Idiot’s guide to the stuff at the top of your XHTML documents

***Martin Stacey***

The purpose of this document is to tell you what you need to know about the stuff at the top of a web page you never want to think about, for getting pages to be correct XHTML 1.0 Strict or XHTML 1.1 or XHTML5 and to be shown correctly on screen. You can just cut and paste or use templates, but you might get bitten if you leave things out or start mixing different versions of XHTML.

Look elsewhere for technical explanations of what the different types of doctypes actually do, and what the various components of the stuff at the beginning are really for.

The code samples are from the templates provided for *CTEC 1412* that you can use to get started writing web pages without needing to type stuff you’ll always need to include. HTML editors that provide templates don’t all provide all the meta tags etc that I think you need to have.

**Doctypes**

The purpose of a doctype declaration is to tell the browser or other program what sort of a document this is, so that it will process in the right way.

For *CTEC 1412* you need to use **XHTML 1.0 Strict**, or **XHTML 1.1**, or **XHTML5**. (Do NOT use ~~XHTML 1.0 Transitional~~, or any form of HTML without the X.)

XHTML uses XML doctype declarations that tell the browser the name of the doctype and where in the web to find the XML doctype definition that specifies exactly what is and is not legal for a document of that type. This is an XHTML 1.0 Strict doctype declaration.

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"

"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">

This is an HTML5 doctype declaration – designed to be much simpler.

<!DOCTYPE html>

If the **file extension** is ***.xhtml*** the browser will think this is an xml document, look up the doctype definition, and use it. If the **file extension** is ***.html*** the browser will think this is a web page and show it without worrying about the doctype definition.

However Internet Explorer 11 will try to figure out from the doctype declaration what version of IE you wrote your page for, and impersonate an old, out-of-date version that doesn’t understand some advanced features of HTML and CSS.

**XML Prolog**

<?xml version="1.0" encoding="UTF-8"?>

The XML prolog is optional, but strongly encouraged in XHTML documents. When it’s present it *must be the first line* of the document, with absolutely nothing before it. You will need it to get XHMTL5 to be recognized by validator.nu as an XHTML5 document.

**Namespace declarations**

XML documents need to include a declaration of the *namespace* they are using: what the character sequences are that constitute allowable names for elements, attributes and so on. The <html> element gets the namespace declaration for everything within it, i.e. the entire page. HTML including HTML5 doesn’t need this, but you need to put this in for XHTML documents including XHTML5. *xmlns* stands for XML Name Space. The *xmlns* bit is essential; I haven’t seen anything yet for which *xml:lang* or *lang* is actually required.

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

**Meta Elements**

Meta elements are HTML elements in the head of the document, that tell the browser something about what to do with the page.

**Specifying the character set**

You can usually get away without specifying the character set the page is using (which should always be UTF-8). It’s good practice to name it, but you have to do it the right way.

XHTML 1.1 and XHTML 5 want the character set specified in the XML prolog in the first line of the page.

<?xml version="1.0" encoding="UTF-8"?>

You can use a *meta* element to specify the character set the page is using. You can leave out the character encoding, but the W3C validator will give you a warning. XHTML 1.0 and 1.1 want it done this way.

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

This is **disallowed in HTML 5** (and thus in XHTML 5).

If you need to specify the charset in HTML 5 in a *meta* tag, use this one; however you shouldn’t need to. (*meta charset* isn’ t allowed in XHTML 1.0 or 1.1.)

<meta charset="utf-8" />

You should *always* use the **utf-8** character set.

**Correct behaviour on mobile devices**

The viewport meta tag tells the browser about how to handle the relationship between the width of the page and the width of the screen.

<meta name="viewport" content="width=device-width; initial-scale=1.0" />

This tells the browser to render the page for the device width the device says it has, rather than for the actual number of hardware pixels, which is wise for mobile devices that claim fewer pixels than they actually have, so as to match *device independent pixels* more closely to what people actually see. (For instance, an iPhone 5 says it has a 320 pixel width to fit the physical size of the device, even though it has a much higher resolution screen.) For big screens, claimed widths correspond to hardware pixels.

**Getting Internet Explorer to behave**

IE has a nasty habit of trying to figure out which earlier version of IE you wrote your page for, and then impersonating it. The following *meta* element will tell IE to act its age; it isn’t needed for HTML 5 documents which will always be displayed by the most up-to-date version of IE.

<meta http-equiv="X-UA-Compatible" content="IE=edge" />

**HTML5 Template**

All HTML 5 documents need to have to be recognized as HTML 5 by browsers is the very simple doctype declaration <!DOCTYPE html>.

<!DOCTYPE html>

<html lang="en">

<head>

<title>HTML 5 Template</title>

<meta name="viewport" content="width=device-width; initial-scale=1.0" />

</head>

<body>

</body>

</html>

**XHTML 1.0 Strict Template**

XHTML 1.0 Strict documents use this doctype declaration, as well as the namespace declaration in the html element. This template uses a meta element to specify the charset.

<!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>

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

<title>HTML</title>

<meta http-equiv="X-UA-Compatible" content="IE=edge" />

<meta name="viewport" content="width=device-width; initial-scale=1.0" />

</head>

<body>

</body>

</html>

**XHTML 1.1 Template**

XHTML 1.1 documents use this doctype declaration. The *XML* prolog doesn’t seem to be essential for correct behaviour but it saves putting in another declaration of the charset. You need the namespace declaration in the *html* element.

<?xml version="1.0" encoding="UTF-8"?>

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN"

"http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">

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

<head>

<title>XHTML 1.1 Template</title>

<meta http-equiv="X-UA-Compatible" content="IE=edge" />

<meta name="viewport" content="width=device-width; initial-scale=1.0" />

</head>

<body>

</body>

</html>

**XHTML5 Template**

The stuff at the top labels this first as an XML document, and then as an HTML5 document. If it’s processed as an XML document, the namespace declaration in the *html* element is essential.

<?xml version="1.0" encoding="UTF-8"?>

<!DOCTYPE html>

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

<head>

<title>XHTML 5 Template</title>

<meta name="viewport" content="width=device-width; initial-scale=1.0" />

</head>

<body>

</body>

</html>