FACOLTÀ DI INGEGNERIA
UNIVERSITÀ DEGLI
STUDI DI PALERMO

Sistemi ICT per il Business Networking

The Web

*Docente: **Vito Morreale** (vito.morreale@eng.it)*

The Web

- The World Wide Web ("WWW" or simply the "Web") is a **global information space** which people can read and write via computers connected to the Internet

- **Not a synonym of the Internet**: the Web is actually a service that operates over the Internet

The Web
Internet

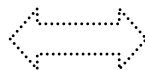**ICT Systems for Business Networking**

The Web

- The World Wide Web is the combination of **four basic ideas**:

  - **Hypertext**: moving from one part of a document to another or from one document to another through internal connections among these documents (called "**hyperlinks**")

  - **Resource identifiers**: locating a particular resource (computer, document or other resource) on the network through a **unique identifier**

  - **Client-server model of computing**: client software or a client computer makes **requests** of server software or a server computer that provides the client with **resources or services** (e.g data or files)

  - **Markup language**: characters or codes embedded in text indicate to a computer how to print or display the text (e.g. in italics or bold type or font)

Web browser and server

- On the Web, a **web browser** (client) retrieves information resources (e.g. web pages and other files) from **web servers** using their network addresses and displays them, typically on a computer monitor, using a markup language that determines the details of the display

  - One can then follow **hyperlinks** in each page to other resources on the Web of information whose location is provided by these hyperlinks

  - The act of following hyperlinks is often called "**browsing**" the Web

  - Web pages are arranged in collections of related material ("**websites**")

Web Browser   Web Server

Main Web standards

- At its core, the Web is made up of three **standards**:
    - the **Uniform Resource Identifier (URI)**: a universal system for referencing resources on the Web (e.g. Web pages, files, etc.)
    - the **HyperText Transfer Protocol (HTTP)**, which specifies how the browser and server communicate with each other
    - the **HyperText Markup Language (HTML)**, used to define the structure and content of hypertext documents
- Tim Berners-Lee now heads the **World Wide Web Consortium (W3C)**, which develops and maintains Web standards that enable computers on the Web to effectively store and communicate different forms of information

How the Web works

1. A user wants to access a web page or other "resource" on the Web. He or she normally **types the URL** of the page into the web browser, or by **following a hypertext link** to that page or resource
2. **Resolution of the URL into an IP address** by the global, distributed Internet database known as the **Domain name system (DNS)**
3. An **HTTP request** is sent to the web server working at that IP address for the page required
4. In the case of a typical **web page**, the HTML text, graphics and any other files that form a part of the page will be **returned** to the client (the web browser) in quick succession
5. The web browser then **renders** the page as described by the HTML, CSS and other files received, incorporating the images, links and other resources as necessary and **produces the on-screen 'page'** that the user sees

Publishing web pages

- The increased opportunity to publish materials on the Web has been facilitating by the emergence of (free and non-free) **web hosting services**
  - which provide individuals, organizations and users with online systems for storing information, images, video, or any content accessible via the Web
- Web hosting companies **provide**
  - **space** on a server they own for use by their clients

Uniform Resource Identifier (URI)

- An Internet protocol element consisting of a short **string** of characters **conform to a certain syntax** that can be used to refer to a resource

4

Uniform Resource Identifier (URI)

- **Uniform Resource Locator (URL)**: a URI that, in addition to identifying a resource, describes its primary access mechanism or network "location"
    - Example: the URL **http://www.wikipedia.org/**
        - identifies a **resource** (Wikipedia's home page)
        - implies that a representation of that resource (such as the home page's current HTML code, as encoded characters) is obtainable via **HTTP** from a network host named www.wikipedia.org.

- **Uniform Resource Name (URN)**: a URI that identifies a resource by name in a particular namespace to talk about a resource **without implying its location**
    - Example: the URN **urn:ISBN:0-395-36341-1** is a URI that allows one to talk about a book, but doesn't suggest where and how to obtain an actual copy of it

---

Uniform Resource Identifier (URI)

- **Syntax**:
    - a URI **scheme name** like "http", "ftp", "mailto", "urn", etc.,
    - a **colon** character ":"
    - a **scheme-specific part**

- **URI reference**: another type of string that represents a URI and the resource identified by that URI
    - An **optional fragment identifier**, preceded by "**#**", may be present at the end of a URI reference
    - The part of the reference before the "#" indirectly identifies a resource, and the fragment identifier identifies some **portion of that resource**
        - In **web document markup languages**, URI references are frequently used in places where there is a need to point specific portions of the same logical document

URI Examples

- Examples of absolute **URIs**
  - http://www.somehost.org/absolute/URI/with/absolute/path/to/resource.txt
  - ftp://www.somehost.com/resource.txt
  - urn:issn:1535-3613
- Examples of **URI references**
  - http://www.example.it/resource.txt#frag01
  - http://www.somehost.edu/absolute/URI/with/absolute/path/to/resource.txt
  - /relative/URI/with/absolute/path/to/resource.txt
  - relative/path/to/resource.txt
  - ../../../resource.txt
  - resource.txt
  - /resource.txt#frag01
  - #frag01

HyperText Transfer Protocol (HTTP)

- The method used **to transfer or convey** information on the Web

- A request/response protocol between clients and servers
  - A **HTTP client** initiates a **request** by establishing a Transmission Control Protocol (TCP) connection to a particular **port** on a remote host (port 80 by default)
  - A **HTTP server** listening on that port waits for the client to send the request message
  - Upon receiving the request, the server **sends back** a status line, such as "HTTP/1.1 200 OK", and a message of its own, the body of which is perhaps the requested file, an error message, or some other information

HyperText Markup Language (HTML)

- A **markup language** designed for the **creation of web pages** with hypertext and other information to be displayed in a web browser
  - **to structure information** — denoting certain text as headings, paragraphs, lists and so on — and can be used to describe, to some degree, the appearance and semantics of a document

Web servers

- The term can mean **one of two things**:
  - A **computer** that is responsible for accepting HTTP requests from Web browsers and serving them Web pages
  - A computer **program** that provides this functionality
- Although Web server programs differ in detail, they all share some basic **common features**
  - Every Web server program operates by **accepting HTTP requests** from the network, and **providing an HTTP response**
  - The HTTP response typically consists of an **HTML document**, but also a **raw text file**, an **image**, or some **other type of document**
  - Usually Web Servers have also the capability of **logging** some detailed information, about client requests and server responses, to log files to collect **statistics** by running log analyzers on log files

Web servers

- The origin of the content sent by server is
  - **static** if it comes from an existing file
  - **dynamic** if it is dynamically generated by some other program or script
- Some common Web or HTTP server programs are:
  - **Apache HTTP Server** from the Apache Software Foundation
  - **Internet Information Services (IIS)** from Microsoft
  - Sun Java System Web Server from Sun Microsystems, formerly **Sun ONE** Web Server, **iPlanet** Web Server, and **Netscape Enterprise Server**
  - **Zeus Web Server** from Zeus Technology
  - …

Web sites

- A **collection of Web pages**, typically common to a particular domain name or sub-domain on the World Wide Web on the Internet
- A **web page** is an HTML/XHTML document accessible generally via HTTP
- The pages of a website will be accessed from a **common root URL** called the homepage, and usually reside on the **same physical server**
- Pages are organised into a **hierarchy**, although the hyperlinks between them control how the reader perceives the overall structure

Web sites

- Any website can contain a **hyperlink to any other website**, so the distinction between individual sites, as perceived by the user, may sometimes be blurred

- Web pages can be viewed or otherwise accessed from a range of **Internet enabled devices** of <u>various sizes</u> (e.g. desktop computers, laptop computers, PDAs and cell phones)

Static Web sites

- One that has content that is **not expected to change frequently** and is **manually maintained** by some person or persons using some type of **editor software**

- There are two broad categories of editor software used for this purpose which are

  - **Text editors** such as Notepad, where the HTML is manipulated directly within the editor program

  - **WYSIWYG editors** such as Microsoft FrontPage and Macromedia Dreamweaver, where the site is edited using a GUI interface and the underlying HTML is generated automatically by the editor software

Dynamic Web sites

- One that may have **frequently changing information**

- When the web server receives a request for a given page, the page is **automatically generated** by the software in direct response to the page request
  - Active Server Pages (ASP)
  - Java Server Pages (JSP)
  - PHP programming language

- Static content may also be **dynamically generated periodically** or if certain conditions for regeneration occur (cached) to avoid the performance loss of initiating the dynamic engine on a per-user or per-connection basis

---

Dynamic Web sites

- Plugins for browsers used to show **active content**
  - Flash, applets written in Java, etc.

- Dynamic HTML also provides for user interactivity and real-time element updating **within Web pages** (i.e., pages don't have to be loaded or reloaded to effect any changes), mainly using the JavaScript
  - **Most modern browsers** support JavaScript

Types of Web sites

- **Blog (or weblog) site**: to log online readings or to post online diaries

- **Business site**: used for promoting a business or service

- **eCommerce site**: for purchasing goods, such as Amazon.com

- **Community site**: a site where persons with similar interests communicate with each other, usually by chat or message boards

- **Database site**: a site whose main use is the search and display of a specific database's content such as the Internet Movie Database

- **Development site**: a site whose purpose is to provide information and resources related to software development

- **Directory site**: a site that contains varied contents which are divided into categories and subcategories, such as Yahoo! Directory and Google

- **Download site**: strictly used for downloading electronic content, such as software, game demos or computer wallpaper

Types of Web sites

- **Game site**: a site that is itself a game or "playground" where many people come to play, such as MSN Games

- **Information site**: contains content that is intended merely to inform visitors, but not necessarily for commercial purposes (e.g. Encyclopedia)

- **News site**: similar to an information site, but dedicated to dispensing news and commentary

- **Search engine site**: a site that provides general information and is intended as a gateway or lookup for other sites

- **Personal site**: run by an individual or a small group (such as a family) that contains information that the individual wishes to include

- **Web portal site**: a website that provides a starting point, a gateway, or portal, to other resources on the Internet or an intranet

- **Wiki site**: a site which users collaboratively edit (such as Wikipedia)

- **Political site**: A website on which people may voice political views

Types of Web sites

- Some sites may be included in **one or more of these categories**
  - For example, a business website may promote the business's products, but may also host informative documents, such as white papers
- There are also numerous **sub-categories**

Web portals

- Some **features** of enterprise portals are:
  - **Single touch point** — the portal becomes the delivery mechanism for all business information services
  - **Collaboration** — portal members can communicate synchronously or asynchronously through threaded discussion and email digests (forums) and blogs
  - **Personalization** — the ability for portal members to subscribe to specific types of content and services
    - Users can customize the look and feel of their environment
  - **Integration** — the connection of functions and data from multiple systems into new components/portlets
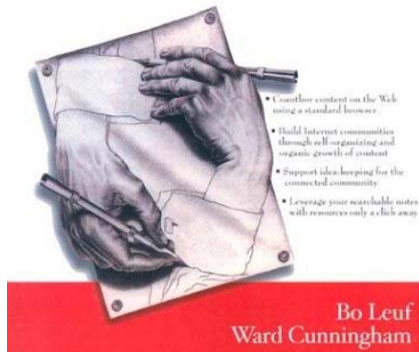
Enterprise Web portals

- Most enterprise portals provide **single sign-on** capabilities to their users

- Enterprise portals may be referred to **by the community they serve**
  - Business-to-employee (**B2E**) portal
  - business-to-customer/consumer (**B2C**) portal
  - business-to-dealer/distributor (**B2D**) portal
  - business-to-business/supplier (**B2B**) portal
  - business-to-government (**B2G**) portal
  - ...

Wiki

- A type of website that allows users to **add and edit content** easily and is especially suited for **collaborative writing**

- The name is based on the **Hawaiian term** wiki, meaning "quick", "fast", or "to hasten" (Hawaiian dictionary)

- A wiki system **records** each individual change that occurs over time, so that at any time, a page can be reverted to any of its previous states

- A wiki system may also provide various **tools** that easily allow the user community to monitor the constantly changing state of the wiki and discuss the issues that emerge in trying **to achieve a general consensus** about wiki content

- Wiki content can also be misleading as users may add incorrect information to the Wiki page

Wiki

The Wiki Way
Quick Collaboration on the Web

- Coauthor content on the Web using a standard browser
- Build Internet communities through self-organizing and organic growth of content
- Support idea-keeping for the connected community
- Leverage your searchable notes with resources only a click away

Bo Leuf
Ward Cunningham

- Some wikis will allow completely **unrestricted access** so that people are able to contribute to the site without necessarily having to undergo a process of 'registration

- A wiki enables **documents to be written collectively** in a very simple markup language using a web browser

---

Wiki

- **Ease** page creation and updating

- Generally there is **no review** before modifications are accepted

- Most wikis are **open to the general public** without the need to register any user account

- Sometimes session **log-in** is requested to acquire a "wiki-signature" cookie for autosigning edits

- More private wiki servers require **user authentication**

- However, many **edits can be made in real-time**, and appear almost instantaneously online

Wiki pages

- Three representations for each page:
  - The **user-editable "source code" (wikitext)**: the format stored locally on the server
    - usually **plain text**, made visible to the user only when the edit operation shows it in a browser form
    - augmented with a simplified markup language to indicate various structural and visual conventions
  - A **template** that defines layout and elements common to all pages
  - The **rendered HTML** code produced by the server <u>on the fly</u> from the source text when a particular page is requested

---

Wiki pages

Wiki syntax (MediaWiki)	HTML	Rendered output

`'''Doctor''? No other title? A ''scholar''? And he rates above the civil authority?"`
 	
`"Why, certainly," replied Hardin, amiably. "We're all scholars more or less. After all, we're not so much a world as a scientific foundation—under the direct control of the Emperor."`
	`<p>`
	`"<em>Doctor</em>? No other title? A <em>scholar</em>? And he rates above the civil authority?"`
	`</p>`
	`<p>`
	`"Why, certainly," replied Hardin, amiably. "We're all scholars more or less. After all, we're not so much a world as a scientific foundation—under the direct control of the Emperor."`
	`</p>`
		"Doctor? No other title? A *scholar*? And he rates above the civil authority?"
		
		"Why, certainly," replied Hardin, amiably. "We're all scholars more or less. After all, we're not so much a world as a scientific foundation—under the direct control of the Emperor."

15

Wiki pages

- Some recent wiki engines use a **"WYSIWYG" editing**
  - Users who do not have the necessary plugin can generally edit the page, usually by directly editing the raw HTML code
- **Simple wikis** allow only basic text formatting, whereas more **complex** ones have support for tables, images, formulas, or even interactive elements such as games
- Because of the difficulty in using several syntaxes, many people are putting considerable effort into defining a **wiki markup standard**

Wiki pages

- Wikis are a true hypertext medium, with **non-linear navigational structures** ➔ Each page typically contains a **large number of links** to other pages
- Most wikis offer at least a **title search**, and sometimes a **full-text search**
- The **MetaWiki search engine** was created to enable keyword-based searches across multiple wikis

Wikipedia

World Wide Web Consortium (W3C)

- International consortium where member organizations, a full-time staff, and the public, work together **to develop standards for the World Wide Web**

- W3C's **mission** is: "To lead the World Wide Web to its **full potential** by developing **protocols** and **guidelines** that ensure **long-term growth for the Web**"

- W3C also engages in **education**, **develops software**, and serves as an **open forum** for discussion about the Web

- The Consortium is headed by **Tim Berners-Lee**, the original **creator** of the Web and primary **author** of the URL, HTTP, and HTML specifications, the principal technologies that form the basis of the Web

17

World Wide Web Consortium (W3C)

- **W3C Process**: Recommendation ➜ Working Draft (WD) ➜ Last Call Working Draft ➜ Candidate Recommendation (CR) ➜ Proposed Recommendation (PR) ➜ W3C Recommendation (REC)

- A Recommendation may be **updated** by separately-published Errata until enough substantial edits accumulate, at which time a **new edition** of the Recommendation may be produced (e.g., XML is now in its third edition)

- W3C also publishes various kinds of informative **Notes** which are not intended to be treated as standards

- The Recommendations are under a **royalty-free patent license**, allowing anyone to implement them

---

Internet hosting service

- A service that allows organizations and individuals to serve **content on the Internet**

- Various **levels of service** and **various kinds of services** offered

- **Full-featured hosting services** include:
    - **Dedicated hosting service**, where the hosting service **owns and manages** the machine, leasing **full control to the client**
    - **Virtual private server**, which appears as a dedicated server but only uses a single physical server
    - **Colocation facilities**, which provide just the Internet connection and climate control, but let the client do his own system administration (the most expensive)

Dedicated hosting service

- A type of Internet hosting where the client leases an **entire server not shared** with anyone

- Organizations have **full control** over the server(s), including choice of operating system, hardware, etc.

- **Server administration** can usually be provided by the hosting company as an add-on service

- Dedicated servers are most **often housed in data centers**

Application-specific hosting services

- Limited or application-specific hosting services include:
  - Web hosting service
  - E-mail hosting service
  - DNS hosting service
  - Game servers

Web hosting service

- A **web hosting service** is a type of Internet hosting service that provides individuals, organizations and users with online systems for storing information, images, video, or any content accessible via the Web

    - free web hosting

    - shared web hosting

    - clustered hosting

    - reseller hosting

    - image hosting service

    - file hosting service

    - remote backup service or online backup service

Free web hosting

- **Free**, usually **advertisement-supported** and of **limited functionality**

- Free web hosts will either provide

    - a subdomain (**yourname.hostservicename.com**)

    - a directory (**www.hostservicename.com/~yourname**)

    - a second-level domain along with the hosting (**www.yourname.com**)

- A few free web hosts have a **basic package for free**, and **enhanced packages (with more features) for a cost**

    - This allows users to try the service for an initial trial (see how it performs compared to other hosts), and then upgrade when (and if) needed

Shared web hosting

- **Shared web hosting**
  - **More than one web site** is hosted on the same server
  - The hosting service must include **system administration** since it is shared by many users that do not want to deal with it
  - The **provider** is generally responsible for **management** of servers, **installation** of server software, **security updates** and other aspects of the service
- **Clustered hosting**
  - provides customers with a "clustered" handling of security, load balancing, and necessary website resources
  - "**virtualizes**" the resources beyond the limits of one physical server, and as a result, a website is not limited to one server
  - the processing power is shared from many servers and the applications are distributed in real-time

Other hosting services

- **Reseller hosting** is a form of web hosting wherein the account owner may split up their alloted hard drive space and bandwidth and resell web hosting

- An **image hosting service** allows individuals to upload images to an Internet website

- A **file hosting service** or **online file storage service** is an Internet hosting service specifically designed to host static content, typically large files that are not web pages
  - Typically they allow web and FTP access. They can be optimized for serving many users (as is implied by the term "hosting") or be optimized for single-user storage (as is implied by the term "storage")

Other hosting services

- A **remote backup service** or **online backup service** is a service that provides users with an online system for backing up and storing computer files
    - Files are usually guaranteed not to be lost
    - The client runs a scheduled software program (typically once a day)
    - This program compresses, encrypts, and transfers to the remote backup service provider's servers
- A **game server** is a server used by game clients
    - Any video game played over the internet generally requires a connection to a game server

Other hosting services

- A **DNS hosting service** is a service that runs Domain Name System servers
- An **e-mail hosting service** is an Internet hosting service that runs e-mail servers
    - E-mail hosting services differ from typical end-user e-mail providers such as webmail sites
    - They cater to large businesses, allow custom configuration, large number of accounts, use of own domain name

References

- **The Web** (provided by the teacher)