

# CT216 Software Engineering Tutorial

Eoin O Fiachain

23rd/30th September 2004

## 1 Tutorial Schedule

The tutorials will occur every Thursday morning at 9am in Room TB302 in the Tower Block. They will cover various aspects of the technologies involved in the CT216 Software Engineering Project.

There will be 5 tutorial sessions in all. Tutorial sessions will be run separately for each of the two groups in the class on alternating Thursdays. Accordingly, each student need only attend one tutorial every fortnight.

The schedule is as follows:

Sep 23	The World Wide Web and HTML (Group 1)
Sep 30	The World Wide Web and HTML (Group 2)
Oct 7	Web Development in PHP (Group 1)
Oct 14	Web Development in PHP (Group 2)
Oct 21	MySQL (Group 1)
Oct 28	MySQL (Group 2)
Nov 4	Integrating MySQL with PHP; Using CVS (Group 1)
Nov 11	Integrating MySQL with PHP; Using CVS (Group 2)
Nov 18	Advanced PHP Concepts (Group 1)
Nov 25	Advanced PHP Concepts (Group 2)

Tutorial notes will be placed in the *Software Engineering Project* folder in the CT216 page on *blackboard.nuigalway.ie*. This folder also contains a collection of links to useful online resources for the software engineering project (tutorials, API documentation, etc.).

Questions are welcome at any point throughout the tutorial. I can be contacted with any further queries at the conclusion of the tutorial or by email at *eoim@eoimofiachain.com*.

## 2 Introduction to The World Wide Web

### 2.1 What is the Internet

The Internet is a publicly available worldwide system of interconnected computer networks that communicate using the Internet Protocol (IP).

### 2.2 What is the World Wide Web

The World Wide Web is a system of internet servers that support specially formatted documents (called HTML pages). These pages provide links to other HTML pages and to other types of documents and files stored on the internet. Not all Internet servers are part of the world wide web.

There are are three main standards employed in the World Wide Web:

- **The Uniform Resource Locator (URL)**  
Specifies a unique address for each page of information  
e.g. <http://www.nuigalway.ie/somepage.html>
- **Hyper Text Transfer Protocol (HTTP)**  
Specifies how to send information to and from the server
- **Hyper Text Markup Language (HTML)**  
A method of encoding information so it can be displayed on a variety of devices

### 2.3 Uniform Resource Locator (URL)

A URL (also known as a *web address*) is a standardised unique address for some resource (such as a document or image) on the Internet.

Each URL contains a number of elements which collectively uniquely identify a particular resource and would typically take the following form:

`protocol://host:port/path`

For example, the URL representing the the CT216 homepage is:

`http://corrib.it.nuigalway.ie:80/content/teaching/ct216/ct216.html`

Elements found a typical URL would be:

- **protocol**

The protocol (set of rules/standards) indicates how to access the resource. This is typically set to *http* and most web browser applications will automatically assume the HTTP protocol if a specific protocol is omitted from the URL.

Other protocols often used in URLs include *ftp* (File Transfer Protocol) and *https* (HTTP over SSL).

- **host**

This is the internet address of the server that is providing (also called *hosting*) the resource.

It can be specified either by domain name (e.g. *corrib.it.nuigalway.ie*) or by IP (Internet Protocol) address (e.g. *140.203.16.54*).

- **port**

The port is a number representing a specific end-point destination on the internet server. As internet servers can host many different services and protocols, each service is identified by a unique number.

An analogy would be an apartment building. The building has a specific address but each apartment within the building is assigned an additional number in order to differentiate between them. When a letter arrives in the post it can be assigned to the appropriate apartment.

Services on an internet server are differentiated in a similar manner using the port number. The primary HTTP service on an internet server is typically assigned the port number *80* by default. Therefore if the *http* protocol is employed in a URL and no port number is specified then port *80* will be assumed.

- **path**

The path identifies a specific resource (file) on a web server. Web sites typically involved a large number of files which are divided into directories and subdirectories.

## 2.4 Hyper Text Transfer Protocol (HTTP)

HTTP is the primary method used to communicate information on the World Wide Web. It uses the *TCP/IP* networking protocols.

It is based on the request/response paradigm where one host acts as a *client* and another as a *server*. The client connects to the server and sends a *request* for a particular file. The server then sends a *response* containing the file (or an error message).

The particular file to be requested is identified by a URL. The response can contain many types of data but will often contain a HTML file.

The following is a sample request and response using [www.google.com](http://www.google.com).

### Client Request:

```
GET / HTTP/1.1
Host: www.google.com
```

### Server Response:

```
HTTP/1.1 200 OK
Content-Length: 3059
Server: GWS/2.0
Date: Sat, 11 Jan 2003 02:44:04 GMT
Content-Type: text/html
Cache-control: private
Connection: keep-alive
```

*followed by HTML text*

Each HTTP request is considered independent from each other and a new TCP/IP connection to the server is made on each request. We will learn about how to use *cookies* and *sessions* in later tutorials in order to identify requests from the same source.

### 2.4.1 HTTP Query String

HTTP URLs may have an additional element called a *Query String*. This is placed after the *path* element following a question mark (?) character.

This is used to pass further information (or parameters) to a particular resource on a web server.

The following is an example of a URL containing a Query String obtained after I searched for Galway on google.

```
http://www.google.ie/search?hl=en&ie=UTF-8&q=galway&btnG=Search&meta=
```

The query string is `hl=en&ie=UTF-8&q=galway&btnG=Search&meta=`, which contains a number of parameters for the google search query.

The parameters are parsed into name/value pairs separated by the `&` and `=` characters. In the preceding example the first parameter would have the name 'hl' and the value 'en'.

Certain characters, known as *unsafe characters*, are not allowed in the query string. These are converted into an alternative format `%xx` where `xx` is the ASCII code of the character. Unsafe characters include `=`, `&`, `%` and `+`.

## 2.5 Hyper Text Markup Language (HTML)

HTML is a markup language designed for presenting information on the World Wide Web. It is an Internet standard maintained by the World Wide Web Consortium (W3C).

There are various versions of the HTML standard the most recent of which is HTML 4.01. XHTML (a very similar standard based on XML) is a newer standard which looks set to become the dominant standard in years to come.

HTML pages can contain hyperlinks to other resources and HTML pages on the internet.

The links section on the CT216 page contains links to many online HTML tutorials that will provide a guide to writing HTML pages. Understanding of HTML is absolutely essential to the CT216 Software Engineering project, and is an important pre-requisite for web programming in PHP.

Some of the main elements of HTML will be summarised at the end of the tutorial but a thorough understanding of HTML can only be obtained through practice. It is highly recommended that students familiarise themselves with writing basic HTML documents before the next tutorial.

### 2.5.1 Writing HTML

The *Notepad* application can be used to write HTML documents. Simply, write the HTML markup code in the editor and save the file with a *.html* or *.htm* extension. Double-click on the file's icon in *Windows Explorer* and the file will be opened in your default web browser.

After making changes to the HTML code, make sure to save the file in *Notepad* and select the *refresh* icon in your web browser to display the updated page.

## 2.6 What is a Web Server

This is a computer that stores HTML pages and makes them accessible to other computers on the World Wide Web through the HTTP protocol. It is also known as a HTTP server. Common HTTP servers include Apache and Microsoft's Internet Information Services.

A web server would typically be left running continuously in order to ensure a website is permanently accessible. A single web server often hosts multiple websites with different host names through a process called *virtual hosting*.

## 2.7 What is a Web Browser

A web browser is an application that allows the user to access the World Wide Web. It acts as a HTTP client and connects to a HTTP server, requests a resource, waits for a response, and processes the response accordingly. If the response is a HTML file the web browser will parse the HTML markup and display the file appropriately.

Internet Explorer is the dominant web browser but other web browsers include Mozilla, Mozilla Firefox and Netscape Navigator.