

Introduction

In this lesson, we are going to review the differences between HTML versions and talk about why it's best to use HTML5 today. In most cases, there is no good reason to use earlier versions of HTML but it's important to know how we got here.

The W3C and the WHATWG

The **W3C (World Wide Web Consortium)** is an international community that develops open standards to ensure the long term growth of the web. They standardize languages like HTML and CSS. The specifications passed through a process of working draft, candidate recommendation, proposal recommendation, and finally, W3C recommendation, in which the specification can be considered final and published for the browser makers and web developers.

The **WHATWG (Web Hypertext Application Technology Working Group)** was founded by individuals of Apple, Mozilla foundation, and Opera software in 2004. At that time HTML4 was already seven years old and apparently no progress was being made in improving the language to tackle the new challenges that were arising, like web applications and the ever more complex websites.

The W3C was concerned with **XHTML 2.0**, a version of XHTML that's was even not backwards compatible with the current HTML. Slowly, the web community noticed that they didn't need to wait for the W3C to do something about the HTML, and that the browser developers and web professionals could have a say in the future of web development. After many years of work and collaboration of the WHATWG and companies like Mozilla, Apple, Microsoft, and Opera, HTML5 became a reality, and its concepts were being published all around the world.

As a result, the W3C has cancelled development on XHTML 2.0 and recently HTML5 has finally been published as a W3C recommendation on October 2013. The web community showed that it can indeed change the fate of HTML.

HTML4

First let's cover HTML4, a version that was published as a recommendation in 1997. Open up the `"html4.html"` file to see an example.

This particular version, **HTML4.01** was published as a recommendation in 1999. The tags are in uppercase here but that was not mandatory. Many people used the tags in uppercase because it helped to set the text apart from regular text. The practice of using lowercase tags became more popular after XHTML in which it is mandatory.

Here you can see that the rules are as loose as the current version of HTML:

- The tags can be in any case
- Tags like `p` can be left open in some cases, although this is not a recommended practice
- Void tags like `BR` do not have closing indications

Also note here that the `<doctype>` tag and the `<meta>` tag, that defines the character set, are much more complex:

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01//EN" "http://www.w3.org/TR/html4/strict.dtd">
[...]
```

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

The `<doctype>` has a link to the **DTD (Document Type Definition)** with all the specifications for the language.

There are more differences from HTML5, but these are the main ones. This version of the language is based on another one called **SGML (Standard Generalized Markup Language)**.

XHTML 1.1

In an effort to simplify the implementation of HTML and its use by browsers and developers as well as making it more easily extensible, the W3C began work on an **XML (Extensible Markup Language)** permutation of HTML: **XHTML (Extensible Hypertext Markup Language)**. The version 1.0 was published as a W3C recommendation in the year 2000.

XHTML had three DTDs:

- The **Strict** DTD didn't consider valid many presentational HTML tags and attributes, and was a good way to learn to write good code. This way it also encouraged the use of CSS which at that time was not widely used.
- The **XHTML Transitional** DTD was a more forgiving version.
- The **XHTML Frameset** DTD was intended to use with frames, something that you almost don't see anymore.

Open up "*xhtml1.html*" file to see an example of a slightly newer version, **XHTML1.1** that was published as a recommendation in 2001. You can see some things are a little different: XML is much stricter with code and that taught developers to write better code. All tags have to be closed – even void tags are self-closing:

```
<br/>
```

There's also the declaration that this is an XML document:

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

The `doctype` is very complex as well:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN" "http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">
```

The HTML tag contains a lot of links to specifications, and the `meta` defining the charset is the same as HTML4's.

Rules

As XHTML is based on XML, the rules of this language are applied:

- All tags have to be written in lower case.
- All attributes must have quoted values.
- All tags must close, even tags that have no content.
- Nesting of tags must be correct.
- For example, we can't nest one tag into another and close them incorrectly:

```
<strong><span></strong></span>
```

- The `style` and `script` tags have to have `type` attributes, specifying `text/css` or `text/javascript`. Also they have to include `CDATA` tags.

XHTML, along with the Web standards movement, brought a lot of good practice to web development but it's also had its problems.

One of them is that XML is supposed to be a strict language and theoretically browsers should throw an error when any part of the code is incorrect. But that, of course, never happens. Also, for the document to be true XML, it would need to be served to the browser as an XML file. Notice that the content attribute in the meta tag is set to `text/html` whereas the correct type for this kind of document should be `application/xhtml+xml`:

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

So, many websites used XHTML without fully committing to it.

XHTML5

There's also an XHTML-like implementation for HTML5 that is called XHTML5 polyglot. Open up `xhtml5.html` file to see an example. There are links in resources section to get you started if you are interested.

We have reviewed the main versions of HTML that were, and still are, very widely used. HTML5 is a combination of the WHATWG's work together with web professionals, browser makers, and later, the W3C. It's a much simpler approach to mark up languages, and it even offers an XHTML implementation if you want to work that way.