

1. Wohlgeformtheit



<https://wiki.selfhtml.org/wiki/XML/Regeln/Wohlgeformtheit>

Ein XML-Dokument ist **wohlgeformt**, wenn es den grundlegenden Syntaxregeln von XML folgt. Dabei spielt es keine Rolle, ob eine DTD definiert ist. Die wichtigsten Regeln sind:

Grundregeln für Wohlgeformtheit:

- **Prolog:** Das Dokument sollte mit der XML-Deklaration beginnen, z. B.:

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

Diese Deklaration definiert die XML-Version und die Zeichencodierung.

- **Ein einziges Wurzelement:** Das gesamte XML-Dokument muss genau ein Wurzelement enthalten, das alle anderen Elemente umfasst.

Richtig:

```
<Bibliothek>  
  <Buch></Buch>  
</Bibliothek>
```

Falsch:

```
<Buch></Buch>  
<Buch></Buch>
```

Hier fehlt das Wurzelement bzw. wird doppelt verwendet (falls im Dokument das Wurzelement ist)

- **Start- und End-Tags:** Jedes Start-Tag (<Element>) muss ein passendes End-Tag (</Element>) haben.

Richtig:

```
<Titel>XML Grundlagen</Titel>
```

Falsch:

```
<Titel>XML Grundlagen
```

- **Verschachtelung:** Elemente dürfen sich nicht überlappen.

Falsch:

```
<a><b></a></b>
```

- **Attribute:** Alle Attribute müssen Werte in Anführungszeichen haben.

Richtig:

```
<Verlag ort="Berlin"></Verlag>
```

Falsch:

```
<Verlag ort=Berlin></Verlag>
```

2. Gültigkeit

<https://wiki.selfhtml.org/wiki/XML/DTD>

https://wiki.selfhtml.org/wiki/XML/DTD/Elemente_und_Verschachtelungsregeln#Beispiel_einer_g.C3.BCltigen_Anwendung

Ein XML-Dokument ist **gültig**, wenn:

1. Es wohlgeformt ist.
2. Es den in einer DTD definierten Regeln entspricht.

Prüfung auf Gültigkeit:

- Die DTD gibt die erlaubte Struktur des Dokuments vor. Das XML-Dokument muss sich daran halten.
- Jedes in der DTD definierte Element muss korrekt verwendet werden, einschließlich der Attribute, Reihenfolge und Häufigkeit der Elemente.

3. Wichtige DTD-Konstrukte

DTD (Document Type Definition) definiert die Struktur eines XML-Dokuments, einschließlich erlaubter Elemente, Attribute und Werte.

Häufigkeitsoperatoren

- **+** (**mindestens einmal**): Das Element muss mindestens einmal vorkommen.

```
<!ELEMENT Bibliothek (Buch+)>
```

- `<Bibliothek>` muss mindestens ein `<Buch>` enthalten.

- ***** (**beliebig oft, auch nullmal**): Das Element kann beliebig oft auftreten, auch gar nicht.

```
<!ELEMENT Zusatzinfo (Seitenzahl*, Untertitel*)>
```

- `<Seitenzahl>` und `<Untertitel>` dürfen beliebig oft oder gar nicht vorkommen.

- **?** (**null- oder einmal**): Das Element ist optional und darf höchstens einmal auftreten.

```
<!ELEMENT Seitenzahl (#PCDATA)?>
```

- `<Seitenzahl>` ist optional und darf höchstens einmal vorkommen.

- **|** (**Alternativen**): Erlaubt die Wahl zwischen mehreren Elementen.

```
<!ELEMENT Zusatzinfo (Seitenzahl | Untertitel)>
```

- `<Zusatzinfo>` muss entweder ein `<Seitenzahl>` oder ein `<Untertitel>` enthalten, aber nicht beide.

Text und Zeicheninhalte

- **#PCDATA (Parsed Character Data):**

Erlaubt Textinhalt in einem Element, der von XML-Parsern interpretiert werden kann.

Beispiel:

```
<!ELEMENT Titel (#PCDATA)>
```

- `<Titel>` kann Text enthalten, z. B.:

```
<Titel>XML für Anfänger</Titel>
```

- **CDATA (Character Data):**

CDATA-Abschnitte enthalten rohen Text, der von XML-Parsern nicht interpretiert wird.

Beispiel:

```
<code><![CDATA[<tag>Inhalt</tag>]]></code>
```

- Alles innerhalb von `<![CDATA[` und `]]>` wird wörtlich behandelt.

Attribute

- Attribute werden mit `<!ATTLIST>` definiert.

Beispiel:

```
<!ATTLIST Verlag ort CDATA #REQUIRED>
```

- `<Verlag>` muss ein Attribut `ort` enthalten, das beliebigen Text (CDATA) akzeptiert.

Attribute-Anforderungen:

- **#REQUIRED** : Das Attribut muss vorhanden sein.
- **#IMPLIED** : Das Attribut ist optional.
- **#FIXED** : Das Attribut hat einen festen Wert.

4. Regeln für die DTD-Prüfung

- **Reihenfolge der Elemente:** Die Reihenfolge der Elemente im XML-Dokument muss exakt der DTD entsprechen, wenn keine Alternativen (`|`) angegeben sind.
- **Wiederholung:** Die Häufigkeit muss mit den DTD-Regeln (`+` , `*` , `?`) übereinstimmen.
- **Pflichtattribut:** Attribute, die mit `#REQUIRED` markiert sind, müssen vorhanden sein.
- **Datentypen:** Textinhalt muss mit `#PCDATA` oder `CDATA` übereinstimmen.

Ein XML-Dokument ist **ungültig**, wenn es diese Regeln verletzt, auch wenn es wohlgeformt ist.

