Knowledge Update Course

for Second Knowledge Update Course and IT Teachers for Secondary School Computer and IT Teachers

Day Three

Web Page Design Web Authoring and Management

Simon Hui

Email: iccshui@polyu.edu.hk

Industrial Centre

The Hong Kong Polytechnic University 31 July 2003



Internet

Rise of the Internet is a co-evolution of

- Technologies
- Services
- Human interactions



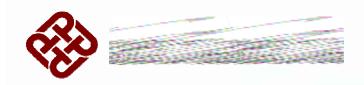


Technologies

The fundamental communications

- 1947 Transistor
- 1950 Mainframe
- 1970 Optical Fiber
- 1971 Microprocessor
- 1970s Dialup Communications
- 1981 IBM PC
- 1980s Local Area Network
- 1990s World Wide Web





Services

- 1837 Telegraph
- 1870 Telephone
- 1890s Radio broadcasting
- 1920s Teletype
- 1930s Television broadcasting
- **1969 ARPANET**



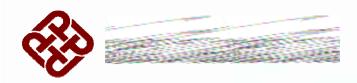


The ARPANET

•ndustrial

- The global Internet's intellectual predecessor was the Advanced Research Projects Agency Network (ARPANET) of the U.S. Department of Defense.
- The formation of this agency was part of the U.S. reaction to the then Soviet Union's launch of Sputnik in 1957.
- The ARPANET Completion report ends with the suggestion that the ARPANET is fundamentally connected to and born of computer science.
- A "Network Information Center" (NIC) for the ARPANET is established for the ARPANET Stanford Research Institute in 1969.
- The Campus Computing Network of UCLA obtained a connection to the ARPANET.

 We are



A Simple Statement of the History of Internet

■ The Internet can be traced back to the 1970s with experiment conducted by the U.S. and a few Universities in U.S.A. to connect a few super computers.





Protocol

- Protocol is the special set of rules that end points in a telecommunication connection use when they communicate.
- The word protocol is derived from Greek *protocollon*, which was a leaf of paper glued to a manuscript volume, describing its contents.



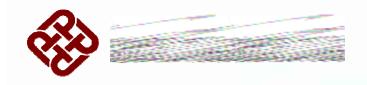


Internet Protocols

•ndustrial

- On the Internet, there are the TCP/IP protocols, consisting of:
- Transmission Control Protocol (TCP), which uses a set of rules to exchange messages with other Internet points at the information packet level
- <u>Internet Protocol</u> (IP), which uses a set of rules to send and receive messages at the Internet address level
- Additional protocols that are usually packaged with a TCP/IP suite, including the <u>Hypertext Transfer Protocol</u> (HTTP) and <u>File Transfer Protocol</u> (FTP), each with defined sets of rules to use with corresponding programs elsewhere on the Internet
- There are many other Internet protocols, such as the Border Gateway Protocol (BGP) and the Dynamic Host Configuration Protocol (DHCP).

 We are



Internet Protocol IP

- The Internet Protocol (IP) is the method by which data is sent from one computer to another on the Internet.
- Each computer (known as a <u>host</u>) on the Internet has at least one <u>IP address</u> that uniquely identifies it from all other computers on the Internet.
- When you send or receive data (for example, an email note or a Web page), the message gets divided into little chunks called packets.
 Each of these packets contains both the sender's Internet address and the receiver's address.



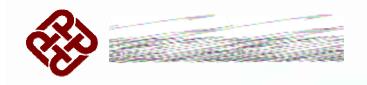


Packet

•ndustrial

- Any little chunks of data is sent first to a gateway computer.
- Gateway computer reads the destination address and forwards the packet to an adjacent gateway.
- The gateway that receives packets in turn reads the destination address and so forth across the Internet until one gateway recognizes the packet as belonging to a computer within its immediate neighborhood or domain.
- That gateway then forwards the packet directly to the computer whose address is specified.

 We are

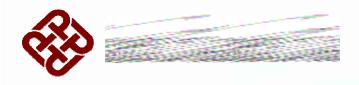


IP is Connectionless

- IP is a connectionless protocol.

 There is no continuing connection between the end points that are communicating.
- Each packet that travels through the Internet is treated as an independent unit of data without any relation to any other unit of data.
- Right Order of Packets Packets with lost order means a message is in a mess.
 Packets do get put in the right order is because of TCP.





Transmission Control Protocol

■ TCP is a connection-oriented protocol that keeps track of the packet sequence in a message.

■ TCP connection is established and maintained until the message to be exchanged by the application programs at each end have been exchanged.





TCP/IP

- TCP/IP is two layered.
- Transmission control protocol manages assembling of file into smaller packets transmitted over.
- The other end received by a TCP layer that reassembles the packets into the original file.
- The lower layer, Internet protocol handles the address part of each packet so it gets to the right destination.

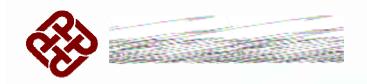




Use of TCP/IP

- For example, when an <u>HTML</u> file is sent to you from a Web server.
- Transmission Control Protocol (TCP) program layer in that server does
 - 1. divides the file into one or more packets,
 - 2. numbers the packets,
 - 3. forwards them individually to the IP program layer.
- Although each packet has the same destination IP address, it may get routed differently through the network.
- At the other end (the <u>client</u> program in your computer), TCP reassembles the individual packets and waits until they have arrived to forward them to you as a single file.





HTTP

- HTTP is the set of rules for exchanging files. HTTP is an application protocol.
- Files can contain references to other files whose selection will elicit additional transfer requests.
- Any Web server has an HTTP daemon.
- Http daemon waits for HTTP requests and handle them when they arrive.
- Web browser is an HTTP client, sending requests to server machines.
- Browser user enters file requests by either "opening" a Web file (typing in a Uniform Resource Locator) or clicking on a hypertext link)
- Browser builds an HTTP request and sends it to the <u>Internet Protocol</u> address indicated by the URL.
- The HTTP daemon in the destination server machine receives the request and, after any necessary processing, the requested file is returned are ndustrial



HTTP1.1

- The latest version of HTTP is HTTP 1.1.
- World Wide Web application protocol that runs on top of the Internet's TCP/IP suite of protocols.
- HTTP is developed by Internet Engineering Task Force (IETF).





Hypertext Protocol Daemon

- Each web server has an HTTPD.
- HTTPD waits in attendance for requests to come in from the rest of the Web.
- It forward requests to other processes as appropriate.





HTTP in a Nut Shell

∙ndustrial

■ HTTP combines <u>hypertext</u> with the speed of today's electronic networks.

Who Invent HTTP

■ Tim Berners-Lee, with a small team, of CERN, the high-energy particle physics laboratory in Geneva, Switzerland, in 1991.

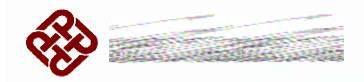
What is Hypertext

- Hypertext is the organization of information units into connected associations.
- An instance of such an association is called a <u>link</u> or hypertext link.

What is the Web

■ World Wide Web is nothing more (or less) than an enormous amount of information content connected by an enormous number of hypertext links.

We are



Internet Domain Name System

- The Internet's domain-name system (DNS) allows users to refer to web sites and other resources.
- DNS uses easier-to-remember domain names (such as "www.icann.org") rather than the all-numeric IP addresses (such as "192.0.34.65") assigned to each computer on the Internet.
- Each domain name is made up of a series of character strings (called "labels") separated by dots.
- The right-most label in a domain name such as org is referred to as its "top-level domain" (TLD).

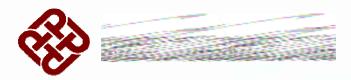




Top Level Domain

- TLDs with two letters (such as .de, .mx, and .jp) have been established for over 240 countries and external territories and are referred to as "country-code" or ccTLDs.
- Most TLDs with three or more characters are referred to as "generic" TLDs, or "gTLDs".





Generic Sponsored and Unsponsored TLDs

- The Sponsor is responsible for developing policies on the delegated topics so that the TLD is operated for the benefit of a defined group of stakeholders.
- Sponsored: .aero .coop .edu .gov .mil .museum
- Unsponsored: .biz .com .info .int .name .net .org .pro





The DNS Authority

∙ndustrial

- The Authority is the Internet Corporation for Assigned Names and Numbers (ICANN)
- ICANN is the non-profit corporation that was formed to assume responsibility for

the IP address space allocation,
protocol parameter assignment,
domain name system management,
and root server system management functions.

These functions are previously performed under U.S. Government contract by IANA and other entities.



Web Page Addressing

ndustrial

- A web page has its location on the Internet.
- A client program can access it with its address.

Point of Content

- Points of content, whether it be a page of text, a video or sound clip, a still or animated image, or a program, all have identifiable locations.
- Internet space is inhabited by many points of content.
- W3C describe their locators as Uniform Resource Identifier.



Uniform Resource Identifier

A URI typically describes:

- The way used to access the resource
- The specific computer that the resource is housed in
- The specific name of the resource (a file name) on the computer

For example, this URI:

http://www.myschool/images/school_logo.gif





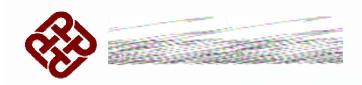
Uniform Resource Locator

- The most common form of URI is the Web page address.
- The Web page address is a particular form or subset of URI.

For example, this URL:

http://www.myschool.edu/myclass/index.html

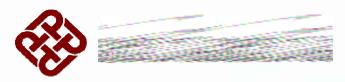




The Way to Access Resource

- http identifies a file that can be accessed using the Web protocol application, <u>Hypertext Transfer</u> Protocol.
- The HTTP is housed on a computer named " www.myschool.edu "
- This computer is called as a Web server.
- A name "www.myschool.edu "can be mapped to a unique Internet address
- Client computer accesses this server through that unique Internet address.
- Client computer access a server's resource by the Internet address and the resource's location path.





Resource Location in Host Computer

In order that resources can be accessed by a client computer, a computer named "www.myschool.edu" can store these resources in its identifiable locations such as:

/myclass
..index.html
..page1.html
..page2.html
/image
..school_logo.gif
..campus.gif
..building.gif
/reports
..report1.html
..report2.html

We are Industrial entre

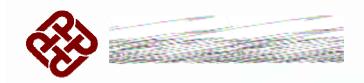


Absolute and Relative URLs

- It is sometimes inconvenient to type out the protocol and the location path every time in a Web page.
- A base URL can be declared in a Web page.
- When a base URL is declared, all the resources are only required to be specified in the location path only. The server name does not need mentioned again.

```
<hbody>
<hr/>
<hr/>
<br/>
<BASE href="http:///www.thissite.org/"></hr>
</hr>
<br/>
<br/>
<br/>
<br/>
<br/>
<br/>
<br/>
<hr/>
<br/>
<
```





Absolute vs Relative URLs

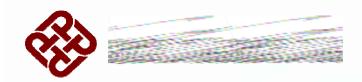
Absolute URLs

- Use a complete Internet address
- Specify the transfer protocol first
- < a href=http://www.myschool.edu/myclass/page1.html">Go
 here

Relative URLs

- Related to the address of the calling page
- No transfer protocol is specified
- Go here



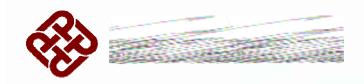


Relative Addressing Examples

- Relative URL resolves to full URL.
- If relative URL is an image
-
- It would expand to the full URL as

"http://www.myschool/images/school_logo.gif"

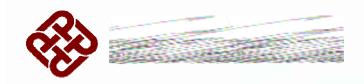




Domain Name

- A web address is also called a domain name.
- Like a memorable street address, domain name help people find a site online.
- Web address should be easy to remember for friends or customers.
- Once your domain is registered, you can build a public web site at your own web address (e.g., www.mouse.com) and create a personalized email address (e.g., micky@mouse.com).





Domain Name Registration

- Domain name can be registered through online services companies, through your ISP or directly from ICANN, The Internet Corporation for Assigned Names and Numbers.
- One of ICANN's activities is to work with other partners. These include agreement with.
- Network Solution (now VeriSign) to operate the .com, .net and .org.
- Some other companies responsible for .biz, .info and .name.
- There are over 150 ICANN-accredited registrars.





Web Site

∙ndustrial

- A Web site is a collection of Web files on a particular subject that includes a beginning file called a home page. The Web site for IBM has the home page address of http://www.ibm.com.
- A Web site is rather easily confused with a Web server.
 A server holds the files for one or more sites.
 On one hand, a very large Web site may reside on a number of servers that may in different geographic locations.
- IBM Web site consists of thousands of files spread out over many servers in world-wide locations.
- Typically a company's site resides on a commercial space provider's server.
- Web Presence is also a name for the same meaning of Web site, with less geographic implications.

 We are



Web Hosting

•ndustrial

- Web hosting (also known as Webhosting, Web site hosting, and hosting) is the business of housing, serving, and maintaining files for one or more Web sites
- Web sites occupy computer space.
- More important is speed provided for Web site files to connect to the Internet.
- Most hosting services offer connections on <u>T-carrier</u> system lines.

Typically, an individual business hosting its own site would require a similar connection and it would be expensive.

We are



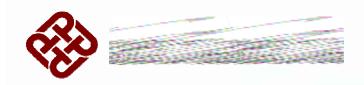
Internet Access Provider

- Many companies share the cost of a fast Internet connection for serving files with IAPs.
- Some IAPs offer subscribers free space for a small Web site that is hosted by one of their computers.
 Geocities offers registered users free space for a Web site.

These services are free, but are also very basic.

Types of Hosting are: Virtual hosting Dedicated hosting





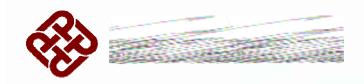
Virtual hosting

- In virtual hosting, company does not occupy an entire computer space, but each company has its own domain name and set of e-mail address.
- Servers are expensive; some hosting companies let customers to configure a computer space as if they own individual server dedicated entirely to their site.

Without involvement from the ISP. You can

- Have your own domain name
- Have your own IP address
- Administer your own file directories,
- Add e-mail accounts and address assignments,
- Assign multiple domain names that resolve to a basic domain name,
- Manage your own logs and statistics analysis,
- Maintain passwords.





Dedicated hosting

- Dedicated hosting is the provision of a <u>dedicated</u> <u>server</u> machine that is dedicated to the traffic to your Web site.
- Only very busy sites require dedicated hosting.
 Dedicated
- Many companies purchase their own servers and place them on a site that provides fast access to the Internet. This practice is called colocation.





Collocation

ndustrial

On the Internet, collocation is used to mean the provision of space for a customer's telecommunications equipment on the service provider's premises.

For example, a Web site owner could place the site's own computer servers on the premises of the Internet service provider (ISP).

Or an ISP could place its network <u>routers</u> on the premises of the company offering switching services with other ISPs.

We are



Internet Connections

- ISPs, sometimes called Internet access providers, specify their bandwidth such as the mbps or types of circuits or the number of T1 lines equivalent (1.5 Mbps).
- In Hong Kong, licensed ISPs have links connecting to HKIX. http://www.hkix.net
- HKIX (Hong Kong Internet eXchange) is a cooperative project initiated, coordinated and operated by Information Technology Services Centre, The Chinese University of Hong Kong.
- HKIX interconnect the Internet Access Providers (IAPs) in Hong Kong so that we can exchange intra-Hongkong traffic locally without routing through US. Currently 71 HK IAPs connected to HKIX. http://www.hkix.net/hkix/connected.html





Types of IAP

- IAPs such as HKNET, Pacific Supernet, with end to end control of their own overseas circuits.
- IAPs such as HK Broadband Network, PCCW Powerb@se, with direct international circuits connected to overseas IAPs.
- IAPs such as iLink, iCable, without direct international circuits connected to the overseas, they rely on the other two types of IAPs to connect circuits on their behalf.



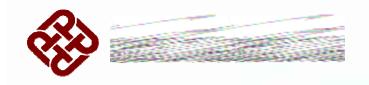


Bandwidth of IAPs

ndustrial

- IAP is self-sufficient. For example, they should have their own primary DNS, Email, WWW and News servers.
- The primary global Internet connectivity of IAPs is independent of HKIX facilities.
- Local IAP can connect to a bigger IAP with slow lines.
- Most IAPs have obtained more than a PNETS licence from OFTA, Office of the Telecommunication Authority.
- http://www.ofta.gov.hk/
- PNET is Public Non-exclusive Telecommunications
 Service licence

 We are



The WWW Ways to Search Search Engines and Directories

- With the advent of the WWW in 1994, data started migrating from Gropher and FTP servers to Web servers.
- A new industry of cataloging and indexing the Web came. Two main ways of organizing the Web are there.
- The ways are the Directories and Search Engines.

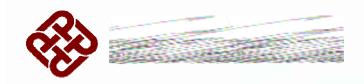




Directories

- Handpicked category. A directory doesn't search the Web.
- A directory is very organized, and very easy to use.
- Yahoo, catalogs close to 2 million individual Web pages.
- People seem to like it.
- Yahoo directory content represents less than 1/10th of 1% of the total web pages on the web.

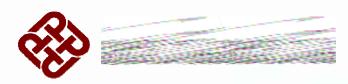




Search Engines

- Scouring the Web Automatically.
- Not powered by human hands.
- A search engine uses spider or crawler to roam the Web automatically, feeding back what it finds back to a massive bank of computers.
- The largest search engine index (Google) contains over 1.3 billion entries, less than one half of pages on the web.





Popular Search Engines / Directories

AltaVista

About

AOL Search

Ask Jeeves

Excite

Open Directory

·Yahoo

The Big Ten

Google

HotBot

The big and the powerful

AltaVista, Google, HotBot, Northern Light

Lycos

MSN Search

Northern Light

WebCrawler

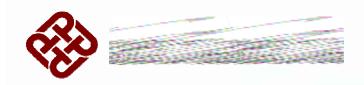
The easy-to-use

About.com, Ask Jeeves!, Yahoo!

The average

Excite, Lycos, and MSN search





Internet Society

- The Internet Society is an international non-profit organization.
- It acts as a guide and conscience for the workings of the Internet.
- It was founded in 1992 and is based on Reston, Virginia.
- It supports the Internet Architecture Board (IAB).
- IAB supervises technical and other issues.

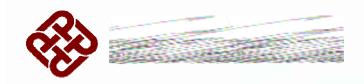




Internet Architecture Board

- IAB oversees of the technical evolution of the Internet
- One activity is the Internet Engineering Task Force (IETF)
- Another is Internet Research Task Force (IRTF)





IETF and IERF

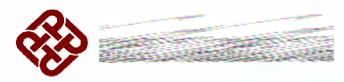
Internet Engineering Task Force

- Its members are drawn from the Internet Society's individual and organization membership.
- Its members include the Web's chief creator Tim Berners-Lee.

Internet Research Task Force

- ITRF works on network technology.
- New applications of the Internet are generating every instance.
- **■** ITRF oversees conflict of interests.
- It gives reasonable solutions to the many issues and concerns.
- It brings diverse interests together.

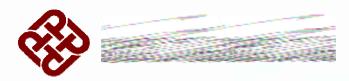




The World Wide Web Consortium

- The W3C is an industry consortium.
- It promotes standards for the evolution of the Web.
- It produces specifications and reference software for interoperability between WWW products. It is funded by industrial members.
- It is vendor-neutral, and its products are freely available to all.





Internet Assigned Numbers Authority

- IANA is the organization under the Internet Architecture Board.
- IANA is responsible for the registry for any "unique parameters and protocol values" for Internet operation.

 These include port numbers, character sets, and MIME (Multi-Purpose Internet Mail Extensions) media access types.
- http://www.iana.org/





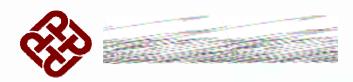
Internet Corporation for Assigned Names and Numbers

- ICANN is the private (non-government) non-profit corporation with responsibility for...
- IP address space allocation, protocol parameter assignment, domain name system management, and root server system management functions,

Since its beginning, ICANN has had to deal with controversial issues such as what new top-level domain names should be permitted.

http://www.icann.org/





Language for Web Page Publishing

- Hypertext Markup Language, HTML is the set of markup symbols or codes inserted in a file intended for display on a World Wide Web browser page.
- HTML is based on SGML, Standard Generalized Markup Language.

SGML

- SGML emphasis on descriptive rather than procedural markup.
- SGML has document type concept.
- SGML is system independent.





SGML Basics - Meta

Meta

Meta is a prefix that in most information technology usages means "an underlying definition or description".

Metadata

Metadata is a definition or description of data.

Metalanguage

Metalanguage is a definition or description of language.





SGML Basics – Mark Up Tag

Mark Up

- You insert a sequence of characters or other symbols in a text or word processing file.
- Mark ups indicate how the file:

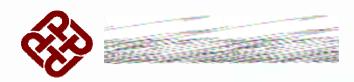
Should look when it is printed or displayed.

Describe the document's logical structure.

Tags

The markup indicators are often called "tags."





Standard Generalized Markup Language

- SGML is an international standard metalanguage. ISO 8879:1986.
- It is a standard markup definition for document structure.
- It defines how you can define markup.
- It is device-independent, system-independent methods of representing texts in electronic form.
- It is the parent language of other mark up languages.

•ndustrial

■ It is the tag-set building rules -- for HTML and for most other descriptive tag-sets.



SGML Descriptive Markup

- SGML simply provide names to categorize parts of a document.
- SGML texts are comprised of plain ASCII text, combined with items in angle brackets, e.g. <title>A Micky's Story</title>.

Examples

<para>

"the following item is a paragraph"

\end{list}

"this is the end of the most recently begun list"





Tags

■ Tags are in angle brackets, tags designate structural divisions.

<title> A Micky's Story</title>

<scene>On the Island</scene>

OR

- Tags convey information about renditional and typographical elements
- changes in typeface, line breaks, etc.





Viable Tags

•ndustrial

- Tags are composed of plain text ASCII characters.
- They are easy to deliver across a network.
- Italic Font typically lost when the text is transferred out of WordPerfect and into another format because WordPerfect is designed for desktop publishing. Tags that have meaning with WordPerfect do not necessary have meaning with other computer systems.
- SGML tags are simply other letters and characters typed in as part of the text.
- SGML tags travel with the text if it moves from computer system to computer system. We are



SGML Features

SGML Data Independent

- SGML ensures documents encoded be transportable.
- It let you transport document from one hardware and software environment to another without loss of information

SGML Display Independent

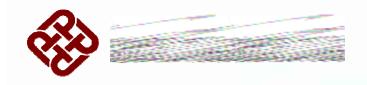
- The actual display of such a document may vary, depending on the output medium and style preferences.
- SGML compiler can interpret any document by reference to its document type definition (DT/D)ne



SGML Based Language

- Hypertext Markup Language is an example of an SGML-based language.
- HTML has its Document Type Definition.
- Extensible Markup Language is another application of SGML.
- XML has its own ways to define document.





Document Type

■ The type of a document is formally defined by its constituent parts and their structure.

For example

- The definition of a report, consisted of a title and possibly an author, followed by an abstract and a sequence of one or more paragraphs.
- Anything lacking a title, according to this formal definition, would not formally be a report.





Document Type Definition

- A DTD is a metalanguage for a certain type of document.
- DTD defines the legal building blocks of a document of a type.
- DTD is introduced by SGML.





SGML Basic Types

- CDATA (sequence of characters)
- ID and NAME ([A-Za-z], [0-9], [-_:.])
- IDREF and IDREFS (ID token or spaced ID tokens)
- NUMBER ([0-9])

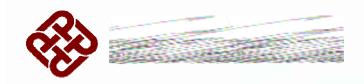
Refer to

http://www.w3.org/TR/REC-html40/types.html#type-cdata

HTML DTD

http://www.w3.org/TR/REC-html40/sgml/dtd.html





Hypertext Markup Language

- HTML is the set of <u>markup</u> symbols or codes inserted in a file intended for display on a World Wide Web browser page.
- HTML is a formal Recommendation by the World Wide Web Consortium.
- Microsoft's Internet Explorer and Netscape's Navigator also provide some additional non-standard codes.
- The current version of HTML is HTML 4.01
- http://www.w3.org/TR/html401/





HTML Versions

Previous Versions

- HTML 2.0 [RFC1866]
- HTML 3.2 [HTML32]

HTML 4

- HTML 4 supports more multimedia options, scripting languages, style sheets, better printing facilities, and documents that are more accessible to users with disabilities.
- HTML 4 is an SGML application.





HTML DTD

- The bulk of the HTML DTD consists of the declarations of *element types* and their attributes.
- The <!ELEMENT keyword begins a declaration and the > character ends it.

For Example

<!ELEMENT UL - - (LI)+>

UL (unorder list),

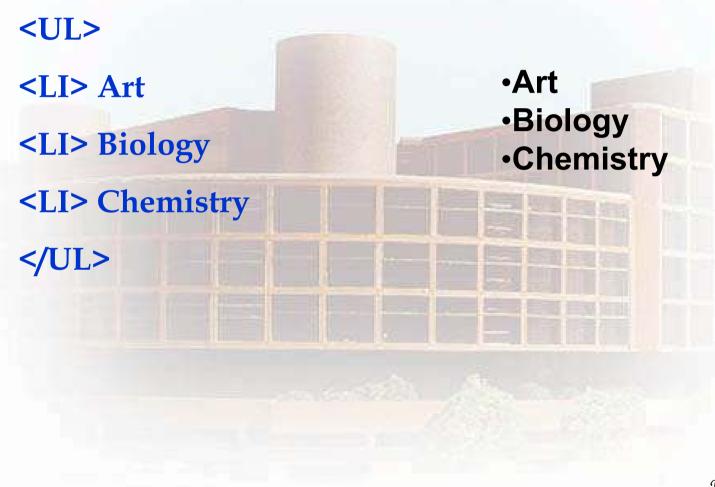
LI (list item),

-- (both start and end tags are required.)





Example Use of DTD Unordered List



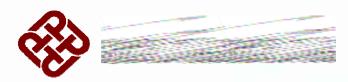




Features in HTML 4.0

- The <u>cascading style sheet</u>, the ability to control Web page content at <u>multiple</u> levels.
- The ability to create richer forms.
- Support for frames.
- Enhancements for tables that make it possible to use captions to provide table content for Braille or speech users.
- The capability to manage pages so that they can be distributed in different languages.





Extensible Hypertext Markup Language

- W3C describes XHTML "a reformulation of HTML 4.0 as an application of the Extensible Markup Language (XML)."
- XHTML is the follow-on version of HTML 4.

Refer to

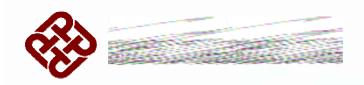
W3C HTML Specification

http://www.w3.org/TR/REC-html40/

Extensible Hypertext Markup Language XHTML 1.0

http://www.w3.org/TR/xhtml1/





Global Structure of HTML

An HTML 4 document is composed of three parts:

- a line containing HTML version information,
- a declarative header section (delimited by the <u>HEAD</u> element),
- a body, which contains the document's actual content.
 The body may be implemented by the BODY element or the FRAMESET element

Refer to

http://www.w3.org/TR/REC-html40/struct/global.html#h-7.1





HTML Structure Example

ABC

·Art

·Biology

<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN">

<HTML>

<HEAD>

<TITLE>Subjects</TITLE>

</HEAD>

<BODY>

ABC

 Art

 Biology

 Chemistry

</BODY>

</HTML>

•Chemistry





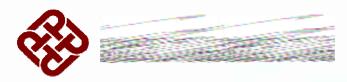
The Head Element

■ Information in the declarative Head section, within Head tags pair is not displayed in general.

```
<HEAD>
<TITLE>Subjects</TITLE>
</HEAD>
```

- User agents, such as browsers, do not generally render elements that appear in the <u>HEAD</u> as content.
- User agents may, however, make information in the HEAD available to users through other mechanisms.





Information Inside Head Declaration

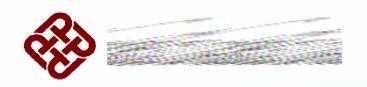
The <u>HEAD</u> element contains information about the current document such as

- The document's title, keywords that may be useful to search engines.
- The document's other data that is not considered document content.

More Information Inside Head Declaration

- Base
- Meta
- Script





Base

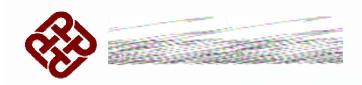
- In HTML, links and references to external elements by a resource indicator or locator.
- Base must appear in HEAD section before any element that refers to an external source.

The relative URI
"../mouse/micky.gif"
would resolve to:

http://www.cartoon.com/ mouse/micky.gif

- <HTML>
- <HEAD>
- <TITLE>Our Products</TITLE>
- </HEAD>
- <BODY>
- <P>Please visit Micky Mouse
- </BODY>
- </HTML>



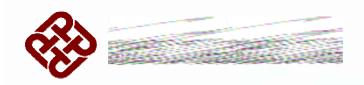


Script

- A client-side *script* is a program that may accompany an HTML document or be embedded directly in it.
- Some languages have been conceived expressly as script languages.

<script language = "JavaScript 1.2">





Meta

- Information about a document rather than document content
- Machine-readable

```
<HTML>
```

<HEAD>

<TITLE>Industrial Centre</TITLE>

<META name="keywords" content="Technology Training Services">

<META name="description" content="Innovation Centre">

<HEAD>

. . .





Use of Meta Tag

•ndustrial

- Meta can be used by search engines to classify and identify the document without downloading the entire document
- Meta is used identify properties of a document: author, expiration date, a list of key words, etc.
- The syntax use of Meta is like this...

<META [http-equiv][name][content][scheme] >

Such as <META name="Author" content="Simon Hui">

Refer to

http://www.w3.org/TR/REC-html40/struct/global.html#h-7.1



Meta and Search Engines

■ A common use for meta is to specify keywords that a search engine may use to improve the quality of search results.

<META name="keywords" content="vacation, Greece, sunshine">



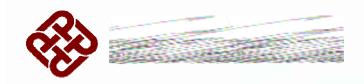


Meta and Language

- Several META
 elements can provide
 language-dependent
 information about a
 document
- Search engines may filter on the lang attribute to display search results using the language preferences of the user.

- <-- For speakers of US English -->
 <META name="keywords" lang="en-us"</p>
 content="vacation, Greece, sunshine">
 - <-- For speakers of British English -->
 <META name="keywords" lang="en"
 content="holiday, Greece, sunshine">
 - <-- For speakers of French -->
 <META name="keywords" lang="fr"
 content="vacances, Grèce, soleil">





Meta and Default Information

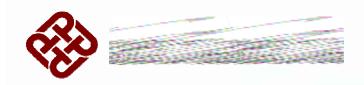
The meta element may be used to specify the default information for a document in the following instances:

- The default scripting language.
- The default style sheet language.
- The document character encoding.

For Example: character Encoding

<META http-equiv="Content-Type" content="text/html; charset=ISO-8859-5">





Meta and Scheme

- Meta Scheme provides user agents more context for the correct interpretation of meta data
- The syntax use of Meta is like this...

<META [http-equiv][name][content][scheme] >

For example,

<META scheme="ISBN" name="identifier" content="0-8230-2355-9">





The Body Element

Body Tags enclose the main portion of the document

<BODY>

Inside is the document content.

</BODY>

■ The syntax use of Body is

<BODY [bgcolor=value][onload=action][onunload=action][events]>





Intrinsic Events

<BODY

[bgcolor=value][onload=action][onunload=action][events]>

Intrinsic events are changes that likely to occur and how scripts are bound to events.

Examples are:

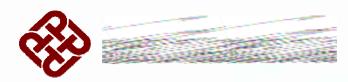
Onload, onunload, onclick, ondbclick, onmousedown, onkeypress, etc.

Intrinsic events work with your scripts.

Refer to

http://www.w3.org/TR/REC-html40/interact/scripts.html#events





Presentation of Body of Document

- The body of a document contains the document's content.
- The content may be presented by a user agent in a variety of ways.

For example,

- for visual browsers, you can think of the body as a canvas where the content appears: text, images, colors, graphics, etc.
- For audio user agents, the same content may be spoken.





The Document Body

The document body consists of many elements:

- The element identifiers: id and class attributes
- Block-level and inline elements
- Grouping elements: the DIV and SPAN elements
- Headings: The H1,H2,H3,H4,H5,H6 elements
- The ADDRESS element





Grouping Elements

Block-level

DIV (Division) element is used to structure a HTML document into a series of divisions.

Inline

SPAN element groups content to be marked inline.

DIV and SPAN are structural element important to positioning abilities.

The most common use for DIV and SPAN is marking sections of text formatting or positioning with style sheets.

The above declaration is not obvious, go to example ...





Division and Span

```
<BODY>
<DIV id="client-mouse" class="client">
 <P><SPAN class="client-title">Client information:</SPAN>
 <TABLE class="client-data">
   <TR><TH>Last name:<TD>Micky</TR>
   <TR><TH>First name:<TD>Mouse</TR>
   <TR><TH>Email:<TD>micky@mouse.org</TR>
 </TABLE>
</DIV>
<DIV id="client-duck" class="client">
 <P><SPAN class="client-title">Client information:</SPAN>
 <TABLE class="client-data">
   <TR><TH>Last name:<TD>Donald</TR>
   <TR><TH>First name:<TD>Duck</TR>
   <TR><TH>Email:<TD>donald@duck.com</TR>
  </TABLE>
</DIV>
</BODY>
```

Client information:

Last name: Micky

First name: Mouse

Email: micky@mouse.org

Client information:

Last name: Donald

First name: Duck

Email: donald@duck.com





Element Identifiers

■ ID attribute assigns a name to an element. This name must be unique in a document.

<P id="myparagraph"> This is a uniquely named paragraph.

- Class attribute assigns one or more class names to an element.
- The element may be said to belong to these classes.
- A class name may be shared by several element instances
- For example, in your HTML code,

Your style sheet may has included:

P.left {text-align: left}

P.right {text-align: right}





Headings

■ The H1,H2,H3,H4,H5,H6 elements

<body>

<h1>Forest</h1>

<h3>Animal</h3>

<h6>Tiger</h6>

</body>

Forest

Animal

Tiger





Address Tags

The address tag behaves much like the paragraph tag.

<ADDRESS>

Micky
Mouse

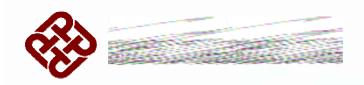
DisneyLand

**Hong Kong
**

</ADDRESS>

<u>Micky Mouse</u> DisneyLand Hong Kong





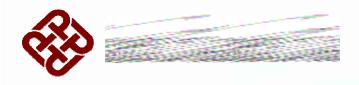
Framesets and HTML Bodies

 Documents that contain framesets replace the BODY element by FRAMESET element.

Frames

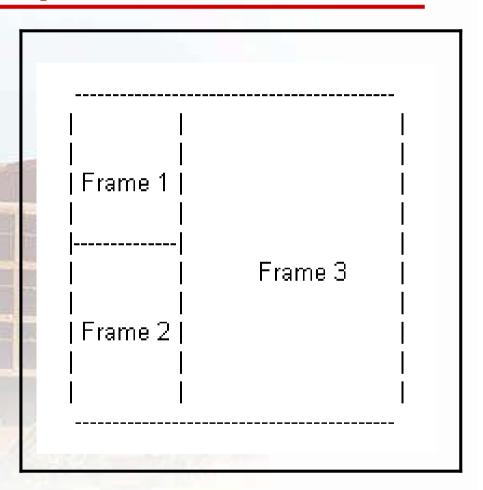
- HTML frames allow you to present documents in multiple views.
- The views may be independent windows or subwindows.
- Multiple views offer designers a way to keep certain information visible, while other views are scrolled or replaced.





Multiple Views of Frame

- For example, within the same window,
- one frame might display a static banner,
- a second a navigation menu, and
- a third the main document that can be scrolled through or replaced by navigating in the second frame.

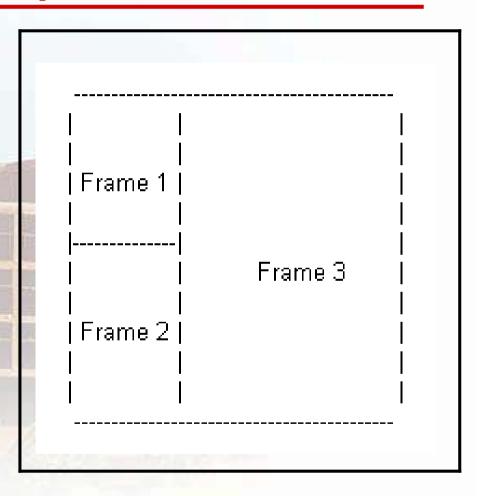






Multiple Views of Frame

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD</pre>
HTML 4.01 Frameset//EN"
"http://www.w3.org/TR/html4/frameset.dtd">
<HTML>
<HEAD>
<TITLE>A simple frameset document</TITLE>
</HEAD>
 <FRAMESET cols="20%, 80%">
   <FRAMESET rows="100, 200">
    <FRAME src="contents_of_frame1.gif">
    <FRAME src="contents_of_frame2.html">
    </FRAMESET>
   <FRAME src="contents_of_frame3.html">
 </FRAMESET>
</HTML>
```







Frameset Divide

Rows Divide Vertically

<FRAMESET rows="50%, 50%">

...the rest of the definition...

</FRAMESET>

This indicate the frameset has 2 rows, each frame occupy 50% of the vertical screen size.

Columns Divide Horizontally

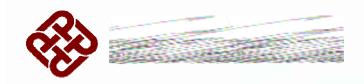
<FRAMESET cols="1*,250,3*">

...the rest of the definition...

</FRAMESET>

This indicates the frameset has 3 columns, the 2nd frame occupied horizontally 250 pixels, the rest is shared by the 1st and 3rd frames.

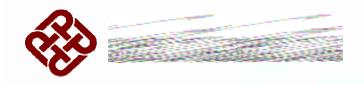
We are not not a strict the strict of the stri



Dynamic HTML

- Dynamic HTML is not a standard defined by the World Wide Web Consortium (W3C).
- DHTML was used by Netscape and Microsoft to describe the new technologies the 4.x generation browsers would support.
- DHTML is the combination of HTML, style sheets and scripts that allows documents to be animated.





Concepts of DHTML

- An object-oriented view of a Web page and its elements
- Cascading style sheets and the layering of content
- Programming that can address all or most page elements
- Dynamic fonts





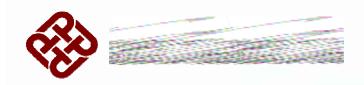
Object-oriented View of Web Page

- Each page element (division or section, heading, paragraph, image, list, and so forth) is viewed as an "object".
- W3C calls it the "Document Object Model".

Style Sheet and Layering

- For Web pages, a style sheet describes the default background color or image, hypertext link colors, and possibly the content of page.
- Style sheets help ensure consistency across all or a group of pages in a document or a Web site.
 We are

ndustrial



Dynamic Fonts

- Netscape includes <u>dynamic fonts</u> as part of dynamic HTML.
- This feature lets Web page designers include font files containing specific font styles, sizes, and colors as part of a Web page.
- Fonts are downloaded with the page.

 That is, no longer the font choice is dependent on what the browser provides.





Cascading Style Sheet

- Although DHTML is not a W3C standard, Cascading Style Sheet, level 1 is recommended by World Wide Web Consortium.
- CSS is implemented in the latest versions of the Netscape and Microsoft Web browsers.
- CSS specifies the possible style sheets or statements that may determine how a given element is presented in a Web page.

CSS1, Refer to

http://www.w3.org/TR/REC-CSS1-961217.html





Other W3C Style Sheets

W3C Recommends CSS1, CSS2, XPath, XSLT.
 CSS especially is widely implemented in browsers.

Refer to

http://www.w3.org/Style/

- CSS1 has around 50 properties (for example color and font-size).
- In 1998, CSS2 includes all CSS1 properties and adds around 70 of its own, such as properties to describe aural presentations and page breaks.

CSS2 builds on CSS1. Refer to http://www.w3.org/TR/REC-CSS2/





The Use of CSS

CSS gives more control over the appearance of a Web page to the page creator than to the browser designer or the viewer.

More Than One Style

Q. What style will be used when there is more than one style specified for an HTML element?

A. Styles will "cascade" into a new "virtual" Style Sheet by the following rules, where number one has the highest priority:

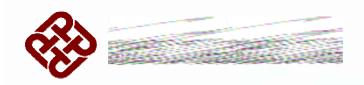
Inline Style (inside HTML element)

Internal Style Sheet (inside the <head> tag)

External Style Sheet

Browser default



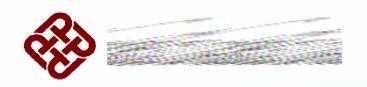


Inline Style Sheet

- An inline style (inside an HTML element) has the highest priority, which means that it will override every style declared inside the <head> tag, in an external style sheet, and in a browser (a default value).
- Technically not a sheet implementation.
- This option uses style sheet syntax

<TAG style="attribute1:value1; ...">





Style Sheet Rule

A rule consists of two parts:

- Selector the part before the left curly brace
- Declaration the part within the curly braces

Selector and Declaration

- The selector is the link between the HTML document and the style.
- Selector specifies what elements are affected by the declaration.
- The declaration is that part of the rule that sets forth what the effect will be.



A declaration has two parts separated by a colon:

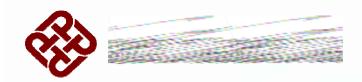
```
H1 { color: green }
Property Value
```

For Example,

H1 {color: green; text-align: center;}



http://www.w3.org/Style/LieBos2e/enter/



Example of Use of Style Sheet

```
<HTML>
<TITLE>Micky's home page</TITLE>
 <STYLE>
   H1, H2 { color: green }
 </STYLE>
<BODY>
<H1>Micky's Home Page</H1>
<P>Micky has:
 <UI.>
   <LI>Big eyes
   <LI>Short tail
   <LI>Beautiful smile
 </UL>
<H2>Home of Micky</H2>
<P>Micky will stay in Hong Kong in 2005.
</BODY>
</HTML>
```

Micky's Home Page

Micky has:

- Big eyes
- Short tail
- Beautiful smile

Home of Micky

Micky will stay in Hong Kong in 2005.





HTML Tree Structure

Internal / Embedded Style Sheet

<HTML>

<HEAD> ... </HEAD>

<STYLE> ... </STYLE>

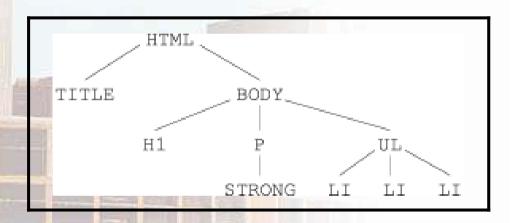
<BODY> ... </BODY>

</HTML>

Like internal style sheet, the linked (external) style sheet contains a set of STYLE selectors and declarations.

External style sheet should be saved with a .css extension.

The tree structure of this document is:



External / Linked Style Sheet

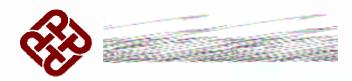
The linked style sheet is a .css file.

<HEAD>

<LINK rel=stylesheet

href="filename.css" type="text/css">
</HEAD>

The industrial indus



An Example of an External Style Sheet File

- An external style sheet can be written in any text editor.
- The file should not contain any html tags.
- Your style sheet should be saved with a .css extension.
- Style Attributes

Fonts, such as size, bold, italic, etc.

Color and Background

Text, such as values for spacing, underlining etc.

Margins, Padding, and Borders

Classification, include attributes for list items in HTML tag.

hr {color: red} p {margin-left: 20px} body {background-image: url("images/backgd.gif")}





Tell Browser about CSS

You can tell browser which style sheet language is being used.
When the browser loads a document, it checks to see if it understands the style sheet language. If it does, it will try to read the sheet, otherwise it will ignore it.

```
<HTML>
<TITLE>Micky's home page</TITLE>
<STYLE TYPE="text/css">
H1 { color: green }
</STYLE>
<BODY>
...
</BODY>
</HTML>
```





Multiple Style Attributes

Multiple attributes for an element separated by ;

P { text-align: center; color: black; font-family: arial }

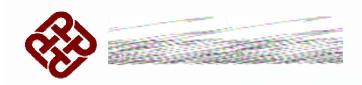
Add classes for element can group foundation attributes with additional style sheets. P.message {font-size: 12pt; font-weight: bold}

P.note {font-size: 14pt}

The classes note and message share commons of element P.

Using together rendered

Style P.message will be 12pt, bold, black, arial and center aligned.



Extensible Markup Language

We are

·ndustrial

- XML is a flexible way to create common <u>information</u> formats.
- XML aims to share both the format and the <u>data</u> on the World Wide Web, intranets etc.
- XML is similar to the language of today's Web pages, the Hypertext Markup Language

XML and SGML

- SGML is the W3C standard for how to create a document structure.
- HTML is an application of SGML, XML is a simpler and easierto-use subset of SGML.
- XML is not an application of SGML.

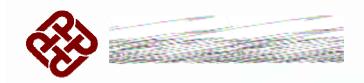


W3C Goals of Using XML

W3c XML specifications set out:

- Straightforwardly usable of the Internet
- Support a wide variety of applications
- Easy to write programs that process XML documents
- Keep minimum optional features
- Human legible XML documents





Why Use XML

- Share data without application-specific format by a user-defined DTD.
- Groups or organizations employing the same DTD then know each can use or apply data created by any of the other groups.
- DTD can sent with XML data and provide the recipient with all information needed to interpret and use the data.
- XML data is not depended on specific features of computing platform.
- Upgrade, extend a system or application as needed so far as these platform changes still can use the data without modifying the data.

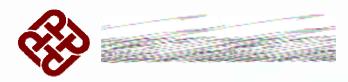




Extensible

- In HTML, the markup symbols are not defined by users.
- XML is "extensible" because the markup symbols are unlimited and self-defining.

```
<?xml version="1.0" standalone="yes">
<HTML> ...
<UL>
                           <friend>
 <LI>First Name: Micky
                             <firstname>Mickey</firstname>
 <LI>Last Name: Mouse
                             <lastname>Mouse</lastname>
 <LI>Address: DisneyLand,
  Hong Kong
                              <address>DisneyLand, Hong
                             Kong</address>
</UL>
                                                       We are
                           </friend>
                                                            ndustrial
</HTML>
```



Strict Conformance Rules in XML

- Data reusability requires well-formed
- Declaration
- Only one root element
- Tags with start and end exist in pairs
- Nested tags be in order, e.g. <A>....
- ABC not equal Abc and not equal to abc
- Attributes in " or '

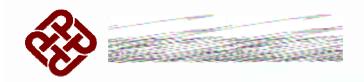




XML Does Not Do Anything

- HTML tags tell a client program to present its content. (i.e. HTML Do something)
- Maybe it is a little hard to understand, but XML does not DO anything.
 XML is created to structure, store, and relocation of information.
- XML itself is not related to any application such as present or modify its data.
- XML is not an application of SGML, XML is a sub-set of SGML.





The Co-exist of XML and HTML

- XML is not a replacement for HTML.
- XML is used to describe the data, while HTML will be used to format and display the same data.

XML can Separate Data from HTML

- When HTML is used to display data, the data is stored inside your HTML.
- With XML, data can be stored within HTML or in separate XML files.





Within HTML or in Separate XML file

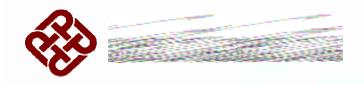
Data Stored within HTML

■ XML data can be stored inside HTML pages as "Data Islands". You can concentrate on using HTML only for formatting and displaying the data.

Data Stored in Separate XML Files

■ In separate XML files, you can concentrate on using HTML for data layout and display, and be sure that changes in the underlying data will not require any changes to your HTML.





The Use of DTD in XML

- DTD is sent with XML data with all information needed to interpret and use the data.
- DTD can reside in the XML document, or stored in separate file.
- DTD can be linked to from within a XML document.
- Similar to CSS can be linked to from within a HTML document.



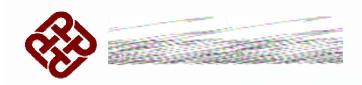


DTD XML Building Blocks

Seen from a DTD point of view, all XML documents (and HTML documents) are made up by the following simple building blocks category:

- Elements
- Tags
- Attributes
- Entities
- PCDATA
- CDATA





DTD Declaring an Element

In the DTD, XML element declaration has the following syntax:

<!ELEMENT element-name category>

or

<!ELEMENT element-name (element-content)>

Example of XML DTD

<!ELEMENT note (to,from,heading,body)>

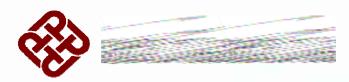
<!ELEMENT to (#PCDATA)>

<!ELEMENT from (#PCDATA)>

<!ELEMENT heading (#PCDATA)>

<!ELEMENT body (#PCDATA)>

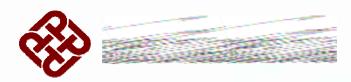




Example of XML with DTD Inside

•ndustrial

```
<?xml version="1.0" standalone="yes">
<!DOCTYPE note [
<!ELEMENT note (to,from,heading,body)>
               (#PCDATA)>
<!ELEMENT to
<!ELEMENT from (#PCDATA)>
<!ELEMENT heading (#PCDATA)>
<!ELEMENT body (#PCDATA)>
]>
<note>
<to>Micky</to>
<from>Minnie</from>
<heading>New Home</heading>
<br/><body>We will stay in Hong Kong</body>
                                                      We are
</note>
```



Example of XML with DTD Outside

```
<?xml standalone="no" version="1.0">
<!DOCTYPE note SYSTEM "note.dtd">
<note>
<to>Micky</to>
<from>Minnie</from>
<heading>New Home</heading>
<body>We will stay in Hong Kong</body>
</note>
```





References

- Fallon, Thomas, The Internet Today, Prentice Hall, 2001.
- Miller, Michael, Special Edition Using The Internet and Web, Que, 2002.

