dodgy sums
Juncker's inadequate package
Riots in middle America
Airlines: low-cost takes on long-haul
The man who bought half Ukraine's debt

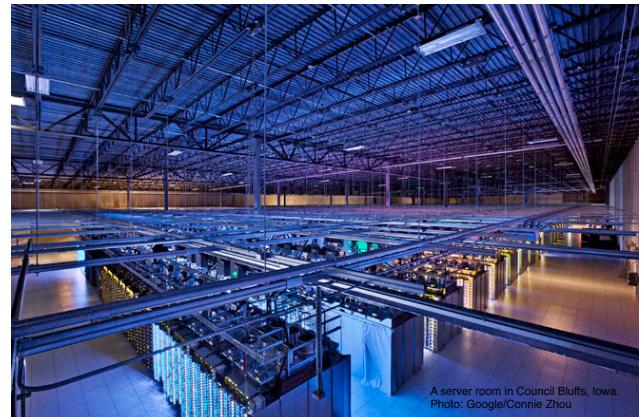


Should governments break up digital monopolies?



A central cooling plant in Google's Douglas County, Georgia, data center.
Photo: Google/Connie Zhou.

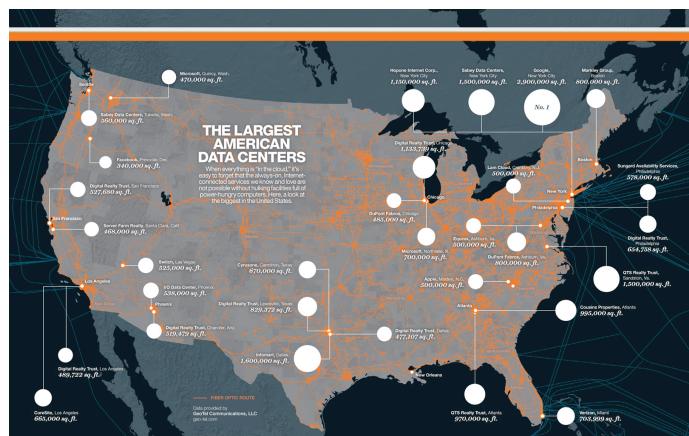
Google Throws Open Doors to Its Top-Secret Data Center. Wired Magazine, STEVEN LEVY. 10.17.12
<http://www.wired.com/wiredenterprise/2012/10/ff-inside-google-data-center/all/>



A server room in Council Bluffs, Iowa.
Photo: Google/Connie Zhou



Outside the Council Bluffs data center, radiator-like cooling towers chill water from the server floor down to room temperature.
Photo: Google/Connie Zhou



To Google σε αριθμούς (\$)- 2007

- Αξία: 160 δις \$
- Έσοδα (2007): 16 δις \$
- Κέρδη (2007): 4.3 δις \$
 - "All that money comes 50 cents at a time"
 - "Costs are mostly fixed, so any incremental revenue is profit... The machinery that represents the fixed costs is Google's secret sauce."
 - "Google has built, in effect, the world's largest supercomputer. It consists of enormous data-centers around the world. The details are Google's best-guarded secret".
 - The result is to provide a "cloud" of computing power that is flexible enough "automatically to move load around between data-centers."
- Google's official strategy: "search, ads, and apps"

M. Λικαΐδης, ΕΠΙ4425

37

Wikipedia on Google

Google search

From Wikipedia, the free encyclopedia

Google is owned by Google, Inc., whose mission statement is to "organize the world's information and make it universally accessible and useful". The largest search engine on the web, Google receives several hundred million queries each day through its various services.

Contents	[edit]
1 The search engine	
1.1 Pages indexed	
1.2 Search results	
1.3 Advanced search operators	
1.4 Non-Web sources of data	
1.5 Google optimization	
1.6 Uses of Google	
1.7 Google dance	
1.8 Search engine features	
1.9 Error messages	
2 Search products	
3 Google jargon	
4 Google games	
5 See also	
...	

* Find out more about navigating Wikipedia and finding information *

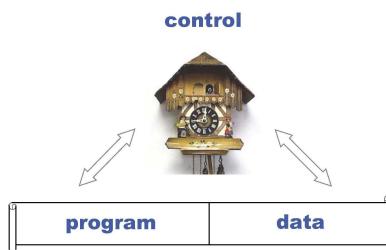


M. Λικαΐδης, ΕΠΙ4425

38

Η θεμελιώδης «εξίσωση» της Πληροφορικής

- Computing = data + program + control



M. Λικαΐδης, ΕΠΙ4425

39

EPL 61

http://www.metafrasi.gr/ iCal Exchange Listserv UCY Mobility Portal md

- World-Wide Web Consortium

Journals

- ACM Transactions on Internet Technology
- World-Wide Web Journal, by Baltzer.

Proceedings of the International WWW Conferences (IW3C Committee)

- The Fifteenth (2006) International WWW Conference
- The Fourteenth (2005) International WWW Conference
- The Thirteenth (2004) International WWW Conference
- The Twelfth (2003) International WWW Conference
- The Eleventh (2002) International WWW Conference
- The Tenth (2001) International WWW Conference
- The Ninth (2000) International WWW Conference
- The Eighth (1999) International WWW Conference
- The Seventh (1998) International WWW Conference
- The Sixth (1997) International WWW Conference
- The Fifth (1996) International WWW Conference
- The Fourth (1996) International WWW Conference

Other Conferences of Interest

- USENIX Symposium on Internet Technologies and Systems (USITS)
- Workshop on Web Caching and Content Distribution
- Performance and Architecture of Web Servers (PAWS)

40

Google: το πρόγραμμα

Google: ο «έλεγχος» (circa 1999)



M. Λικαΐδης, ΕΠΙ4425

42

Proceedings of the Seventh World Wide Web Conference

Next | Contents | Indexes

The Anatomy of a Large-Scale Hypertextual Web Search Engine

Sergey Brin and Lawrence Page

Computer Science Department,
Stanford University, Stanford, CA 94305, USA
sergey@cs.stanford.edu and page@cs.stanford.edu

Abstract
In this paper we present Google, a prototype of a large-scale search engine. Google is designed to crawl and index the Web efficiently and produce near-instantaneous search results. It is probably the first search engine which is able to crawl and index the entire Web (as of June 2000, Google's index contains over 24 million pages). Available at <http://infolab.stanford.edu/>. To engineer a search engine is a challenging task. Search engines index tens to hundreds of millions of web pages involving a comparable number of distinct terms. They answer tens of millions of queries every day. Despite the importance of large-scale search engines on the web, very little academic research has been done on them. Furthermore, due to rapid advances in technology and new applications, creating a web search engine today is a very different task than it was a few years ago. This paper provides an in-depth analysis of the Google search engine—the first such document available to date. We learn a lot from the problems of building a general search engine. There are new technical challenges involved with using the additional information present in hypertext to produce better search results. This paper addresses this question of how to build a practical large-scale system which can exploit the additional information present in hypertext. Also we look at the problem of how to effectively deal with unstructured hypertext collections where anyone can publish anything they want.

Keywords
World Wide Web, Search Engines, Information Retrieval, PageRank, Google

M. Λικαΐδης, ΕΠΙ4425

41

Google: τα δεδομένα

- 11.5 (?) δισεκατομύρια έγγραφα κατανεμημένα σε εκατομύρια διαδικτυακούς κόμβους, με εκατοντάδες εκατομύρια χρήστες διεθνώς... (2007)

M. Δικαιόκος, ΕΠΙ4425

43

'Άλλοι «καθοδηγητές»;

- Google
- Facebook
- Amazon
- Apple
- Microsoft
- Wikipedia
- ???
- ???
- ???

M. Δικαιόκος, ΕΠΙ4425

44

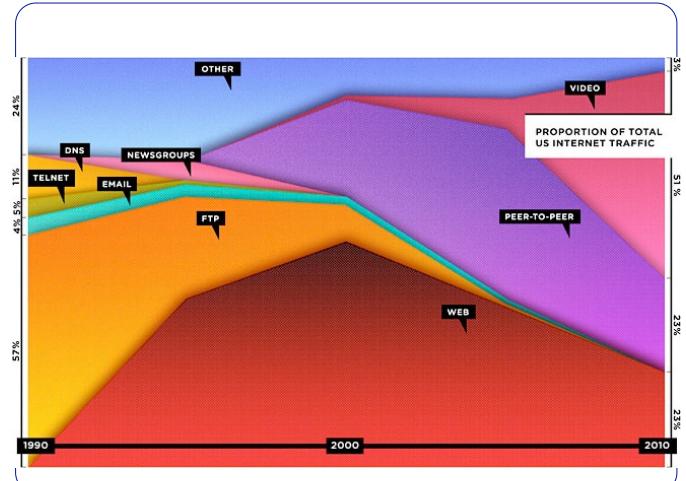
"The Web Is Dead. Long Live the Internet"



"The Web is dead." Anderson & Wolff, WIRED, Aug. 2010
http://www.wired.com/magazine/2010/08/ff_webrip/

M. Δικαιόκος, ΕΠΙ4425

45



The End of the Web, Search, and Computer as We Know It

"the web — soon to become the cybersphere — will no longer resemble a chaotic cobweb. It's already started to happen. Instead, billions of users will spin their own tales, which will merge seamlessly into an ongoing, endless narrative: the earth telling its own story."

D. Gelernter

<http://www.wired.com/opinion/2013/02/the-end-of-the-web-computers-and-search-as-we-know-it/>

M. Δικαιόκος, ΕΠΙ4425

47

THE WALL STREET JOURNAL

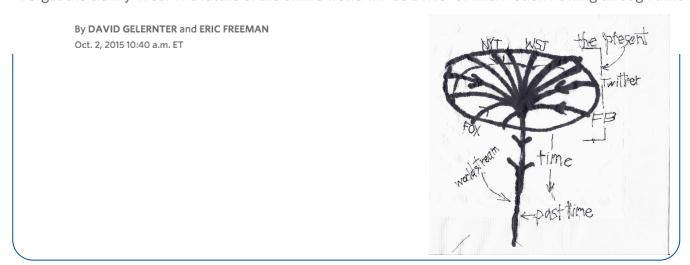


LIFE | IDEAS | ESSAY

The Future of the Internet Is Flow

Forget the clunky Web. The future of the online world will be a river of information flowing through time

By DAVID GELERNTER and ERIC FREEMAN
Oct. 2, 2015 10:40 a.m. ET

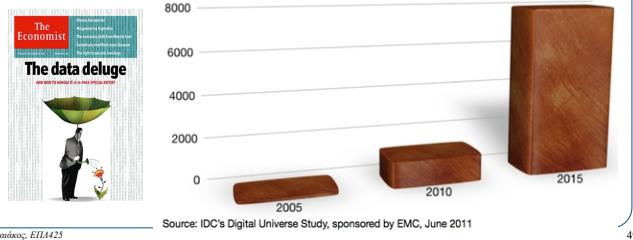


M. Δικαιόκος, ΕΠΙ4425

Data Deluge: flooding by information

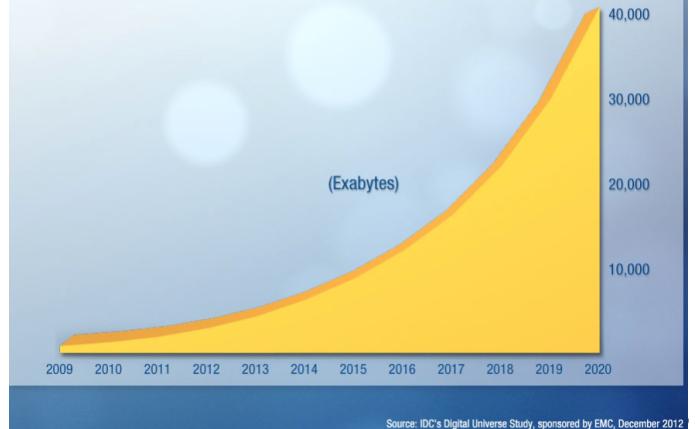
- Our world is creating more bits than ever: in 2011, more than 1.8 zettabytes (1.8×10^{21}), growing by a factor of 9 in 5 years
- This data is transported and made available through the Internet

A Decade of Digital Universe Growth:
Storage in Exabytes



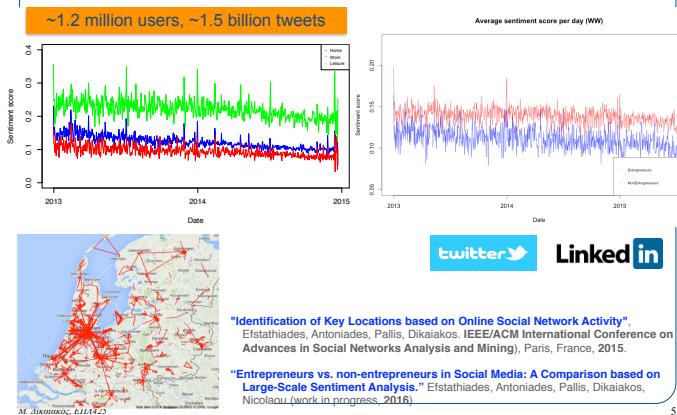
49

The Digital Universe: 50-fold Growth from the Beginning of 2010 to the End of 2020



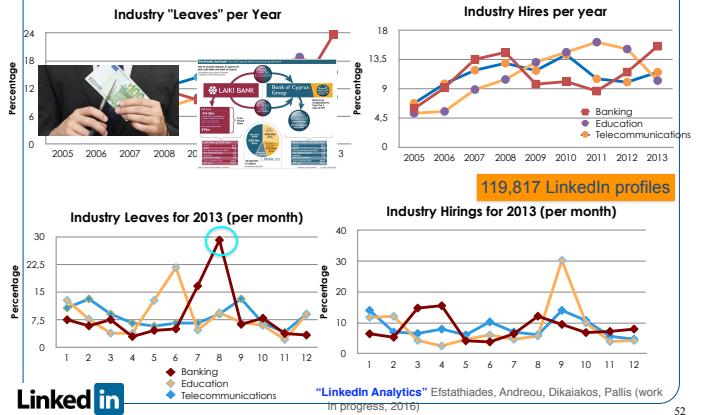
50

From pure & applied to social sciences



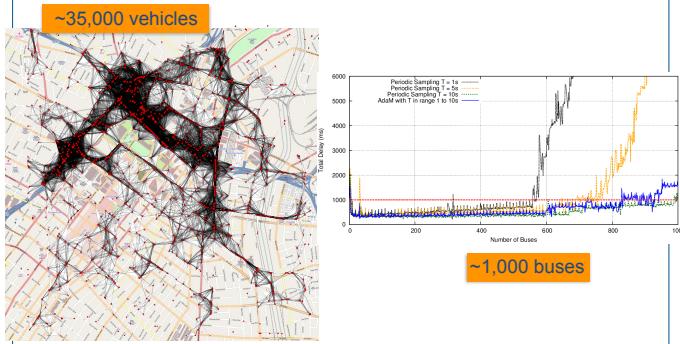
51

From pure & applied to social sciences



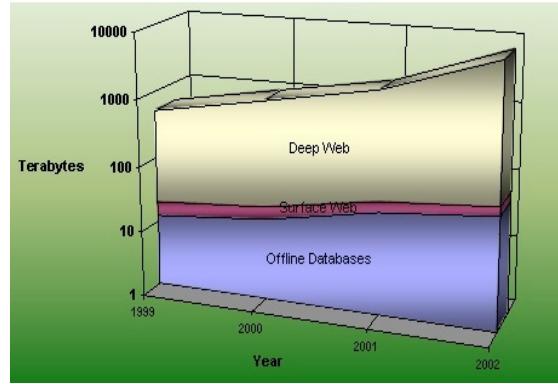
52

From pure & applied to social sciences



53

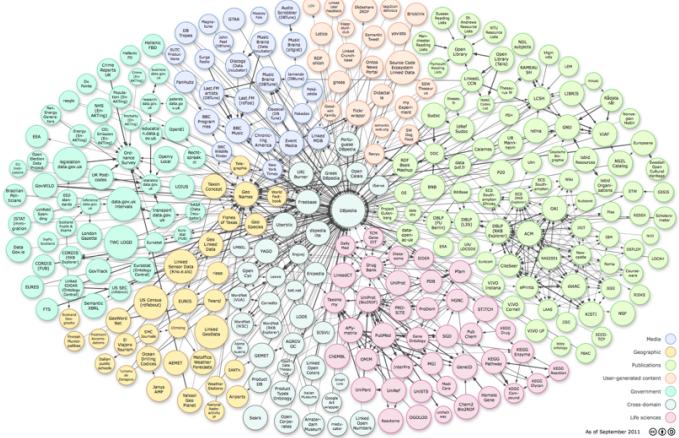
Deep Web



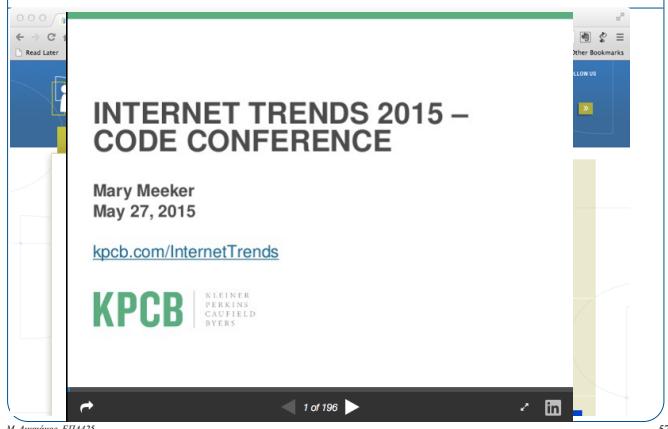
M. Assilakos, EIT4425

54

Linked Data



Τρέχουσα κατάσταση και τάσεις



M. Δικαιόκος, ΕΠΑ425

Διαδίκτυο, Ιστός, ΠΠΠ

- Διαδικτυακό Σύστημα Ανάκτησης Υπερμεσικών Πληροφοριών
 - A wide-area **hypertext, multimedia information retrieval system** that provides access to a large universe of documents and **data**
 - Ενιαίος τρόπος πρόσβασης και απεικόνισης πληροφοριών στο Διαδίκτυο
 - A **uniform way** of **accessing** and **viewing** some **information** on the Internet
 - Ένας χώρος μέσα στον οποίο κάθε πληροφοριακό στοιχείο αποκτά **ταυτότητα**, μέσω της οποίας μπορούμε να αποκτήσουμε πρόσβαση επάνω του
 - A world in which **information has a reference** by which it can be **accessed**

M. Δικαιάκος, ΕΠΑ42:

56

ΟΙ «ΟΡΑΜΑΤΙΣΤΕΣ»



M. Λικαιάκος, ΕΠΛ42.

58

*Ποιός θεωρείται ο επινοητής του
Web;*

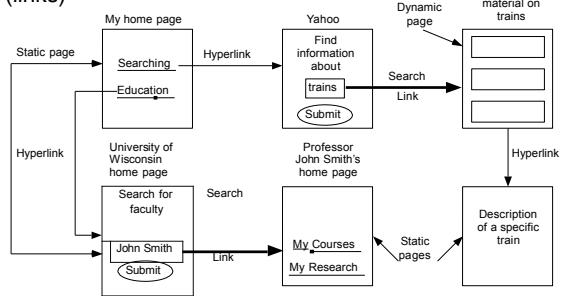
- O Doug Engelbart
 - O Vannevar Bush
 - O Tim Berners Lee
 - O Alan Turing

EPA425

Ο Ιστός ως γράφος

Μοντέλο Δεδομένων Παγκόσμιου Ιστού

- Κατευθυνόμενος γράφος (graph) όπου οι κόμβοι είναι έγγραφα/αρχεία (documents) και οι ακμές είναι ζεύξεις (links)



M. Λικαΐδης, ΕΠΙ4425

61

Η γραφοειδής μορφή του Ιστού

- Κόμβοι = στατικές ιστοσελίδες ~1 δις (circa 2009)
- Ακμές = στατικές ζεύξεις (υπερσύνδεσμοι) ~10 δις
- «Αραιός» γράφος: ~ 7 ακμές/σελίδα κατά μέσο όρο
- Μερικά ερωτήματα:
 - Είναι συνεκτικός γράφος; Μπορούμε πάντοτε να τον διασχίσουμε από τον ένα κόμβο στον άλλο;
 - Μπορεί η γνώση της συνδεσμολογίας των ζεύξεων να βελτιώσει την αποδοτικότητα της αναζήτησης (search);
 - Αν παρακολουθήσουμε την αλλαγή του γράφου ανά το χρόνο, τι συμπεράσματα μπορούμε να βγάλουμε για τις διεργασίες παραγωγής ιστιακού περιεχομένου;

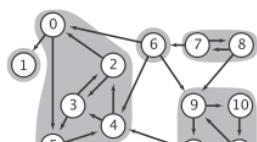
A. Broder, R. Kumar, et al. "Graph structure in the web." 9th WWW Conference, 2000. <http://www9.org/w9cdrom/160/160.html>

M. Λικαΐδης, ΕΠΙ4425

62

Μερικοί αλγόριθμοι για γράφους

- Ισχυρή Συνεκτική Συνιστώσα (Strongly connected component/SCC)**
 - Ένας μέγιστος υπογράφος ενός κατευθυνόμενου γράφου, τέτοιος ώστε όλες οι κόμβων (u, v), υπάρχει ένα κατευθυνόμενο μονοπάτι από το u στο v και ένα κατευθυνόμενο μονοπάτι από το v στο u.
- Ασθενής Συνεκτική Συνιστώσα (Weakly connected component/WCC)**
 - Ένας μέγιστος υπογράφος ενός κατευθυνόμενου γράφου, τέτοιος ώστε όλες οι κόμβων (u, v), υπάρχει ένα μη κατευθυνόμενο μονοπάτι από το u στο v και ένα κατευθυνόμενο μονοπάτι από το v στο u.
 - Ο υπολογισμός των WCC και SCC ενός γράφου μπορεί να γίνεται με γραμμική πολυπλοκότητα.



A digraph and its strong components

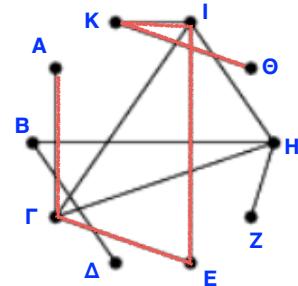
M. Λικαΐδης, ΕΠΙ4425

ρου μικρότερου μονοπατιού» ανάμεσα σε βους του γράφου, ή αλλοιώς κόμβων, τους οποίους πρέπει να με από έναν κόμβο σε κάποιον άλλο, ήσεις, παρακάμψεις ή βρόχους.

63

Πόση είναι η διάμετρος του εικονιζόμενου γράφου;

- 1
- 6
- 5
- 3



ΕΠΙ4425

Το πρόβλημα της κλίμακας

- Ένας τυπικός αλγόριθμος υπολογισμού της διαμέτρου ενός γράφου χρειάζεται **Ο(κόμβοι x ακμές)** βήματα για να τερματίσει.
 - Για τον ιστό: ~ (ιστοσελίδες x ζεύξεις)
- Για 1 δις ιστοσελίδες, 10 δις ζεύξεις και περίπου 0.1 μsec/βήμα, θα χρειαστούν 1δις δευτερόλεπτα, δηλαδή 10 εκατομύρια ημέρες...
- Η σάρωση του Ιστού από την AltaVista τον Μάιο 1999 έδωσε:
 - 220 εκατομύρια ιστοσελίδες (μετά την αφαίρεση αντιγράφων)
 - Ένα τεράστιο WCC με ~186 εκατομύρια ιστοσελίδες
 - Ένα τεράστιο SCC με ~56 εκατομύρια ιστοσελίδες
 - Το αμέσως επόμενο SCC είχε ~156 Κ ιστοσελίδες
 - Ανεξάρτητες σαρώσεις έδωσαν παρόμοια αποτελέσματα

M. Λικαΐδης, ΕΠΙ4425

65

Το πρόβλημα της «προσβασιμότητας»

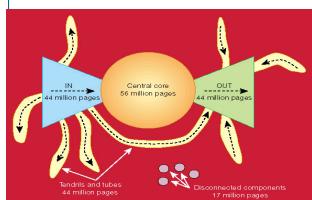
- Πόσες ιστοσελίδες είναι προσβάσιμες από μια τυχαία ιστοσελίδα;
- Ξεκίνα από μια τυχαία ιστοσελίδα
 - Βρεις τους γείτονές της και βάλε τους στη λίστα
 - Επανέλαβε τη διαδικασία για κάθε γείτονα, αποφεύγοντας τους βρόχους και τα αδιέξοδα
- Υπολόγισε τον αριθμό των προσβάσιμων ιστοσελίδων σαν συνάρτηση της απόστασης d από την αρχική ιστοσελίδα.
- Πειραματισμός: ξεκινώντας από 1000 τυχαίες ιστοσελίδες, κατασκευάστε για την κάθε μιά το BFS δένδρο της.
- Παρατήρηση:
 - Κάποιες αναζητήσεις «πεθαίνουν» γρήγορα
 - Κάποιες άλλες μας οδηγούν σε «συνδυαστική έκρηξη» (combinatorial explosion), φθάνοντας σε 100 εκατομύρια ιστοσελίδες.

M. Λικαΐδης, ΕΠΙ4425

66

Ανατομία του Παγκόσμιου Ιστού

- Τυχαιότητα;
- Κατανομή εισερχομένων ακμών
- Κατανομή εξερχομένων ακμών
- Χαρακτηριστικά κόμβων
 - Υπάρχουν κάποιοι κόμβοι πιο “κεντρικοί” από άλλους
- Δυναμική συμπεριφορά
 - Power laws, scale-free behavior



A. Broder, R. Kumar, et al.
“Graph structure in the
web.” 9th WWW
Conference, 2000.