



# ΕΠΛ602: Θεμελιώσεις Τεχνολογιών Διαδικτύου

Φθινοπωρινό Εξάμηνο 2007  
Πέμπτη 18:00-20:30, ΧΩΔ01-003  
Παρασκευή 19:30-21:00, ΘΕΕ01-201

Διδάσκων: Μάριος Δικαιάκος

Βοηθός: Ανδρέας Καμηλάρης

<http://www.cs.ucy.ac.cy/~epl602>  
<http://www.cs.ucy.ac.cy/courses/EPL602>



# Στόχος: Μελέτη Βασικών Πτυχών του Παγκόσμιου Πληροφοριακού Περιβάλλοντος (WWW)



Chenoweth

Dallesport

City of The Dalles

Celilo

The Dalles

Image State of Oregon  
© 2008 Tele Atlas

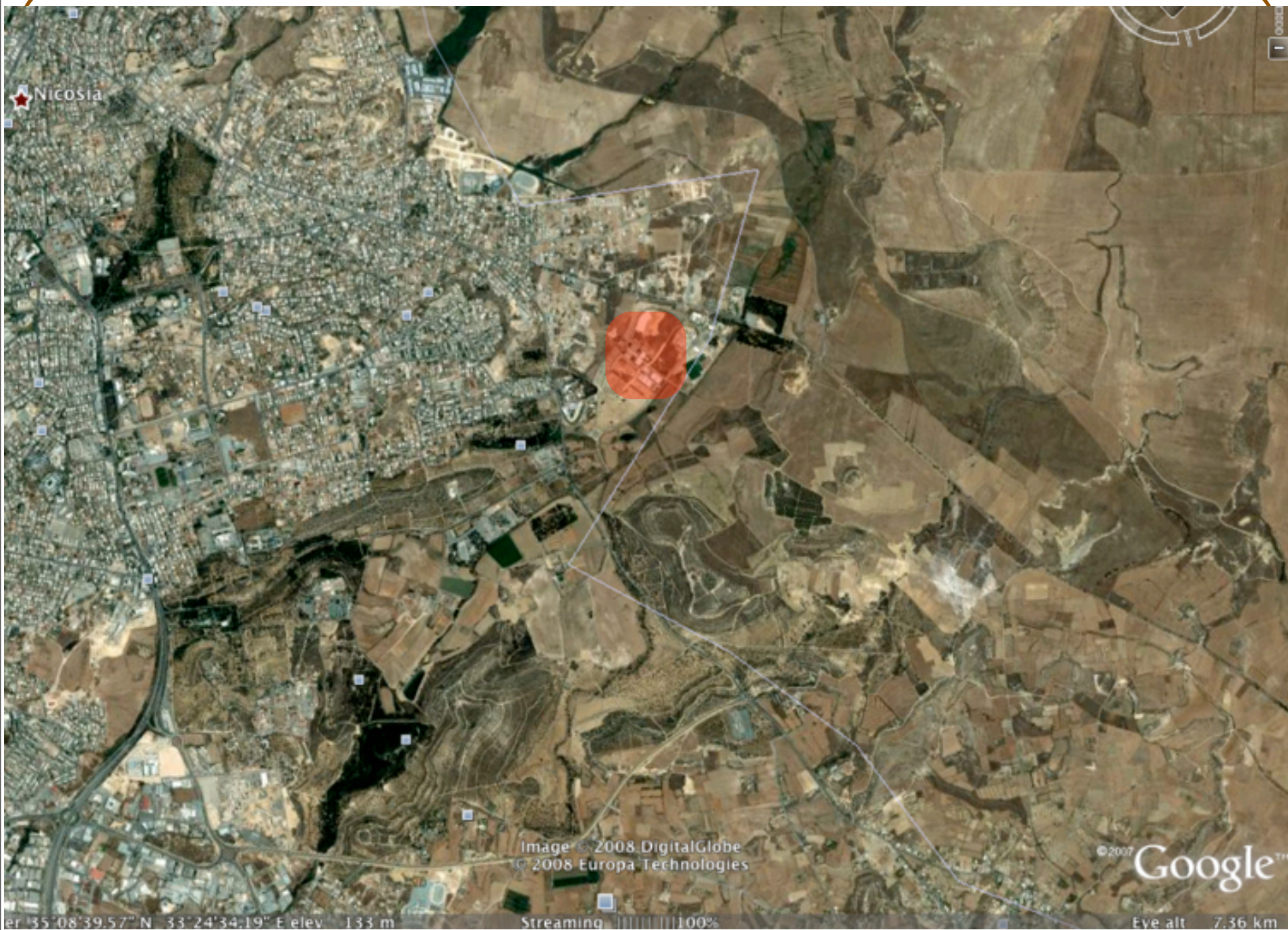
© 2007 Google™

inter 45°36'46.53" N 121°10'41.30" W elev 64 m

Streaming 100%

Eye alt 7.64 km





Nicosia

Image © 2008 DigitalGlobe  
© 2008 Europa Technologies

© 2007 Google™

er 35°08'39.57" N 33°24'34.19" E elev 133 m

Streaming 100%

Eve alt 7.36 km



# The Economist

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](#) **New!** [Diversification](#) [Copyright](#) [more »](#)

Don't be evil

Google Search

I'm Feeling Wary

Search: ☒ the web ☐ me

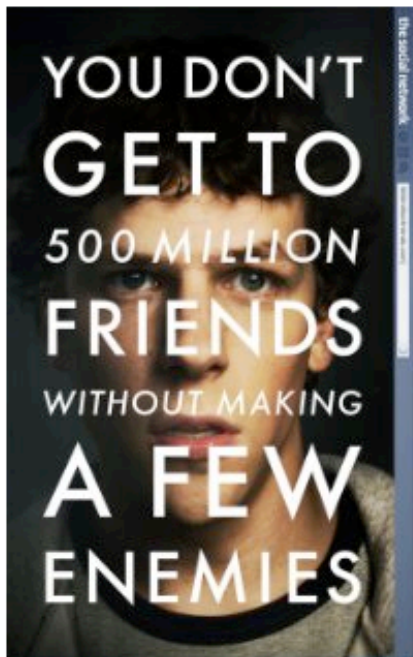
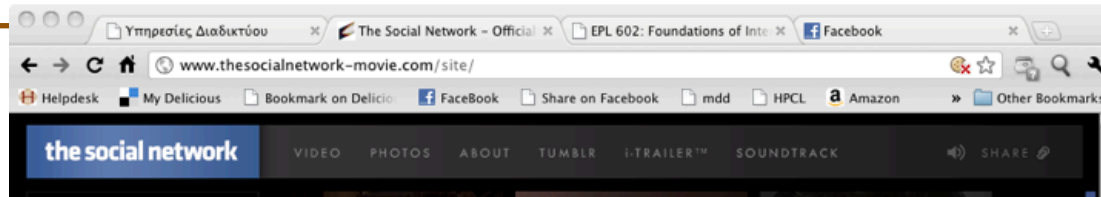


# To Google to 2007



- 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 datacenters 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 datacenters.”
- Google’s official strategy: “search, ads, and apps”

# Who is Google afraid of?



## The Social Network (2010)

[More at IMDbPro](#) »

**PG-13** 120 min - [Biography](#) | [Drama](#) | [History](#)  
- [1 October 2010 \(USA\)](#)

★★★★★☆☆☆☆☆ 8.2/10

Users: (71,330 votes) 367 reviews | Critics: 352 reviews  
Metascore: 95/100 (based on 42 reviews from [Metacritic.com](#))

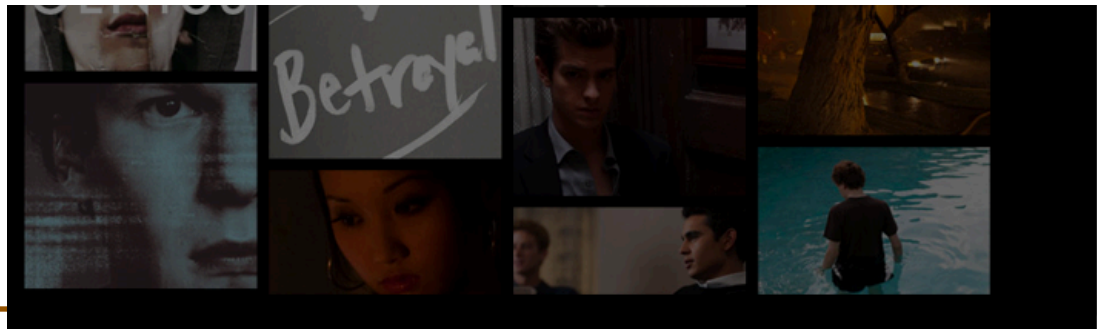
A chronicle of the founding of Facebook, the social-networking Web site.

**Director:** [David Fincher](#)

**Writers:** [Aaron Sorkin](#) (screenplay), [Ben Mezrich](#) (book)

**Stars:** [Jesse Eisenberg](#), [Andrew Garfield](#) and [Justin Timberlake](#)

[Watch Trailer](#) »

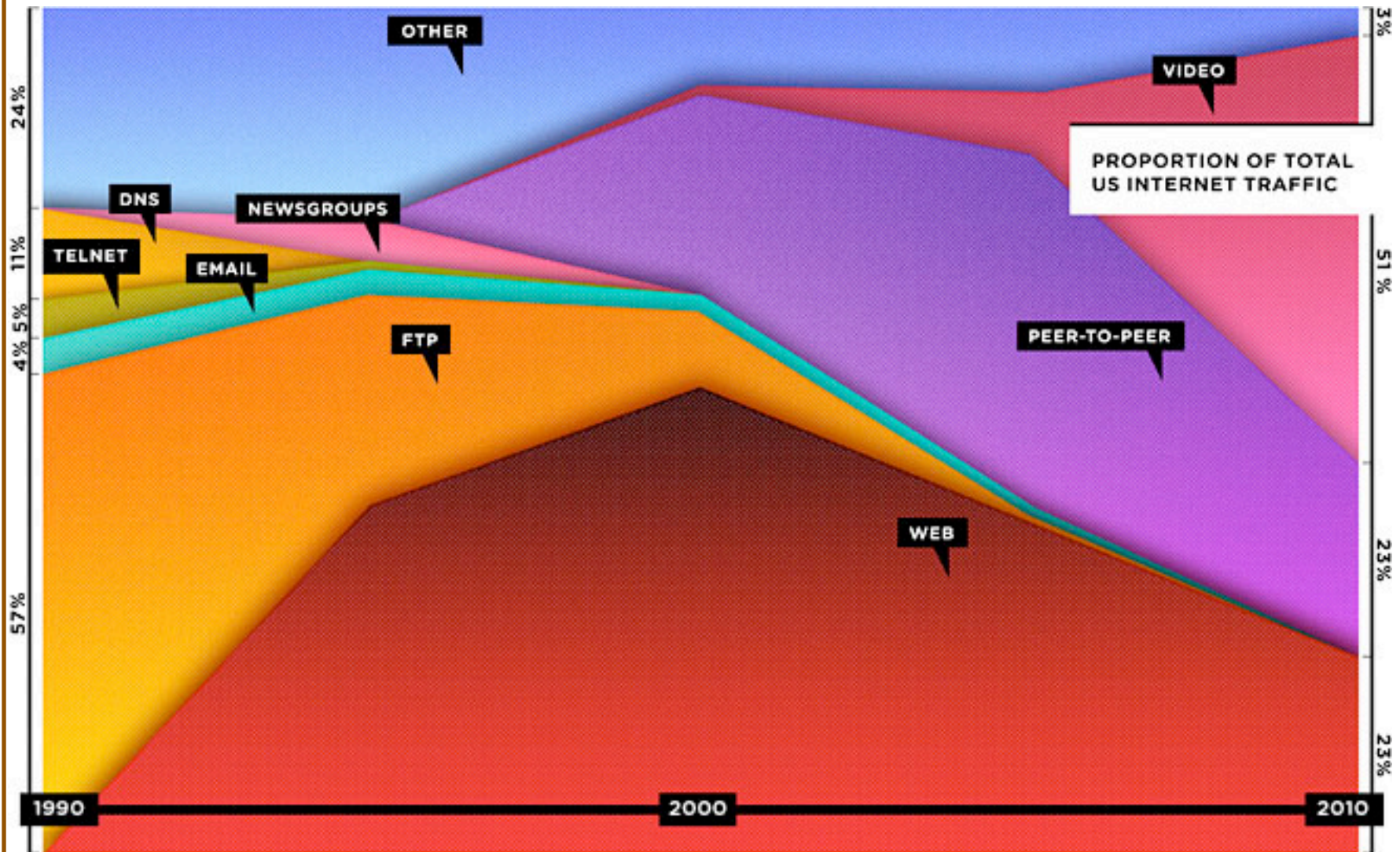


# “The Web Is Dead. Long Live the Internet”





# “The Web Is Dead. Long Live the Internet”



# Wikipedia on Google



## Google search

• Find out more about navigating Wikipedia and finding information •

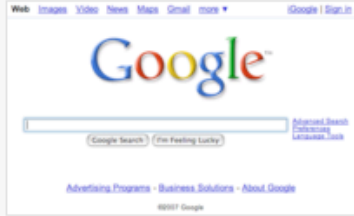
From Wikipedia, the free encyclopedia

**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 [\[hide\]](#)

- 1 The search engine
  - 1.1 PageRank
  - 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

### Google search



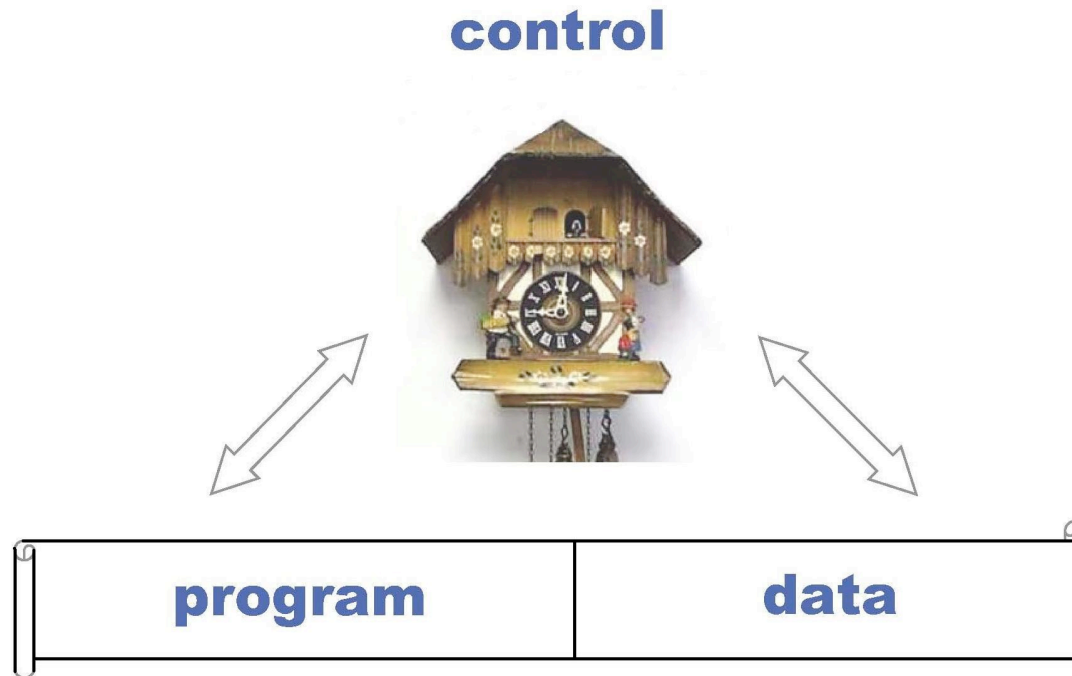
The Google homepage in Safari

<b>Developer:</b>	Google Inc.
<b>Available language(s):</b>	multilingual (~100)
<b>Genre:</b>	Search Engine
<b>License:</b>	Freeware
<b>Website:</b>	<a href="http://www.google.com">www.google.com</a> 

# Η θεμελιώδης εξίσωση



Computing = data + program + control







- [World-Wide Web Consortium](#)

### Journals

- [ACM Transactions on Internet Technology](#)
- [World-Wide Web](#) Journal, by Baltzer.

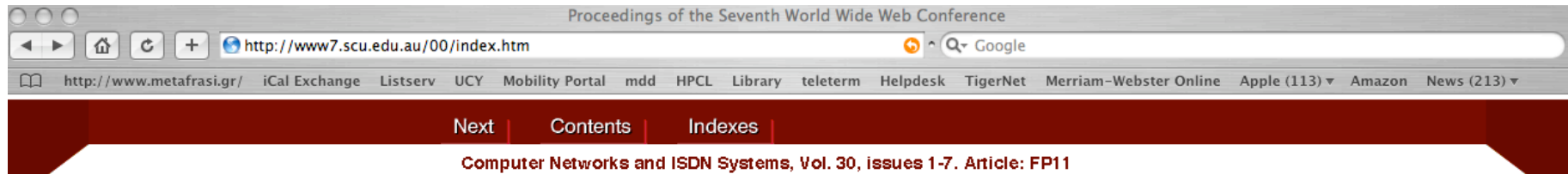
### Proceedings of the International WWW Conferences ([IW3C Committee](#))

- [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 \(PAWS\)](#)

# Google: το πρόγραμμα



## 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](mailto:sergey@cs.stanford.edu) and [page@cs.stanford.edu](mailto:page@cs.stanford.edu)*

### Abstract

In this paper, we present Google, a prototype of a large-scale search engine which makes heavy use of the structure present in hypertext. Google is designed to crawl and index the Web efficiently and produce much more satisfying search results than existing systems. The prototype with a full text and hyperlink database of at least 24 million pages is available at <http://google.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 advance in technology and web proliferation, creating a web search engine today is very different from three years ago. This paper provides an in-depth description of our large-scale web search engine -- the first such detailed public description we know of to date. Apart from the problems of scaling traditional search techniques to data of this magnitude, 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 uncontrolled hypertext collections where anyone can publish anything they want.

### Keywords

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

# Google: ο έλεγχος (circa 1999)





# Google: τα δεδομένα



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



# **Πώς μελετάμε/διδάσκουμε το ΠΠΠ;**



- Παγκόσμιος Ιστός ή *Παγκόσμιο Πληροφοριακό Περιβάλλον (ΠΠΠ)*:
  - Η ευρύτερα χρησιμοποιούμενη και ριζοσπαστικότερη εφαρμογή στην ιστορία του Υπολογισμού και της ανθρώπινης επικοινωνίας:
    - Στα ΑΕΙ, τροποποίησε τον τρόπο με τον οποίο διδάσκουμε, επικοινωνούμε, δημοσιεύουμε, κάνουμε έρευνα.
    - Δημιούργησε νέους βιομηχανικούς τομείς αλλά και άλλαξε την επικοινωνία και την παροχή υπηρεσιών σε ολόκληρους κλάδους της βιομηχανίας.
    - 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
  - Οι κοινωνικές αλληλεπιδράσεις που προκαλούνται από το ΠΠΠ, δημιουργούν νέες απαιτήσεις από τις εφαρμογές του ΠΠΠ, οι οποίες με τη σειρά τους επηρεάζουν τις απαιτήσεις από τις υποδομές του ΠΠΠ.



# The impact of Computing on the Web



- Computer Networking: protocols, algorithms, architectural principles
- Computer Systems: resource management, engineering of software, architectures, programming systems, parallel computation
- Information Systems: hypertext, data modeling, data representation, data management, information retrieval
- (new area of) Social computing: tagging, ontologies, data integration, information retrieval

# Η μελέτη του Παγκόσμιου Ιστού



Despite the Web's great success as a technology and the significant amount of computing infrastructure on which it is built, it remains, as an entity, surprisingly unstudied.

Hendler, Shadbolt, Hall, Berners-Lee and Weitzner  
CACM, July 2008, vol. 51, No. 7

ΕΦΑΡΜΟΓΕΣ

ΔΙΚΤΥΑ Η/Υ

# Η διδασκαλία του ΠΠΠ



- Similarly, if you look at CS curricula in most universities worldwide you will find “Web design” is taught as a service course, along with, perhaps, a course on Web scripting languages.
- You are **unlikely** to find a course that teaches **Web architecture or protocols**. It is as if the Web, at least below the browser, simply does not exist.
- The **protocols**, **architectures**, and **underlying principles** of the Web per se are rarely covered.

**"Web Science: An Interdisciplinary Approach to Understanding the Web."** James Hendler, Nigel Shadbolt, Wendy Hall, Tim Berners-Lee, and Daniel Weitzner. *Communications of the ACM*. Vol. 51, No. 7, **July 2008**. pp. 60-69.

# Κατανοώντας το ΠΠΠ



- Ένα σύνολο από πρωτόκολλα, τα οποία μπορούν να μελετηθούν για τις ιδιότητές τους και...
- ένα σύνολο εφαρμογών που μπορούν να αναλυθούν για τις αλγοριθμικές τους ιδιότητες
- Ωστόσο, ο Ιστός *δεν κατασκευάστηκε ακολουθώντας τον παραδοσιακό κύκλο ζωής των προϊόντων της Πληροφορικής:*
  - Προδιαγράφω (specify),
  - Σχεδιάζω (design)
  - Κατασκευάζω (build)
  - Ελέγχω (test)



# Φυσικές Επιστήμες και Πληροφορική



- Μια σχέση που τροποποιήθηκε από το ΠΠΠ:
  - Φυσικές Επιστήμες: αναλυτικές επιστήμες που επιδιώκουν να ανακαλύψουν νόμους που *διέπουν ή προκαλούν* παρατηρούμενα φαινόμενα
  - Πληροφορική: κυρίως συνθετική επιστήμη, στην οποία θεωρητικοί φορμαλισμοί και αλγόριθμοι δημιουργούνται για να επιτευχθούν επιθυμητές συμπεριφορές (προγραμμάτων και μηχανών)
- Ο Παγκόσμιος Ιστός:
  - Είναι ένα φαινόμενο που χρειάζεται να το αναλύσουμε και να ανακαλύψουμε τους νόμους που το διέπουν
  - Είναι μια επινόηση που πρέπει να την «κατασκευάσουμε» βέλτιστα ώστε να διασφαλίσουμε την μελλοντική της ανάπτυξη

# Από το «μικροσκοπικό» στο «μακροσκοπικό»

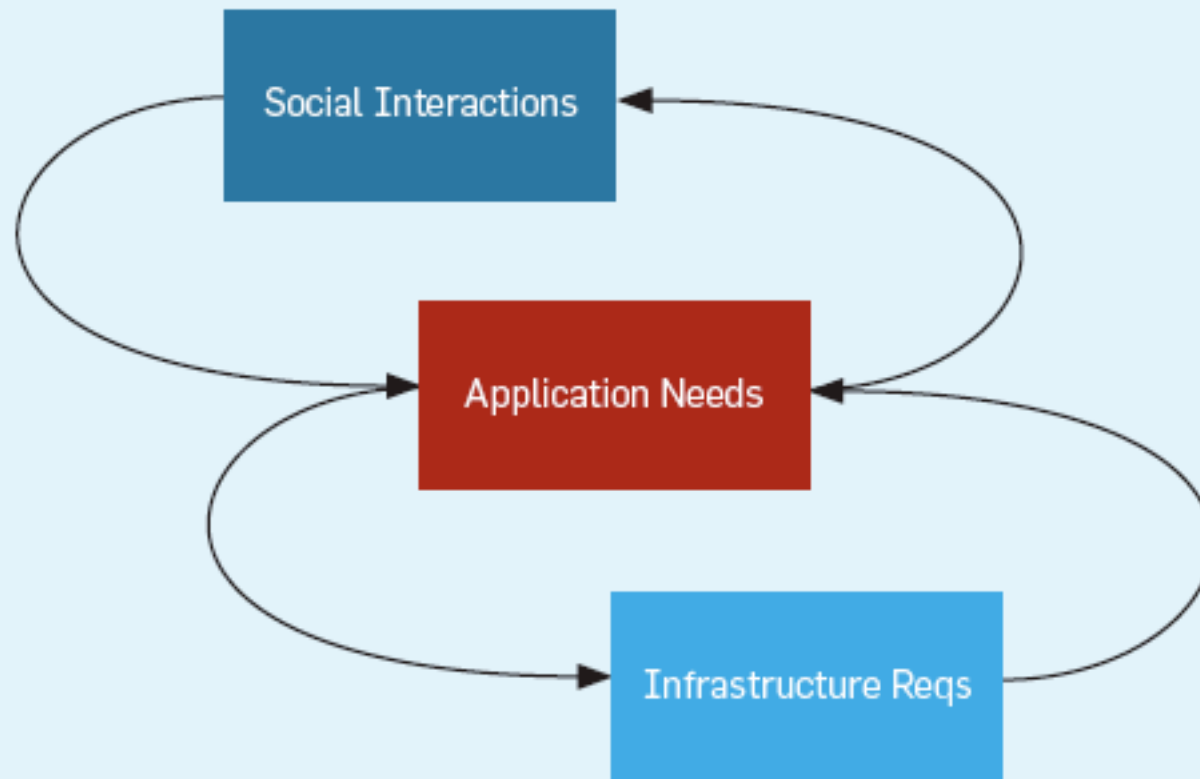


- Σε μικροσκοπικό επίπεδο:
  - το ΠΠΠ είναι μια υποδομή γλωσσών, πρωτοκόλλων και δεδομένων - ένα προϊόν μηχανοτεχνίας (engineering)
    - Πρωτόκολλα - που μελετούνται για τις ιδιότητές τους
    - Εφαρμογές - που μελετούνται για την αλγοριθμική συμπεριφορά τους
    - Υποδομή - που μελετάται για την αρχιτεκτονική και την επίδοσή της
- Το μακροσκοπικό ΠΠΠ προκύπτει από την χρήση «μικροσκοπικών» ΠΠΠ τεχνολογιών από πολλούς χρήστες για την αλληλεπίδραση με άλλους χρήστες και συστήματα
  - Η συσσώρευση πολλών χρηστών δημιουργεί νέες προκλήσεις που δεν μπορούμε εκ των προτέρων να προβλέψουμε

# Από το Mosaic στο Google



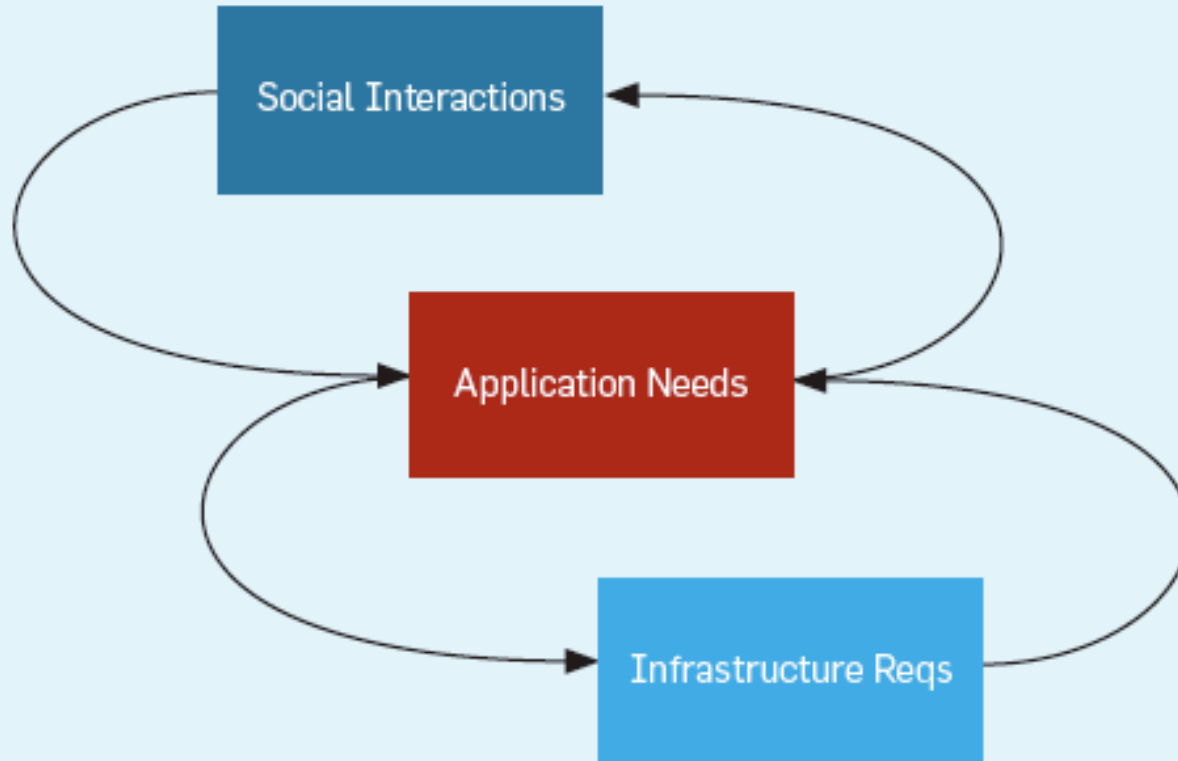
- Mosaic: the first popular Web browser (circa 1992) - παράδειγμα «μίκρο»-συστήματος
  - Πάνω από 1M καταφορτώσεις τον πρώτο χρόνο
  - Αντίστοιχα παραδείγματα: Youtube, Flickr, Facebook
- Από το «μίκρο»-σύστημα στο «μάκρο»-σύστημα
  - Εκατομύρια χρήστες
  - Διαφορετική ανάλυση
  - Νέες ανάγκες (π.χ. αναζήτηση πληροφοριών)
  - Τεχνικές για εξαπάτηση των αλγορίθμων διαβάθμισης των μηχανών αναζήτησης (search spam)
  - Νέες ανάγκες για καλύτερες τεχνολογίες αναζήτησης



**"Web Science: An Interdisciplinary Approach to Understanding the Web."** James Hendler, Nigel Shadbolt, Wendy Hall, Tim Berners-Lee, and Daniel Weitzner. *Communications of the ACM*. Vol. 51, No. 7, July 2008. pp. 60-69.



# Ένας δύσκολος γάμος: iPhone κ. ATT



# Τι χρειαζόμαστε;



- New ways to understand how to design systems to produce the effect we want.
- Currently: design and build in the *micro*, hoping for the best
  - but how do we know if we've built in the right functionality to ensure the desired macroscale effects?
  - How do we predict other side effects and the emergent properties of the macro?
- Further, as the success or failure of a particular Web technology may involve aspects of social interaction among users, *understanding the Web requires more than a simple analysis of technological issues but also of the social dynamic of perhaps millions of users.*
- **"Web Science: An Interdisciplinary Approach to Understanding the Web."** James Hendler, Nigel Shadbolt, Wendy Hall, Tim Berners-Lee, and Daniel Weitzner. *Communications of the ACM*. Vol. 51, No. 7, **July 2008**. pp. 60-69.

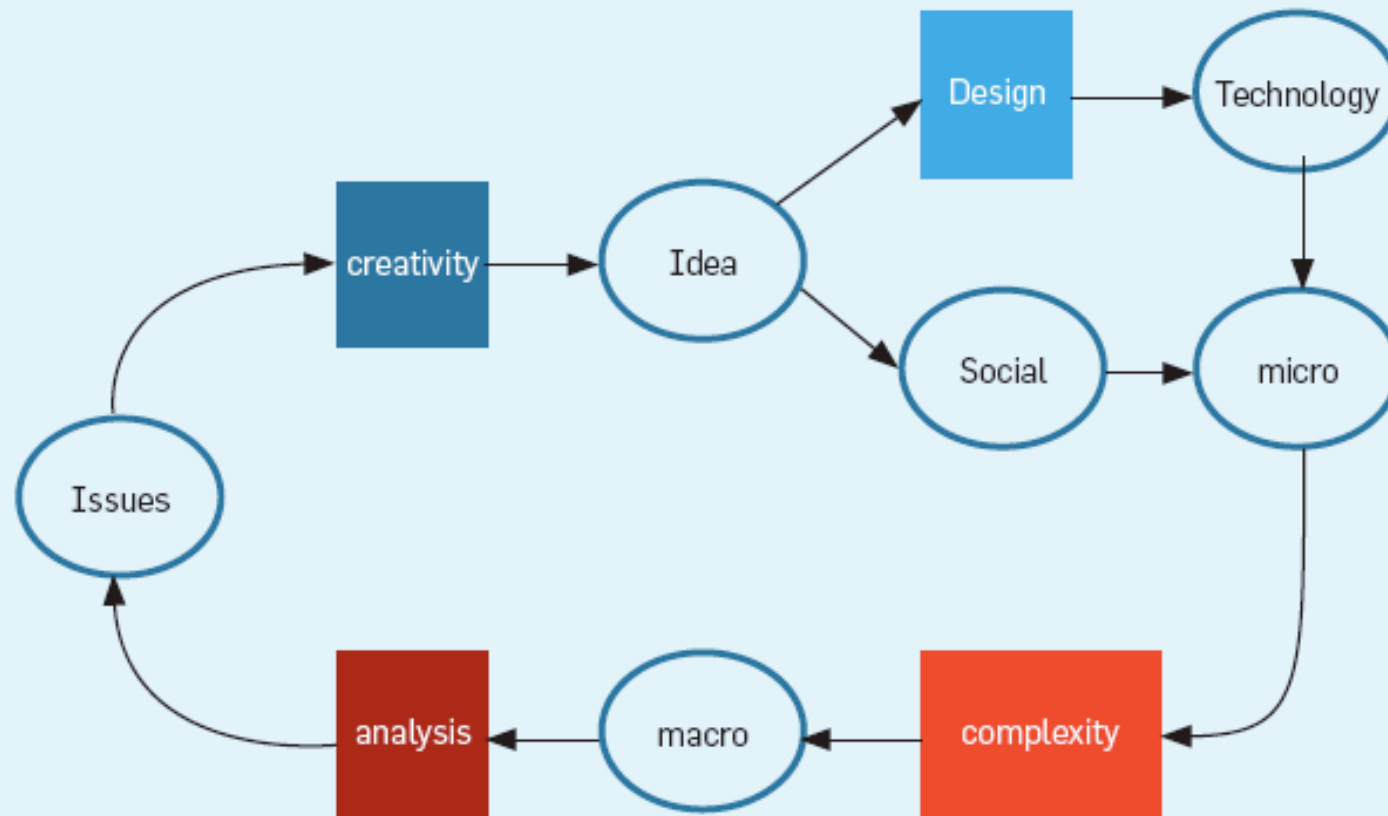
—

# Web Science



- Interdisciplinary field, “taking the Web as its primary object of study.
- Takes into account the significant interplay between:
  - social interactions enabled by the Web’s design
  - scalable and open applications development mandated to support Web-mediated social interactions
  - architectural and data requirements of large-scale applications

**Figure 2: The Web presents new challenges to software engineering and application development.**



**"Web Science: An Interdisciplinary Approach to Understanding the Web."** James Hendler, Nigel Shadbolt, Wendy Hall, Tim Berners-Lee, and Daniel Weitzner. *Communications of the ACM*. Vol. 51, No. 7, July 2008. pp. 60-69.



# Παγκόσμιος Ιστός και Διαδίκτυο



- Διαφορετικές όψεις/αφαιρέσεις-abstractions του Παγκόσμιου ΠΠ:
  - Διαδικτυακή
  - Αποθήκη Δεδομένων
  - Κατανεμημένο Υπολογιστικό Σύστημα
  - Γράφος
  - Πρόγραμμα
  - Υποδομή επικοινωνίας, συνεργασίας, υπηρεσιών

# Το ΠΠΠ σαν γράφημα - the Web graph



- Τα γραφήματα (γράφοι/graphs) είναι ένα βασικό εργαλείο/αφαιρεσία για την μοντελοποίηση προβλημάτων Πληροφορικής.
- The Web Graph (Ιστογράφημα)
  - Πώς προκύπτει;
  - Πώς αναλύεται;
    - Ποιά είναι η κατανομή των εισερχομένων ακμών (in-degree);: power law
    - Ποιά είναι η κατανομή των εξερχομένων ακμών (out-degree);: power law
    - Αλλάζουν τα χαρακτηριστικά της δομής του ιστογραφήματος, καθώς αυτό μεγαλώνει; όχι
  - Πώς αξιοποιείται η ανάλυση;
    - HITS algorithm, Pagerank algorithm
- Είναι ικανοποιητικός ο μοντελισμός μέσω του ιστογραφήματος;

# Οι «κοινωνικές μηχανές»



- Μια κοινωνική μηχανή (social machine) αποτελείται από ένα τεχνολογικό υπόστρωμα και από κανόνες, πολιτικές και οργανωτικές δομές που χρησιμοποιούνται από τους χρήστες κατά την αξιοποίηση του τεχνολογικού υποστρώματος.
- Παραδείγματα κοινωνικών μηχανών:
  - Wikipedia
    - MediaWiki software
    - Πολιτικές και κανόνες της Wikipedia
  - Blogosphere
    - Ιστιακές εφαρμογές για ιστολόγια
    - Κοινωνικοί μηχανισμοί πίσω από την αξιοποίηση blogrolls, permalinks, trackbacks
  - Online Social Networking Sites
    - Ιστιακές εφαρμογές με την διαχείριση κοινωνικών σχέσεων σαν χαρακτηριστικό πρώτης τάξης (level-1 concern)
    - Οι πολιτικές και οι μηχανισμοί της κοινωνικής δικτύωσης

# Οι «κοινωνικές μηχανές»



- Οι διαδραστικές διαδικτυακές εφαρμογές του σήμερα αποτελούν πολύ πρώιμες μορφές κοινωνικών μηχανών, περιορισμένων και απομονωμένων η μία από την άλλη κατά τη λειτουργία τους.
- Πώς θα δημιουργηθούν οι «κοινωνικές μηχανές» του μέλλοντος;
  - Θεμελιώδεις θεωρητικές ιδιότητες των μηχανών και αλγόριθμοι που χρειάζονται για την κατασκευή τους
  - Αρχιτεκτονικές αρχές για τον σχεδιασμό των μηχανών
  - Ενσωμάτωσή τους με την υπάρχουσα υποδομή του ΠΠΠ
  - Θέματα ασφάλειας, εμπιστοσύνης κλπ.



