Универзитет„Св. Кирил и Методиј“ - Скопје
ФАКУЛТЕТ ЗА ИНФОРМАТИЧКИ НАУКИ
И КОМПЈУТЕРСКО ИНЖЕНЕРСТВО

Lesson 2

(X)HTML

Web Design

February 21, 2014

Location of web documents

Web Root directory

- All the files that should be web accessible are placed in the web root directory
- Can have files and other sub-directories
- For Apache Web Server, the default directory is htdocs located in:
  `/var` (*nix base OS-es), or $C : \backslash ProgramFiles \backslash Apache$ (Windows)
- For Microsoft IIS, it is usually $C : \backslash inetpub \backslash wwwroot$

Index file

- This is the default page that should be renedered when accessing the directory
- You can configure the web server for the default file (e.g. usually Index.html, Default.aspx, etc.) or if it is allowed to show the contents of the directory

Structure of a web site

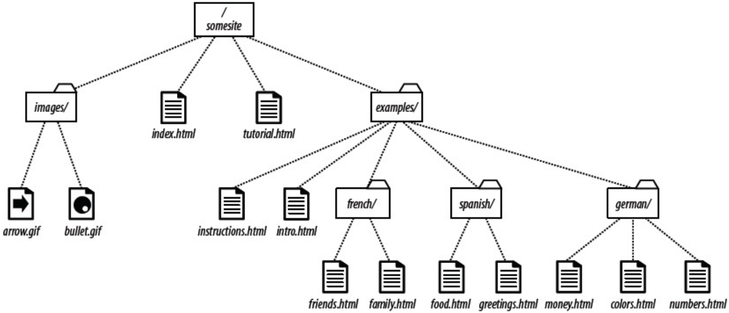

Figure : Example structure of a web site

Relative vs. Absolute Paths

RULE: Always use relative paths!!!

- If you **ARE SURE** that you can not use relative paths, then think about absolute paths (or think about redesining the structure)
- Absolute paths are an issue for web site migration
- You can use them if you are accessing foreign resources (different domain, host, etc)
- Examples:
  - Relative path ("../../daisy.gif")
  - Absolute path ("/var/htdocs/sajtot/sliki/slikata.jpg")
  - Absolute path to a different web site: `http://www.finki.ukim.mk/Content/dataFiles/downloads/IO1920x1080.jpg`

File and directory naming

- You don’t know the hosting environment, or it can be changed
- It is recommended to evade blank character if file/directroy names. Browsers envocode the “space” character as “%20”. Use _ or –.
- Don’t use special characters like ?, %, #, /, :, etc. in file or directory names.
- Use appropriate extensions: .html (or .htm), .gif, .jpg or .jpeg
- Use small caps (*nix based OS-es do not treat small and big caps letters the same)
- File/Directory names should be short, but descriptive :)

Performance

- Web site loading speed heavilly impacts the user experience of the users
- Try to optimize the files (ideal balance between filesize and quality). e.g. The same image can be saved in different resolutions for optimizing bandwith.
- Large files need more time to load.
- **Users with slow or limited connection must not wait for the content!!!**

What is HTML, and XHTML

- HTML is a semistructured language for displaying data in a web page.
- XHTML is a reformulation of HTML for compatibility with the XML standard.
- HTML correct pages don't mean that are XHTML correct.
- XHTML is preferred for beter layout control.

Tags and atributes

HTML document structure is defined in so called **Document Type Definition (DTD)**. Elements are defined in the DTD and can have attributes.

An HTML element is defined by a opening tag, tag attributes, content and closing tag.

If the HTML element doesn't have content, the opening and closing can be made in the same tag.

```
<div id="div1" class="fill" >
    Ova e sodrzirnata vo tagot. Moze da ima i drugi elemementi kako
    <p> Test </p>

    Elementot pocnuva so <div ...> a zavrsuva so </div>
    Atributite na div elementot se:
      id - koe e postaveno na div1
      class - koe e postaveno na fill
    <br /> - e primer na prazen element (bez sodrzina)...
</div>
```

Listing 1: Parts of a HTML element

Rules of well formulated XHTML (1)

- There shouldn't be empty spaces or blank lines before the XML declaration.
- Every element must have opening and closing tag, except if it is an empty element.
- If it's an empty element, it must have appropriate slash before closing (e.g. `<br />`).
- All open and closed tags must be correctly nested. (e.g. `<p>I can <em>fly</em></p>`
- Blank spaces are not allowed between the symbol <and the element name.
- Elements can not have two attributes with the same name.

Rules of well formulated XHTML (2)

- All attribute values should be between quotes (single or double, but used consitently). No spaces are allowed before or after the quotes.
- Comments and processing instructions can not be present in the tags.
- You shouldn't have unrelated <or & symbols in the content.
- The document must have a single element that contains all other elements.
- Attrubute and element names must be in small caps (e.g. `<img>`, `not` `<Img>` `or` `<IMG>`)
- All attributes must have explicit values (e.g. `checked="checked"`)

Rules of well formulated XHTML (3)

- Use **id** instead of **name** as an idenifier.
- Embedded scripts should be contained in CDATA, so they wouldn't be parsed as XML.

```
<script
type="type/javascript">
// <![CDATA[
...  JavaScript goes here...
// ]]>
</script>
```

- Paragraph elements can't be nested (e.g. `<p> <p> Paragraph </p></p>` is not allowed).
- Anchors can't be nested (e.g. `<a> <a> Text </a></a>` is not allowed).
- Forms can't be nested (e.g. `<form> <form> Text </form></form>` is not allowed).
- Labels can't be nested.

Rules of well formulated XHTML (4)

- The **pre** element can't contain **img, object, big, small, sub or sup** elements.
- The **button** element can't contain **input, select, textarea, label, button, form, fieldset, iframe or isindex** elements.

This is a shortened list (The actual list of rules for XHTML is huge).
Use validators and IDE/editor for help with formulating correct XHTML.

Browser rendering modes

Standard and Quirks

Standard mode

- Strict resprecting of the rules defined in the corresponding DTD.
- You can expect that the content will be shown in a consistent way

Quirks mode

- Bad markup (in relation with the DTD) is “tolerated”
- Can be unpredictable!!!

Document Type Definition - DOCTYPE

- Document Type Definition - DTD is a set of rules that define what is allowed in the XHTML document.
- Browsers use the DTD to correctly parse the web page (and use the appropriate rendering mode)
- The DOCTYPE declaration contains an URL to the DTD (rules) file.

DOCTYPE example

- If the browsers detects correct DOCTYPE declaration, it assumes that the HTML code is using the defined DTD and tries to work in standard mode. If the DOCTYPE is invalid, or the HTML doesn't correspond to the DTD, the browser will switch to Quriks mode. This is called **DOCTYPE switching**.
- DOCTYPE declarations are placed in the begining of the documet before the html element.
- Example of HTML 4.01 Strict DTD:

```
 <!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN"
"http://www.w3.org/TR/HTML4.01/strict.dtd">
<html>
...document continues...
```

DOCTYPE for HTML 4 (XHTML 1.0)

Types

- **Strict DTD** - Forbides deprecated elements and attributes (like font or align) .
- **Transitional DTD** - Its prefereable to do not use deprecated elements and attributes. If present, they are “tolerated”.
- **Frameset DTD** - Same as Transitional DTD, but its purpose is for frames instead of single page HTML.
- Validation of (X)HTML:
  - http://validator.w3.org
  - http://htmlhelp.com/tools/validator/

HTML 4 (XHTML 1) vs HTML 5

- HTML 5 is compatible with older versions of (X)HTML
- New elements and attributes:
  - Multimedia support: `<video>`, `<audio>` и `<canvas>`
  - Semantic: `<section>`, `<article>`, `<header>` и `<nav>` and changed/imporved/redefined: `<a>`, `<cite>` и `<menu>`
- Recomendation for disabling certain elements and attributes
- Introduction of APIs for more complicated web applications in HTML 5
- Additions to the HTML DOM model
- Different way of specifying character encoding
- Potential candidate for resolving the problems of different platforms
- More in detail: http://www.w3.org/TR/html5-diff/

Character Encoding

- Encoding used for cyrillic languages is: Cyrillic 1251 (should not be used!)
- Use UTF-8
- If you are using XHTML, you should define character encoding (even if UTF-8 is default)
- If you are using HTML 4, define the encoding in a meta tag

```
<meta http-equiv="content-type" content="text/html; charset=utf-8" />
```

- HTML 5:

```
<meta charset="UTF-8" >
```

Minimal structure of XHTML 1.0 strict

W3C recommendation

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
  "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">

<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="en">

<head>
  <title>An XHTML 1.0 Strict document</title>
  <meta http-equiv="content-type" content="text/html;charset=utf-8" />
</head>
<body>
  <p>... The document content goes here ...</p>
</body>

</html>
```

What can be placed in the head element

The head is the first information that is delivered to the user. Besides the document title, other elements include:

- base – defines default address for all the links
- link – enables linking to external resources
- meta – defines metadata for the document
- script – enables embedding of client scripts
- style – enables embedding of styles

Usage of meta tag

- the meta tag defines metadata for HTML documets
- `<meta>` elements are not rendered on screen
- It should be always in the head element
- It has vast usage, for SEO (Search Engine Optimization), emulating HTTP Response header, additional data, etc.

Atributes of the meta element

- **content** - defines the content of the metadata
- **http-equiv** - enables usage of HTTP headers for the content
- **name** - gives name to the metadata in the content
- **scheme** - specifies the scheme for interpretation of the content
- **charset** – indicates the character set (encoding) of the page (HTML5 ONLY)

Usage of meta

- Description of the content and encoding:

```
<meta http-equiv="Content-Type" content="text/html; charset=ISO-8859-1">
```

- Disable caching the data on the browser

```
<meta http-equiv="pragma" content="no-cache">
```

- Redirect the page after X-seconds

```
<meta http-equiv="refresh" content="3; url=http://www.example.com/newpage.
    html">
```

- Information for search engines

```
<meta name="description" content="The latest movie news">
<meta name="keywords" content="movie, movies, production, genre, review,
    recent">
```

- Expire the web page

```
<meta http-equiv="expires" content="Wed, 21 June 2006 14:25:27 GMT">
```

Questions?