ΕΠΛ 602:Foundation of Web Technologies

Introduction

Introduction

Goal of the course:

Study the principles underlying distributed computing by focusing on a large such system: the Web

Instructor:

Evaggelia Pitoura

www - http:www.cs.uoi.gr/~pitoura

email - pitoura@cs.uoi.gr

Local coordinates to be determined

Distributed System

A distributed system is defined as one in which *hardware* or *software components* at **networked computers** communicate and coordinate their actions only by passing messages.

- Concurrency
- No global clock
- Independent Failures

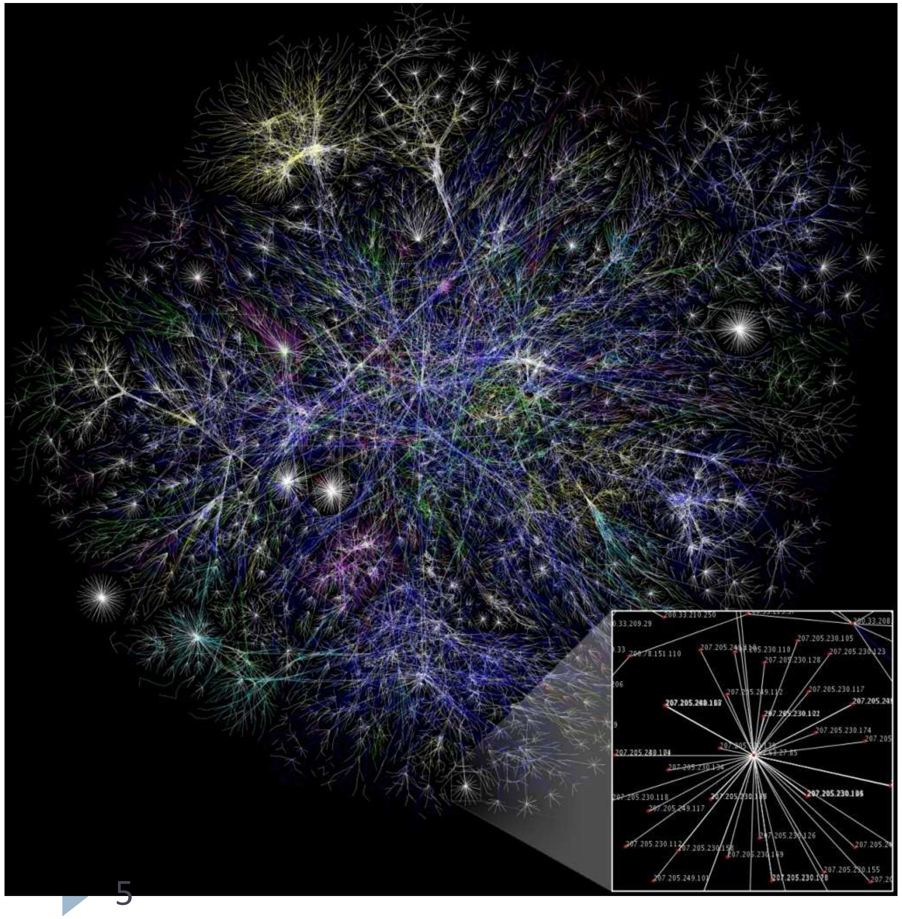
Goal: resource sharing

Distributed System: Examples

Finance and commerce	eCommerce e.g. Amazon and eBay, PayPal, online banking and trading
The information society	Web information and search engines, ebooks, Wikipedia; social networking: Facebook and MySpace.
Creative industries and entertainment	online gaming, music and film in the home, user-generated content, e.g. YouTube, Flickr
Healthcare	health informatics, on online patient records, monitoring patients
Education	e-learning, virtual learning environments; distance learning
Transport and logistics	GPS in route finding systems, map services: Google Maps, Google Earth
Science	The Grid as an enabling technology for collaboration between scientists
Environmental management	sensor technology to monitor earthquakes, floods or tsunamis

Internet

The Internet **is** a global system of interconnected computer networks that use the standard Internet protocol suite (TCP/IP) to serve billions of users worldwide

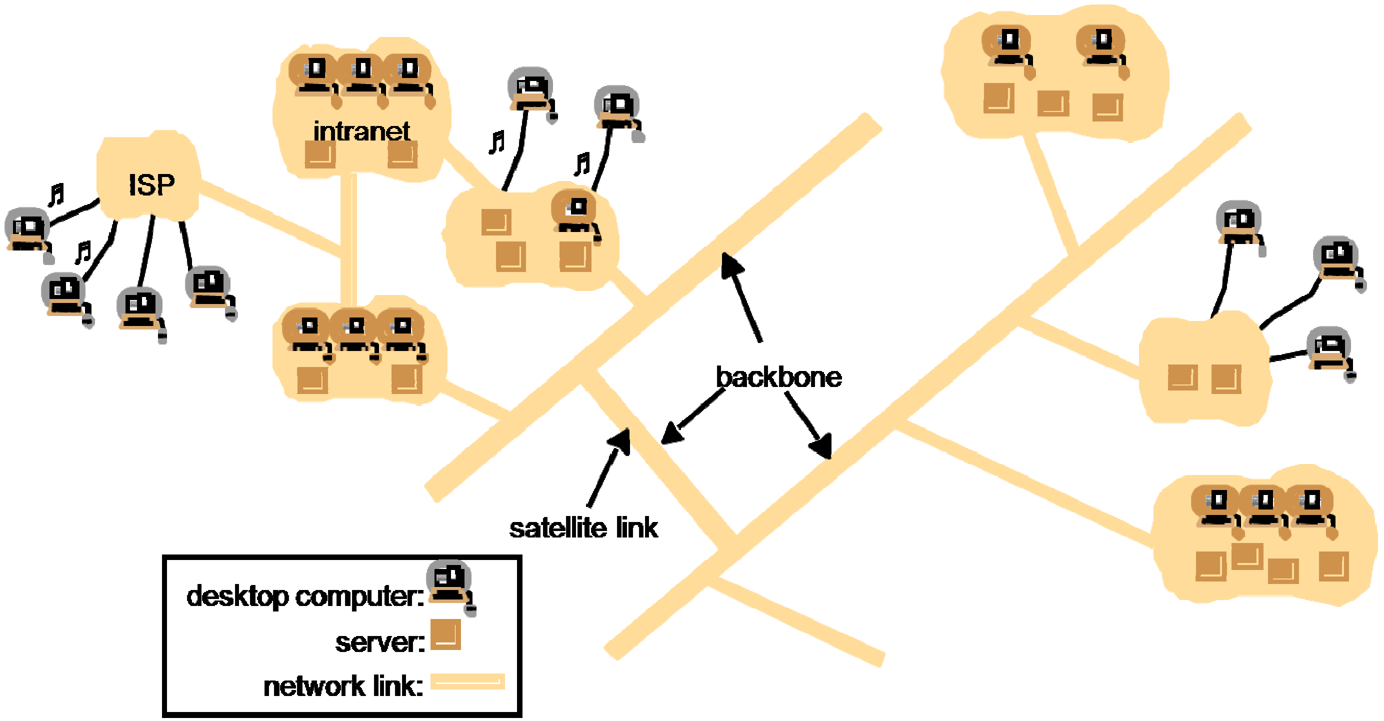


The Internet **carries** an extensive range of information resources and services, such as the inter-linked hypertext documents of the World Wide Web (WWW) and the infrastructure to support email and file transfer

Source wikipedia

- *Suite of protocols*
- *Open-Ended Services*

Internet



The Internet is a very large distributed system that allows users throughout the world to make use of its services.

An **intranet** is a part of the Internet that is separately administered and uses a firewall to enforce its own local security policies. Resources stored on files.

Pervasive and Ubiquitous: Many different types including for example, WiFi, Bluetooth and 3rd generation mobile phones and networks

The Web (WWW)

The **World Wide Web** (or the proper World-Wide Web; abbreviated as **WWW** or **W3**, and commonly known as **the Web**) **is** a collection of textual documents and other resources (**web pages**), linked by hyperlinks and URLs, hosted by **web servers** and viewed or navigated via hyperlinks with **web browsers**.

This system of interlinked hypertext documents **accessed via the Internet** – i.e., an application/service running on the Internet.

- 63 billion pages
- 1 trillion unique web addresses

Source wikipedia

The Web's historic logo designed by Robert Cailliau

The Web (WWW)

Google™

- An underlying physical infrastructure of a very large numbers of networked computers located at data centers around the world
- A distributed file system
- A structured distributed storage
- A lock service (+replication, caching)
- A programming model that allows the management of very large parallel and distributed computations

Outline

1. Short history and functionality of web
2. Short history and functionality of web.2

3. Discussion on issues in web systems

4. Course content and logistics

Web (WWW): History

June 1970 issue of *Popular Science* magazine

Arthur C. Clarke was reported to have predicted that *satellites would one day "bring the accumulated knowledge of the world to your fingertips" using a console that would combine the functionality of the Xerox, telephone, television and a small computer, allowing data transfer and video conferencing around the globe*.

Web (WWW): History

In **1980**, **Berners-Lee** wrote a proposal that referenced ENQUIRE, a database and software project he had built in 1980

November 1990, with *Robert Cailliau*, a more formal proposal to build a "Hypertext project" called "WorldWideWeb" (one word, also "W3") as *a "web" of "hypertext documents" to be viewed by "browsers" using a client–server architecture.*

Estimated that *a read-only web* would be developed *within 3 months* and that it would take *6 months* to achieve "the creation of *new links and new material by readers*, [so that] "*authorship becomes universal*" as well as "the *automatic notification* of a reader when new material of interest to him/her has become available.“

By Christmas 1990, *all tools* for a working Web:

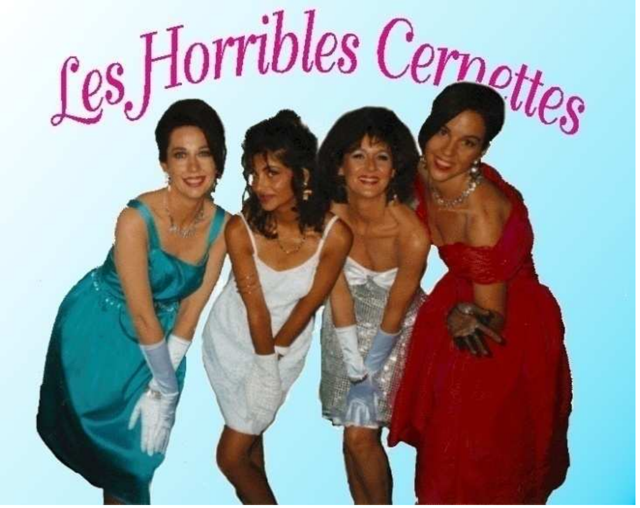
- the first web browser (which was a web editor as well);
- the first web server and
- the first web pages, which described the project itself.

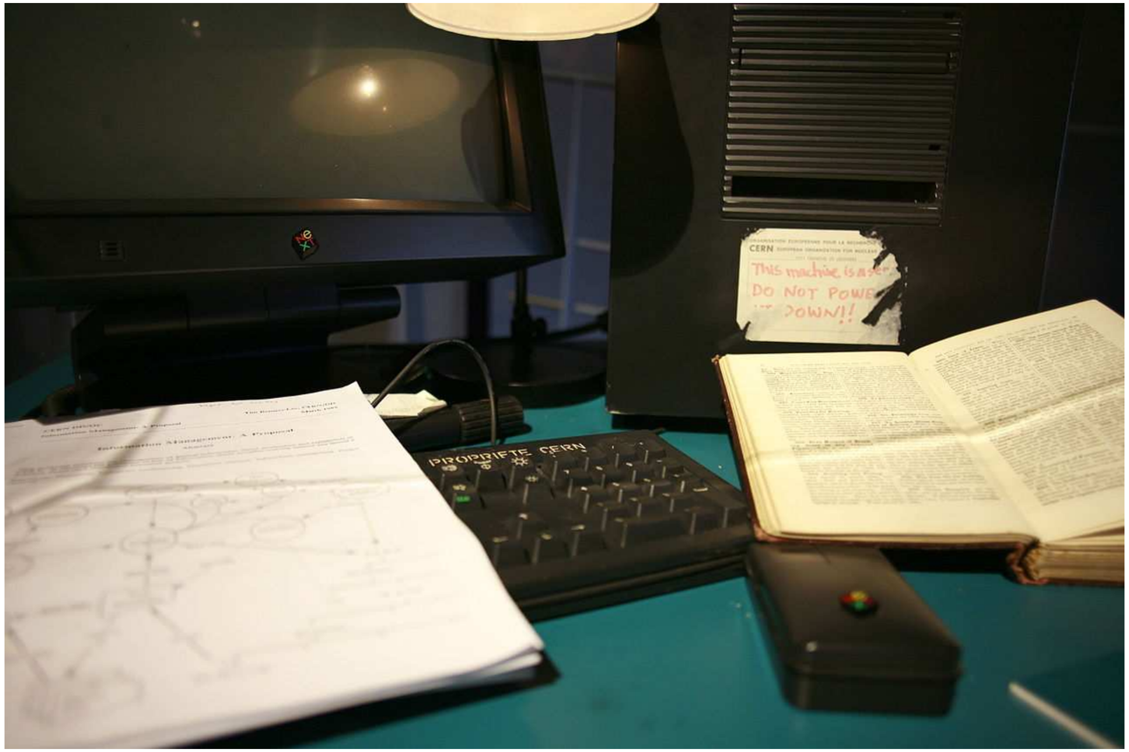
August 6, 1991, post on alt.hypertext newsgroup -> the debut of the Web as a publicly available service on the Internet.

Web (WWW): History

A NeXT Computer - the *world's first web server* and also the first web browser, WorldWideWeb

The *first photo* on the web in 1992, an image of the CERN house band Les Horribles Cernettes.





Web (WWW): History, why in CERN?

Web as a "Side Effect" of the 40 years of Particle Physics Experiments.

After the World War 2. the nuclear centers of almost all developed countries became the places with the highest concentration of talented scientists.

For about four decades many of them were invited to the international CERN's Laboratories.

Web (WWW): History

Berners-Lee's breakthrough: marry hypertext to the Internet

3 essential technologies:

1. a system of globally unique identifiers for resources on the Web and elsewhere, the Universal Document Identifier (UDI), later known as Uniform Resource Locator (**URL**) and Uniform Resource Identifier (URI);
2. the publishing language HyperText Markup Language (**HTML**);
3. the Hypertext Transfer Protocol (**HTTP**)

Web (WWW): History

Differences from other hypertext systems

- ❖ required only *unidirectional links* rather than bidirectional ones.

  (+) possible for someone to link to another resource without action by the owner of that resource

  (+) reduced the difficulty of implementing web servers and browsers (in comparison to earlier systems)

  (-) presented the chronic problem of *link rot (or dead links)*.

- ❖ was *non-proprietary* (unlike, e.g., HyperCard)

  making it possible to develop servers and clients independently and to add extensions without licensing restrictions.

  On April 30, 1993, CERN announced that the World Wide Web would be free to anyone, with no fees due. Coming two months after the announcement that the server implementation of the Gopher protocol was no longer free to use, this produced a rapid shift away from Gopher and towards the Web.

Web (WWW): History

early popular web browser was ViolaWWW for Unix and the X Windowing System.

In **1993**, **Mosaic** web browser, a *graphical browser* developed by a team at the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign (NCSA-UIUC), led by Marc Andreessen.

Prior to the release of Mosaic, graphics were not commonly mixed with text in web pages

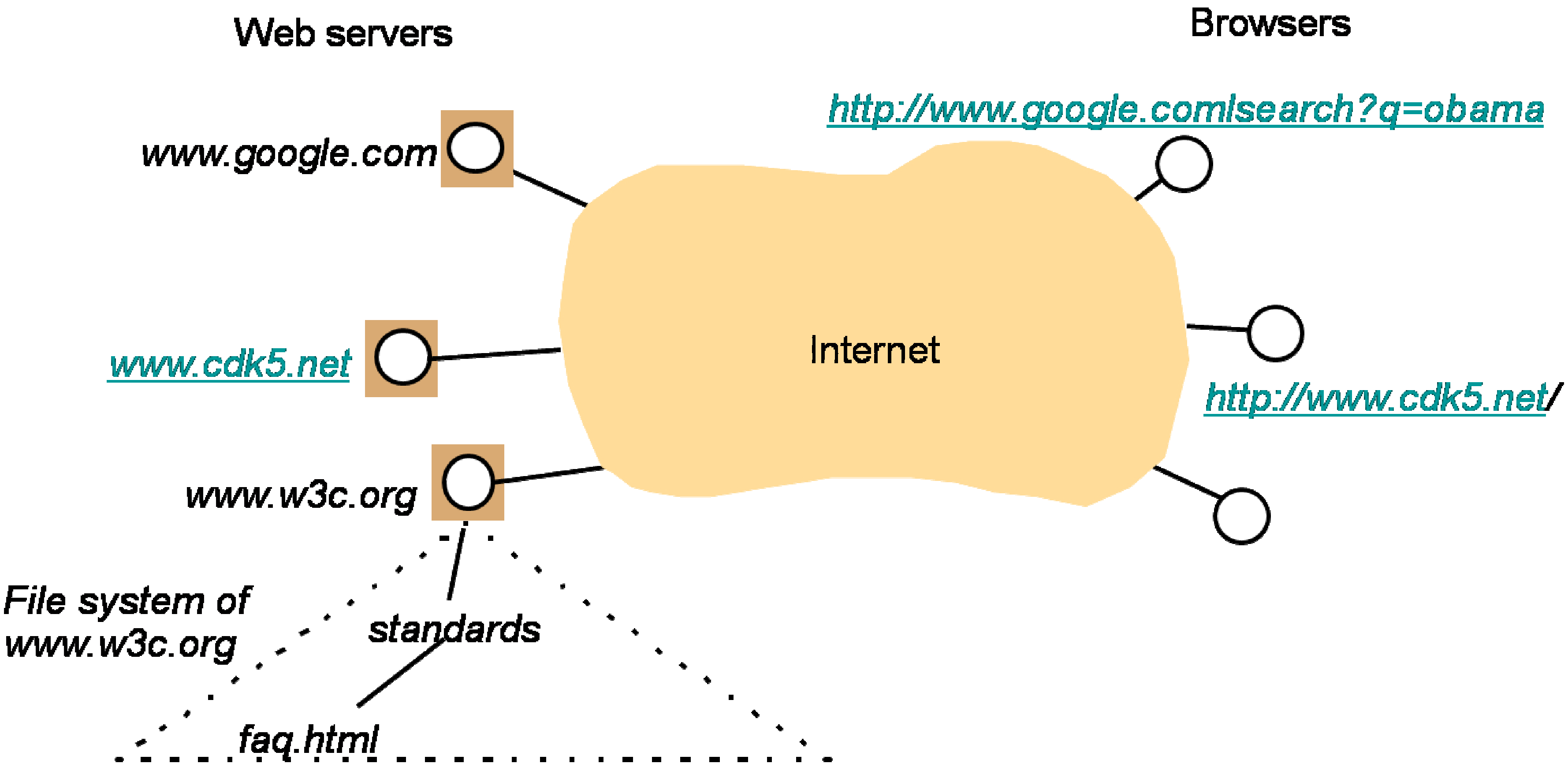
Web (WWW): History

The **World Wide Web Consortium** (W3C) was founded by Tim Berners-Lee after he left (CERN) in October 1994.

World Wide Web and *Internet*

The Web is a collection of documents and both client and server software using Internet protocols such as TCP/IP and HTTP.

Web (WWW): Function



Web (WWW): Function

Viewing a web page

- either by *typing the URL* of the page into a web browser or
- by *following a hyperlink* to that page or resource.

The web browser then initiates a series of communication messages, to **fetch** and **display** it.

Web (WWW): Function

URL http://en.wikipedia.org/wiki/World_Wide_Web .

1.The browser resolves the server-name portion of the URL (*en.wikipedia.org*) into an Internet Protocol address using the globally distributed database known as the Domain Name System (DNS)
This lookup returns an IP address such as *208.80.152.2*.

URL – (DNS) -> IP address

2.The browser then requests the resource by sending an HTTP request across the Internet to the computer at that particular address.
It makes the request to a particular application port in the underlying Internet Protocol; normally port 80. The content of the HTTP request can be as simple as the two lines of text

GET */wiki/World_Wide_Web* HTTP/1.1 Host: *en.wikipedia.org*

Web (WWW): Function

1. The computer receiving the HTTP request delivers it to Web server software listening for requests on port 80.

2. If the web server can fulfill the request it sends an HTTP response back to the browser indicating success, which can be as simple as

```
HTTP/1.0 200 OK Content-Type: text/html; charset=UTF-8
```

followed by the content of the requested page.

Web (WWW): Function

The Hypertext Markup Language for a basic web page looks like

```
<html>
<head>
<title> World Wide Web — Wikipedia, the free encyclopedia </title>
</head>
<body> <p>
The World Wide Web, abbreviated as WWW and commonly known …</p> </body>
</html>
```

The web browser parses the HTML, interpreting the markup (<title>, <p> for paragraph, and such) to draw that text on the screen.

Web (WWW): Function

- Many web pages consist of *more elaborate HTML* which references the URLs of other resources such as images, other embedded media, scripts that affect page behavior, and Cascading Style Sheets that affect page layout.

- A browser that handles complex HTML will make additional HTTP requests to the web server for these other Internet media types.

- As it receives their content from the web server, the browser progressively renders the page onto the screen as specified by its HTML and these additional resources.

Web (WWW): Linking

Most web pages contain hyperlinks to other related pages and perhaps to downloadable files, source documents, definitions and other web resources

In the underlying HTML, a hyperlink looks like

```
<a href="http://www.w3.org/History/19921103hypertext/hypertext/WWW/">Early
archive of the first Web site</a>
```

The hyperlink structure of the WWW is described by the webgraph

Web (WWW): Dynamic Pages

Uses often interact with services

For example: fill a web form

URL: not a file but a program on the server
Input part of the GET, e.g., http//www.google.com/search?q=obama

The server **process** the user input and return the produced HTML as output

CGI programs (programs that web servers run to generate content for their clients)

Service-related code run inside the browser, usually written in Javascript or applets (downloaded with a web page containing a form)

Web (WWW): Web services

Besides web pages – Web resources provide service-specific operations

Suite of protocols

Web (WWW): Caching

Almost all web browsers cache recently obtained data, usually on the local hard drive.

HTTP requests sent by a browser will usually ask only for data that has changed since the last download. If the locally cached data are still current, it will be reused.

Caching helps reduce the amount of Web traffic on the Internet.

The decision about expiration is made independently for each downloaded file

Other components of the Internet can cache Web content.

- Corporate and academic firewalls often cache Web resources requested by one user for the benefit of all.
- Some search engines also store cached content from websites.

Web2.0

A **Web 2.0 site** allows users to interact and collaborate with each other in a social media dialogue as creators **(prosumers)** of user-generated content in a virtual community, in contrast to websites where users (consumers) are limited to the passive viewing of content that was created for them.

Examples of Web 2.0 include social networking sites, blogs, wikis, video sharing sites, hosted services,web applications, mashups and folksonomies.

Web.2: History

The term "Web 2.0" was first used in **January 1999** by **Darcy DiNucci**, a consultant on electronic information design (information architecture). In her article, "Fragmented Future", DiNucci writes:

The Web we know now, which loads into a browser window in essentially static screenfuls, is only an embryo of the Web to come. The first glimmerings of Web 2.0 are beginning to appear, and we are just starting to see how that embryo might develop.
The Web will be understood not as screenfuls of text and graphics but as a transport mechanism, the ether through which interactivity happens. It will [...] appear on your computer screen, [...] on your TV set [...] your car dashboard [...] your cell phone [...] hand-held game machines [...] maybe even your microwave oven.

Web.2: History

In 2003, rise in popularity when **O'Reilly Media** and MediaLive hosted the first Web 2.0 conference.

In their opening remarks, John Battelle and Tim O'Reilly outlined their definition of the "Web as Platform", where software applications are built upon the Web as opposed to upon the desktop.

Web.2: History

Netscape vs Google

Netscape focused on creating software, updating it on occasion, and distributing it to the end users.

Google, a company that at the time did not focus on producing software but instead on providing a service based on data such as the links Web page authors make between sites. Google exploits this user-generated content to offer Web search based on reputation through its "PageRank" algorithm. Unlike software, which undergoes scheduled releases, such services are constantly updated, a process called *"the perpetual beta".*

Encyclopædia Britannica Online and Wikipedia:

Britannica relies upon experts to create articles and releases them periodically in publications, Wikipedia relies on trust in anonymous users to constantly and quickly build content. Wikipedia is not based on expertise but rather an adaptation of the open source software adage "given enough eyeballs, all bugs are shallow", and it produces and updates articles constantly.

Web.2: History

In the **2006** , TIME magazine Person of The Year (You).

TIME selected the masses of users who were participating in content creation on social networks, blogs, wikis, and media sharing sites.
In the cover story, Lev Grossman:
It's a story about community and collaboration on a scale never seen before. It's about the cosmic compendium of knowledge Wikipedia and the million-channel people's network YouTube and the online metropolis MySpace. It's about the many wresting power from the few and helping one another for nothing and how that will not only change the world but also change the way the world changes.

In **2009,** Global Language Monitor declare Web2.0 to be the one-millionth English word

Web.2: Characteristics

Web 2.0 websites allow users to do more than just retrieve information.
They provide the user with more user-interface, software and storage facilities, all through their browser.

This has been called "Network as platform"computing.

The concept of Web-as-participation-platform

Collective intelligence, the wisdom of the crowd, trust, spam, dynamic content, user-generated content, etc

Outline

1. Short history and functionality of web
2. Short history and functionality of web.2

3. Discussion on issues in web systems

4. Course content and logistics

Web: Areas

A large distributed system:

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

Algorithms and Data Structures: the web (or the social network) graph

(new area of) Social computing

The Web: note

- ❖ Design
- ❖Specify
- ❖ Build
- ❖ Test

- ❖Physical Science

The course: Content

1. Basics of Distributed Computing

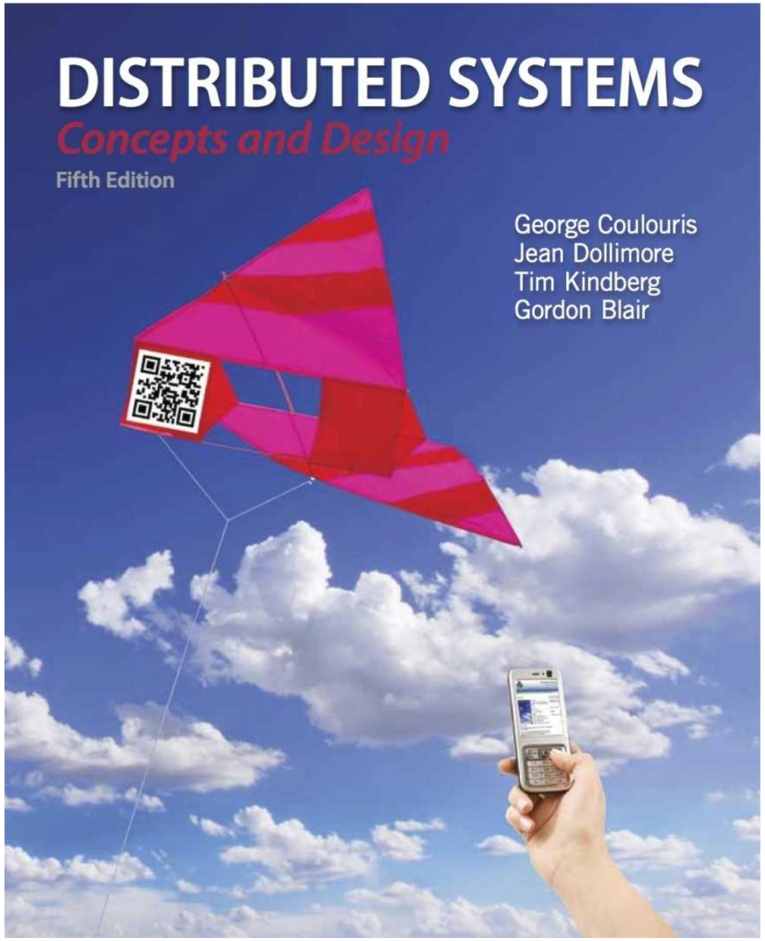
System Models

Networking and Internetworking

Interprocess Communication
Remote Invocation
Indirect communication

Operating System Support
Distributed objects and components

Security
Name Services
Time and Global States
Coordination and Agreement
Transactions and Concurrency Control
Distributed Transactions
Replication



The course: Content

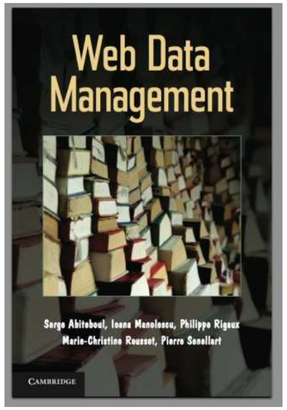
II. Web Basics

Languages
Ajax/Java Script

Data Models
XML
RDF
LinkedData

Web Services/Mashups

Web Search (Google as a Case Study)

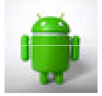

Web Data Management
Serge Abiteboul, Ioana Manolescu, Philippe Rigaux
Marie-Christine Rousset, Pierre Senellart
CAMBRIDGE

Free online at:
http://webdam.inria.fr/Jorge/#

The course: Content

III. Current Trends

Cloud/MapReduce/Hadoop



Mobile Computing

Android programming

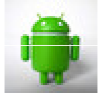
Distributed Overlays

Social Networks

The course: Content

III. Current Trends

Cloud/MapReduce/Hadoop



Mobile Computing

Android programming

Distributed Overlays

Social Networks

The course: Requirements

- Course participation and presentations: 20%
- Term project: 40%
- Midterm (part I)/Final Exam (part II): 20% each

To be refined in the next lecture

Conclusion

Current trends in distributed computing

❖ Distributing Computing as Utility (cloud)
❖ Social Networking
❖ Ubiquitous Computing

Questions? Remarks?