HTML: The Markup Language

3. Documents

This section defines the term document, and provides additional details related to the definition of that term. It is divided into the following parts:

1. The HTML language and HTML and XML syntaxes
2. The HTML namespace and MIME types
3. Conformant documents
4. Case insensitivity in tag names and attribute names

3.1. The HTML language and HTML and XML syntaxes

The term **document** is used in this reference to mean an instance of the HTML language.

The **HTML language** is the language described in this reference; it is an abstract language that applications can potentially represent in memory in any number of possible ways, and that can be transmitted using any number of possible concrete syntaxes.

This reference describes two particular concrete syntaxes for the HTML language: One syntax, which is referred to throughout this reference as the HTML syntax, and another syntax, which is referred to throughout this reference as the XML syntax. Web browsers typically implement two separate parsers for processing documents: an **HTML parser** which is invoked when processing documents in the HTML syntax, and an **XML parser** which is invoked when processing documents in the XML syntax.

The **HTML syntax** is the syntax described in the HTML syntax section of this reference.

The **XML syntax** is defined by rules in the [XML] specification and in the [Namespaces In XML] specification; any syntax-level requirements for documents in the XML syntax described in this reference are intended to be the same as those defined in the XML specification.

3.2. The HTML namespace and MIME types

The **HTML namespace** is defined as `http://www.w3.org/1999/xhtml`. The HTML namespace is the namespace both for documents in the HTML syntax and for documents in the XML syntax.   **jump**

Documents that are served with the `text/html` MIME type must match the descriptions in this reference for characteristics of documents in the HTML syntax.

Documents that have an HTML namespace declaration and that are served with an XML MIME type such as `text/xml`, `application/xml`, or `application/xhtml+xml` must match the descriptions in this reference for characteristics of documents in the XML syntax.

3.3. Conformant documents

There are two types of conformant documents:

- conformant documents in the HTML syntax
- conformant documents in the XML syntax

3.3.1. Conformant documents in the HTML syntax

A conformant **document in the HTML syntax** must consist of the following parts, in the following order:

1. Optionally, a single U+FEFF BYTE ORDER MARK (BOM) character.
2. Any number of comments and space characters.
3. A doctype.
4. Any number of comments and space characters.
5. An html element, with its attributes (if any) and its contents (if any).

> **Note:** The start tag and end tag of the html element can be omitted—as well as, possibly, the start tags and end tags of certain descendants of the html element—in which case the start tag and end tag are considered to be implied.

6. Any number of comments and space characters.

Documents in the HTML syntax must match the syntax described in the HTML syntax section of this reference.

3.3.1.1. Implied start tags and end tags

In documents in the HTML syntax, the start tags and end tags of the html element and particular descendants of the html element can be omitted. In cases where tag omission of those particular elements occurs, the document can still be considered, conceptually, to contain the elements—but with their start tags and end tags *implied*.

The following is an example of a document with implied start tags and end tags for the html, head, and body elements. Note that it is nevertheless a complete, valid

jump

document.

```
<!DOCTYPE html>
<title>A relatively minimal HTML document</title>
<p>Hello World!</p>
```

The DOM tree constructed from that example by a conformant UA would look like this:

```
├DOCTYPE: html
└HTML
  ├HEAD
  │ ├TITLE
  │ │ └#text: A relatively minimal HTML document
  │ └#text: ↵
  └BODY
    └P
      └#text: Hello World!
```

Note that the DOM tree includes the html, head, and body elements whose start tags and end tags are implied in the document.

3.3.2. Conformant documents in the XML syntax

A conformant **document in the XML syntax** must be a **namespace-well-formed** XML document, as defined in the [Namespaces in XML] specification, and its **root element** must be an html element.

> **Note:** Documents in the XML syntax must not make use of any features of the HTML syntax that do not follow XML well-formedness constraints (for example, documents in the XML syntax must not use unquoted attribute value syntax and must not omit tags).

3.4. Case insensitivity in tag names and attribute names

In documents in the HTML syntax:

- Tag names for HTML elements may be written with any mix of lowercase and uppercase letters that are a case-insensitive match for the names of the elements given in the HTML elements section of this document; that is, tag names are case-insensitive.
- Attribute names for HTML elements may be written with any mix of lowercase and uppercase letters that are a case-insensitive match for the names of the attributes given in the HTML elements section of this

jump

document; that is, attribute names are case-insensitive.

In documents in the XML syntax:

- Tag names for HTML elements must exactly match the names of the elements given in the HTML elements section of this document; that is, tag names are case-sensitive.
- Attribute names for HTML elements must exactly match the names of the attributes given in the HTML elements section of this document; that is, attribute names are case-sensitive.

« terminology   syntax »

jump