!DOCTYPE

Standards Details: Intrinsic feature of SGML/XML
Support Key: [**2**|3|3.2|4] [X1|X1.1] [IE6|M|*N6*|O7]

What is it?
Syntax
Doctype Switching

DTD Examples
Tips & Tricks
Browser Peculiarities

= Index DOT Html by **Brian Wilson** =

Main Index | **Element Tree** | **Element Index** | **HTML Support History**

What is it?

The !DOCTYPE declaration is a top-level tag-like reference known as a Public Text Identifier. It should appear at the very beginning of an HTML/XHTML document in order to identify the content of the document as conforming (theoretically) to a particular HTML DTD specification. Historically, including a DTD was not required by browsers; they just interpreted a document according to what elements and attributes it understood. That has now changed (see the section on DOCTYPE switching below.)

The quoted segment within a !DOCTYPE declaration is called a Formal Public Identifier (FPI.) Every distinct DTD variation will have its own unique FPI string.

Syntax

Typical HTML DOCTYPE statement:

<**!DOCTYPE** HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd">
         (1)   (2)   (3) (4)   (5)   (6)   (7)   (8)

Annotated syntax:

<**!DOCTYPE** [Top Element] [Availability] "[Registration]//[Organization]//[Type] [Label]//[Language]" "[URL]">
            (1)   (2)   (3)   (4)   (5)   (6)   (7)   (8)

Explanation

General : A pair of forward slash characters ("//") is used as delimiters between keyword fields in the FPI declaration.

(1) : [Top Element] - Indicates the top level element type declared in the DTD; for HTML it is <**html**>.

(2) : [Availability] - field indicates whether the identifier is a publicly accessible object (PUBLIC) or a system resource (SYSTEM) such as a local file or URL. HTML/XHTML DTDs are specified by PUBLIC identifiers.

(3) : [Registration] - Indicated by either a plus ("+") or minus ("-"). A plus symbol indicates that the organization name that follows is ISO-registered. A minus sign indicates the organization name is not registered. The IETF and W3C are not registered ISO organizations and thus use a "-".

(4) : [Organization] - This is the "OwnerID" - a unique label indicating the name of the entity or organization responsible for the creation and/or maintenance of the artifact (DTD, etc.) being referenced by the DOCTYPE. The IETF and W3C are the two originating organizations of the official HTML/XHTML DTDs.

(5) : [Type] - This is the "Public Text Class" - the type of object being referenced. There are many different keywords possible here, but in the case of an HTML/XHTML DTD, it is "DTD" - a Document Type Definition.

(6) : [Label] - This is the "Public Text Description" - a unique descriptive name for the public text (DTD) being referenced. If the public text changes for any reason, a new Public Text Description string should be created for it.

(7) : [Language] - This is the "Public Text Language"; the natural language encoding system used in the creation of the referenced object. It is written as an ISO 639 language code (uppercase, two letters.) HTML/XHTML DTDs are usually (always?) written in English ("EN".)

(8) : [URL] - This is the optional explicit URL to the DTD being referenced.

!DOCTYPE Switching

The Need

Now that it is many years since the web exploded in popularity, it now has a huge legacy of existing content on untold millions of web pages all over the planet. These pages were all created at different times, before many of the current web standards were in place. Many of these pages also were written by hand, usually by authors that were not aware of what the correct syntax or standards actually were. Many of these same authors also often had (and have) a tendency to write markup, and test the result in a single browser without validating their code. In the past, these browsers had many behaviors that did not comply with the W3C standards, and many pages on the web rely on some of these buggy or non-standards behaviors. Now that the W3C standards receive more and more attention, and the browsers move to be more compliant with these evolving standards, a big problem has arisen.

Old browsers would just read whatever page that was thrown at it and try to render it as best that it could, including any browser extensions that it understood. If a new browser version were to move to a strictly standards-compliant rendering behavior, *millions* of pages on the web would instantly be rendered "incorrectly" (according to the original intent of the author, misguided though that might be.) Authors, and readers alike would most-likely blame the *browser* as being "buggy"...no browser could afford to do this and expect to survive for very long. But - there was a *REASON* that these standards have come to be accepted - they are gigantic leap forward in web page structure and organization. Do we just continue to accept the mish-mash of the historical quagmire that was the HTML in web pages of old?

The Solution

The Macintosh version of Internet Explorer 5.0 began a trend in March 2000 that all the major browsers now support to some degree. To allow for the huge amount of legacy documents *AND* all current and new standards-based documents, browsers now finally pay attention to the !DOCTYPE declaration at the beginning of an HTML/XHTML document. Depending on which !DOCTYPE is used, the browser will either attempt to render the document according to the W3C standards as close as possible or use the legacy rendering behavior that many older pages rely on.

These two rendering modes are referred to generally as "standards mode" and "quirks mode" (it should be pretty obvious which one makes a browser behave as close as possible to the W3C standards. 8-}) Now - how is a browser to choose which rendering mode to use?

The general rules for choosing:

Most older documents often do not have a !DOCTYPE at all, since its presence was not required in order to render the web page in older browsers. Older documents that DO have a !DOCTYPE present usually reference older DTDs. For !DOCTYPEs that reference known-older DTDs and documents that carry no !DOCTYPE at all, quirks mode is always chosen. When we get to HTML 4, things become interesting - some DOCTYPEs select standards mode, and some select quirks mode.

More information

I purposefully will not get into the details of precisely which DTDs produce quirks mode or standards mode, nor will I get into the details of what those exact differences are. The information on these two points would fill many pages and would doubtlessly change over time (not a headache I want to take on.) This sort of information should and *MUST* come from the browser makers themselves, in my opinion (this is one documentation area where the browser makers must, and thankfully HAVE delivered to the authoring community so far. See the Doctype Switching links below for more information on this topic.

Final caveat:

Although all the major browsers now have these two rendering modes, the browsers still have bugs, behaviors and variations between them on what is included when each mode is used. Test your content as widely as possible.

!DOCTYPE Examples

HTML 2.0

```
<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML 2.0 Level 2//EN">
<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML//EN">
<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML 2.0//EN">
<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML Level 2//EN">
```

HTML 3.0

```
<!DOCTYPE HTML PUBLIC "-//IETF//DTD HTML 3.0//EN">
```

HTML 3.2

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 3.2 Final//EN">
```

HTML 4.01

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

XHTML 1.0

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

XHTML 1.1

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

Tips & Tricks

- Some HTML validation tools will need to have the !DOCTYPE declaration present in order to choose the proper DTD to validate against.
- Using a !DOCTYPE declaration also allows tools (such as SGML browsers or document validators) to view your document. Some of these tools can play vital roles in managing HTML documents.
- SGML tools can also use URLs as FPI references to DTDs. The SYSTEM keyword in the Public Text Identifier indicates that the FPI may be a URL reference. Example:

`<!DOCTYPE HTML SYSTEM "http://www.w3.org/DTD/HTML4-strict.dtd">`

Browser Peculiarities

- [Bug described elsewhere] IE6 and Opera 7.0x: If there is anything before the <!DOCTYPE> declaration, quirks mode will be used. In the case of In an XML document (including XHTML), it is syntactically correct and recommended to have the XML declaration (eg: <?xml version="1.0" ?>) come before any other content. This means that XHTML documents that use the XML declaration, even if they carry a current and/or strict DTD, will use Quirks mode.
- After Mozilla 1.0 (Netscape 7.0) was released, a third specialized mode, called "almost standards" mode was created. Please see Mozilla's documentation on !DOCTYPE switching for more details on the authoring implications.
- Historically, browsers did not care about doctypes, because they were expecting the content of the documents to be HTML and they would only interpret what they could in a single way. That has now changed with the adoption of doctype switching behaviors in all the popular browsers (beginning in the versions indicated above) to allow for full standards adoption while also allowing for legacy behavior with older content.
- **Note:** This is not really a browser peculiarity about doctypes, it is a peculiarity about this particular page in Netscape 4.x. If you are using one of these browsers and most of the above content is invisible, you are seeing an interesting and elusive Netscape rendering bug. A fix that works for me is to expand the browser window (if window is not already maximized and resolution permits) at least wide enough for the entire "Annotated Syntax" line to be visible on a single line. I have been trying to track down the reason for this behavior for a long time but it has been exceedingly hard to reproduce... this fix works on *MY* machine at least. Strange bug.

Related Sites

Official References

http://www.rfc-editor.org/rfc/rfc1866.txt
RFC 1866: The HTML 2.0 specification (plain text)
http://www.w3.org/MarkUp/html-spec
The web version of the HTML 2.0 (RFC 1866) specification
http://www.w3.org/MarkUp/Wilbur/
The HTML 3.2 (Wilbur) recommendation
http://www.w3.org/TR/html4/
The HTML 4.0 recommendation
http://www.w3.org/TR/html401/
The HTML 4.01 recommendation
http://www.w3.org/TR/xhtml1/
The XHTML 1.0 recommendation
http://www.w3.org/TR/xhtml11/
The XHTML 1.1 recommendation

!DOCTYPE Switching links

Official References

http://www.mozilla.org/docs/web-developer/quirks/doctypes.html
Mozilla's documentation and table on which DTDs trigger which mode (by David Baron)
http://dbaron.org/mozilla/quirklist
David Baron's list of differences between quirks and standards mode (probably can be called "official")
http://msdn.microsoft.com/library/en-us/dnie60/html/cssenhancements.asp?frame=true
Internet Explorer's "Enhancements in IE6": documentation and mode/behavior differences
http://www.opera.com/docs/specs/doctype/
Opera's documentation and mode/behavior differences

Other References

http://gutfeldt.ch/matthias/articles/doctypeswitch.html
Matthias Gutfeldt's "Doctype switching and standards compliance: An overview"; a great up-to-date discussion of the topic with many additional links for more information.
http://www.hut.fi/u/hsivonen/doctype.html
Henri Sivonen's "Activating the right layout mode using the Doctype declaration"; another great discussion of the topic.

Boring Copyright Stuff...