

HTML/XHTML - DOCTYPE

Autor: Martina Rüter

Dokumenttypen - Grammatik-Regeln für den Browser

SGML (Standardised General Markup Language) => **HTML** (HyperText Markup Language)

XML (eXtensible Markup Language) =>
XHTML (eXtensible Hypertext Markup Language)

Abb. 1: Übersicht der Auszeichnungssprachen: HTML ist aus SGML entstanden, während XHTML ein Teil von XML ist.

HTML 4.01 Strict

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

HTML 4.01 Transitional

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

HTML 4.01 Frameset

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

XHTML 1.0 Strict

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

XHTML 1.0 Transitional

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
```

XHTML 1.0 Frameset

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD XHTML 1.0 Frameset//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-frameset.dtd">
```

XHTML 1.1 DTD

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

Abb. 2: Die wichtigsten DOCTYPs für HTML 4.01 und XHTML 1.0 sowie XHTML 1.1 DTD

Dokumenttypen (DOCTYPE)

HTML und XHTML (Abb. 1) sind Auszeichnungssprachen. Sie werden vom Anzeigeprogramm, dem Browser (z. B. Internet Explorer, Firefox, Mozilla, Opera etc.) am Bildschirm dargestellt. Beide Sprachen haben eine bewegte Vergangenheit, die sich am Versionsstand (HTML 4.01, XHTML 1.1) ablesen lässt.

Damit der Browser die jeweils richtigen Regeln zur Darstellung des Dokuments verwendet, sollte im Quelltext der Dokumenttyp (DOCTYPE) genannt werden.

Die Dokumenttyp-Definition (Abb. 2) erfolgt noch vor dem <HTML>.

Mit dieser Definition werden die „Grammatik-Regeln“ der jeweils verwendeten Version und (in geringem Maße) die Abweichungen vom Standard definiert. Bei den Abweichungen unterscheidet man:

Strict

Der Quelltext liegt im korrekten Format des jeweiligen Versionsstands vor.

Transitional (engl. transition = Übergang)

Es können Abweichungen zum Standard enthalten sein.

Frameset

Der Quelltext definiert ein Frameset.

Dokumenttypen - Grammatik-Regeln für den Browser

Die Variante „Transitional“ wurde geschaffen, da viele Elemente und Attribute in den neueren Versionen (HTML 4.01) nicht mehr enthalten sind (Elemente: z. B.: FONT, CENTER, U; Attribute: z. B.: TEXT, LINK, ALINK, COLOR, BGCOLOR, BACKGROUND, FACE, SIZE, BORDER).

Da diese jedoch weit verbreitet sind, wurde eine Übergangsvariante geschaffen, in der die nicht mehr zu verwendenden Elemente und Attribute noch erlaubt sind.

Eine vollständige Übersicht der „deprecated“ Elemente und Attribute führt das W3C (World Wide Web Consortium): <http://www.w3.org/TR/html4/index/attributes.html>.

Formaler Aufbau der DOCTYPES:

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

HTML: Typ des SGML/XML Dokuments

PUBLIC: Public Identifier

-: Minuszeichen = Organisation, die diese DTD veröffentlicht hat nicht bei der ISO registriert ist (ein Pluszeichen würde eine Registrierung bei der ISO bedeuten)

W3C: Identifikation des Eigentümers der DTD (W3C = World Wide Web Consortium)

DTD: Art des Objekts (DTD = Document Type Definition)

XHTML 1.0 Strict: Bezeichnung des Objekts (XHTML), seine Version (1.0) und die Abart (Strict).

EN: natürliche Sprache (hier: englisch), die dem Dokument zu Grunde liegt.

Es schließt sich die URL zum Regelwerk an.

Überprüfen einer Webseite auf Verwendung überholter Elemente und Attribute:

<http://www.bitvtest.de/werkzeugliste.html#bookmarklets>

Mit den Bookmarklets können zwei Lesezeichen angelegt werden. Rufen Sie eine beliebige Internetseite auf und betätigen Sie die Lesezeichen. Sie funktionieren jedoch nur unter Firefox/Mozilla. !

Ein fehlender oder falscher DOCTYPE führt dazu, dass ein moderner Browser das Dokument nicht im „Standard-Modus“, sondern im „**Quirks-Modus**“ anzeigt. Dieser dient in der Abwärtskompatibilität. Er soll sicherstellen, dass veralteter oder ungültiger HTML-Code weiterhin korrekt angezeigt wird. Dabei werden u. a. Darstellungsfehler alter Browser simuliert.

Die DOCTYPE-Angaben sind beim W3C unter: <http://www.w3.org/QA/2002/04/valid-dtd-list.html> nachzulesen.

Ob Sie den DOCTYPE für HTML oder XHTML verwenden ist eigentlich egal. Bei XHTML sind allerdings die folgenden Regeln zu beachten:

- Alle Elemente und Attribute müssen klein geschrieben werden.
- Alle Werte der Attribute müssen in Anführungszeichen stehen.
- Alle Befehle (TAGs) müssen geschlossen werden, auch solche, die allein stehen. So muss ein Zeilenumbruch in XHTML formal wie folgt aussehen: `
`

Zeichensätze

Eine zweite wichtige Angabe ist, die des korrekten Zeichensatzes im Kopfbereich (HEAD) des Dokuments.

Der Zeichensatz für westeuropäische Sprache lautet:

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

Weitere Zeichensätze sind dem SelfHTML von Stefan Münz zu entnehmen: <http://intra.fidion.de/man/selfhtml70/////thb.htm>.