CS1115/CS5002
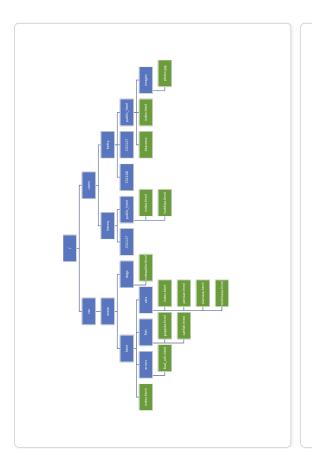# Web Development 1

Dr Derek Bridge

School of Computer Science & Information Technology

University College Cork

Directory tree:

- /
  - var
    - www
      - html
        - errors
          - index.html
          - bad_urls.html
        - fish
          - guppies.html
          - catfish.html
        - cats
          - index.html
          - persian.html
          - siamese.html
          - tonkinese.html
        - dogs
          - rottweiler.html
  - users
    - betty
      - CS1117
        - public_html
          - index.html
          - .htaccess
          - images
            - photo.jpg
      - CS1106
    - benny
      - CS1117
        - public_html
          - index.html
          - hobbies.html

Background: Unix file systems

- **Directories** (what *Windows* calls **folders**):
  - can contain other directories
  - can contain files
- The topmost directory (the **root**) is known as /
- The directory that contains your personal files, the one you log into, is your **home directory**
- A **pathname** is a list of pathnames you must 'travel' through to get to where you want to go (separated by /)
- An **absolute pathname** is one that begins at the root

Absolute pathnames

- Examples:
  - The absolute pathname for the directory called CS1106 is /users/betty/CS1106
  - The absolute pathname for the file called guppies.html is /var/www/html/fish/guppies.html
- **Exercise:** Write down the absolute pathname for benny's file called index.html

Web servers

- Just two or three main Web servers available:
  - Apache: www.apache.org
  - Microsoft Internet Information Server (IIS): www.iis.net
  - Nginx: wiki.nginx.org
- Web server stats: www.netcraft.com

The Apache web server

- Apache is modular, and you decide which of its modules to install, e.g.
  - `mod_mime` to make sure the server includes the correct `Content-Type` header in its response
  - `mod_log_config` to configure logging of HTTP requests
  - `mod_ssl` for encryption using the Secure Sockets Layer/Transport Layer Security protocols
- You enable/disable and configure the modules that you have installed by placing **directives** into configuration files
- The main configuration file is called `httpd.conf`

File and directory names

- Avoid spaces
- Avoid punctuation symbols except underscore, `_`
- For files, use a proper **filename extension**, e.g.:

`.html`	`.css`	`.gif`	`.jpg`	`.png`	`.js`

- Unix is case-sensitive, so best to consistently use lowercase
- Be descriptive but keep them short

Unix permissions

Owner	the person who owns the file/directory
Group	users who belong to the group associated with the file/directory
Other	all users who can login to the system

Read	view the contents of the file
	list the names of the files in the directory
Write	modify the contents of the file or delete it
	add files to the directory, remove files from it, rename files in it
Execute	run the file (if it is a program)
	traverse the directory and, if also have read permission, list the files, view their contents, view their permissions,…

httpd.conf example

```
ServerName www.example.com

Listen 80

DocumentRoot "/var/www/html"
```

- **Question:** What does the ServerName directive specify?
- **Question:** What does the Listen directive specify?
- DocumentRoot:
  - This is where on the server you are storing all HTML files, CSS files, image files, JavaScript programs, etc.
  - By default, if a browser makes a request, the URL pathname is appended to the DocumentRoot pathname
  - E.g. if a browser requests
    http://www.example.com/fish/guppies.html,
    the server will serve /var/www/html/fish
    /guppies.html

Error pages

- Some people/companies like to serve a customized error page, e.g.:
  - Derek Bridge's web site
  - https://www.kualo.co.uk/
- Use an ErrorDocument directive in httpd.conf (or your .htaccess file)

```
ServerName www.example.com

Listen 80

DocumentRoot "/var/www/html"

ErrorDocument 404 /errors/bad_url1.html
```

```
<!DOCTYPE html>
<html>
  <head>
    <title>404 error</title>
  </head>
  <body>
    <p>
      Oops!
    </p>
  </body>
</html>
```

Class exercise

A browser requests the following URLs. What will the server's responses be (assuming file/directory permissions are OK)?

- http://www.example.com/cats/tonkinese.html
- http://www.example.com/cats/manx.html
- http://www.example.com/dogs/rottweiler.html

What will the responses be if the cats directory lacks execute permissions?

Directory listings and index files

- If a browser requests a file, the server serves that file (if it exists and if permissions are OK)
- But what does a server do if a browser requests a directory?
  E.g. http://www.example.com/fish/
  - It may return an error (usually 404, sometimes 403)
  - It may return a default page, usually from the directory
    - Requires mod_dir module to be installed and enabled
    - Name of the file (usually index.html) is given in the DirectoryIndex directive
  - It may return a page that lists the contents of the directory
    - Requires mod_autoindex to be installed and enabled
    - Uses the Options +Indexes directive
    - **Question:** Some people think this is a bad idea. Why?

Class exercise

A browser requests the following URLs. What will the server's responses be (assuming file/directory permissions are OK)?

```
ServerName www.example.com

Listen 80

DocumentRoot "/var/www/html"

DirectoryIndex index.html

<Directory />
    Options +Indexes
</Directory>
```

- http://www.example.com/
- http://www.example.com/fish/
- http://www.example.com/cats/

User web sites

- We want you to have your own web site
- But we prefer if the files are in a directory in your **home directory**, not in the server's document root
- The mod_userdir module makes this possible
- We configure it using the UserDir directive, e.g.:

  `UserDir /users/*/public_html`

  This says your web pages will be in your public_html directory

- A browser can then request benny's web page using the following URL:
  http://www.example.com/~benny/hobbies.html
  The server replaces * by whatever comes after ~ and appends the rest to give:
  http://www.example.com/users/benny/public_html/hobbies.html

- **Question**: What URL will a browser use to request betty's photo? And what will the server map this to?

- You can have your own configuration file, called .htaccess, which can contain some directives, e.g. ErrorDocument, DirectoryIndex, ...