

XML und XSL

Entities

© eden market

Autor: Norman Lahme

Gliederung

1. Einleitung

2. Interne allgemeine Entities
3. Externe geparste allgemeine Entities
4. Nicht geparste Entities
5. Interne und externe Parameterentities
6. Grenzen des Markup
7. Externe Bezeichner

Mit Entities können ...

- ... XML- Dokumente auf mehrere Dateien aufgeteilt werden
 - Arbeitsteilung, Wiederverwendung
- ... Multimedia- Objekte in XML- Dokumente eingebunden werden
- ... DTDs auf mehrere Dateien aufgeteilt werden
 - Arbeitsteilung, Wiederverwendung

Einleitung

- Entities bestehen aus einem Entity- Namen und einem Entity- Inhalt
- Jeder Verweis auf ein Entity (*Entity-Referenz*) wird vom Parser durch den Entity-Inhalt ersetzt

Beispiel

```
<!DOCTYPE Buch [  
<!ELEMENT Buch (#PCDATA)>  
<!ENTITY dtd "Document Type Definition">  
>  
<Buch>Durch eine &dtd; wird ...</Buch>
```

Entity- Arten

Allgemeine vs. Parameter- Entities

→ Wo werden sie verwendet?

- Allgemeine Entities werden in XML- Dokumenten *verwendet*,
- Parameter- Entities in einer DTD

Entity- Arten

Externe vs. interne Entities

→ Wo steht der Inhalt?

- Bei internen Entities steht deren Inhalt in den Entity- Deklarationen
- Externe Entities erhalten ihren Inhalt von anderer Stelle
 - z.B. Datei, Webseite, Datenbank

Entity- Arten

Geparste vs. nicht geparste Entities

→ Text oder Multimedia?

- Ist der Inhalt eines Entities ein zu parsender XML- Text, so heißt das Entity *geparstes Entity* (laut XML-Spezifikation)
- Enthält ein Entity Daten, die nicht geparst werden sollen, so heißt es *nicht geparstes Entity*

Gliederung

1. Einleitung
- 2. Interne allgemeine Entities**
3. Externe geparste allgemeine Entities
4. Nicht geparste Entities
5. Interne und externe Parameterentities
6. Grenzen des Markup
7. Externe Bezeichner

Interne allgemeine Entities

- Sie sind der einfachste Entity- Typ
- Es handelt sich dabei um Abkürzungen
- Sie sind immer geparste Entities

Beispiel

```
<!DOCTYPE Buch [  
<!ELEMENT Buch (#PCDATA)>  
<!ENTITY dtd "Document Type Definition">  
>  
<Buch>Durch eine &dtd; wird ...</Buch>
```

Interne allgemeine Entities

Erlaubte Vorkommnisse für Referenzen:

- In XML- Dokumenten
 - in einem Elementinhalt oder als Attributwert
- In einer DTD
 - Im Inhalt eines anderen allgemeinen Entities
 - Vorwärts- Referenzen sind dabei erlaubt
 - In einem Default- Attributwert
 - Vorwärts- Referenzen sind dabei *nicht* erlaubt

Beispiel

```
<!DOCTYPE Buch [  
<!ELEMENT Buch (#PCDATA)>  
<!ENTITY dtd "Document Type &dtd;">  
<!ENTITY def "Definition">  
>  
<Buch>Durch eine &dtd; wird ...</Buch>
```

Beispiel

```
<!DOCTYPE Buch [  
<!ELEMENT Buch (#PCDATA)>  
<!ENTITY nl "Norman Lahme">  
<!ATTLIST Buch autor CDATA "&nl;">  
>  
<Buch>Blabla...</Buch>
```

Beispiel

```
<!DOCTYPE Buch [  
<!ELEMENT Buch (#PCDATA)>  
<!-- Vorwärtsreferenz ist nicht erlaubt -->  
<!ATTLIST Buch autor CDATA "&nl;">  
<!ENTITY nl "Norman Lahme">  
>  
<Buch>Blabla...</Buch>
```

Beispiel

```
<!DOCTYPE Buch [  
<!ELEMENT Buch (#PCDATA)>  
<!ENTITY dtd "Document Type Definition">  
<!-- Referenz in der DTD ist verboten -->  
&dtd;  
>  
<Buch>blabla ...</Buch>
```

Übung 1 Aufgabe 1

Gliederung

1. Einleitung
2. Interne allgemeine Entities
- 3. Externe geparste allgemeine Entities**
4. Nicht geparste Entities
5. Interne und externe Parameterentities
6. Grenzen des Markup
7. Externe Bezeichner

Externe geparste allgemeine Entities

- Der Inhalt eines externen Entities wird durch einen sog. *externen Bezeichner* angesprochen
- Dieser besteht i.d.R. aus dem Schlüsselwort SYSTEM gefolgt von einer URL (s. später)

Externe geparste allgemeine Entities

Erlaubte Vorkommnisse für Referenzen:

- In XML- Dokumenten
 - in einem Elementinhalt
- In einer DTD
 - Im Inhalt eines anderen allgemeinen Entities
 - Vorwärts- Referenzen sind dabei erlaubt

Beispiel

XML-Dokument

```
<!DOCTYPE Buch [  
<!ELEMENT Buch  
  (#PCDATA)>  
<!ENTITY content  
SYSTEM "content.xml">  
>  
<Buch>&content;</Buch>
```

Datei content.xml

irgend ein Inhalt ...

Beispiel

XML-Dokument

```
<!DOCTYPE Buch [  
<!ELEMENT Buch (#PCDATA)>  
<!ENTITY def SYSTEM "content.xml">  
<!ENTITY dtd "Document Type &def;">  
>  
<Buch>Durch eine &dtd; wird ...  
</Buch>
```

Datei content.xml

Definition

Übung 1 Aufgabe 2

Gliederung

1. Einleitung
2. Interne allgemeine Entities
3. Externe geparste allgemeine Entities
- 4. Nicht geparste Entities**
5. Interne und externe Parameterentities
6. Grenzen des Markup
7. Externe Bezeichner

Nicht geparste Entities

- Nicht geparste Entities sind immer extern und allgemein
 - extern, da es unmöglich ist, Nicht-XML-Informationen in XML-Deklarationen auszudrücken
 - allgemein, da es verboten ist, *Daten* in XML-DTDs auszudrücken

Nicht geparste Entities

- Syntax: wie bei externen geparsten allgemeinen Entities, wobei folgendes angehängt wird:
 - Schlüsselwort NDATA und Namen einer deklarierten *Notation*
 - Diese Notationsdeklaration gibt der Prozessor beim Parsen an die Applikation weiter, die daraus entnimmt, wie sie das Entity verarbeiten soll

Nicht geparste Entities

Erlaubte Vorkommnisse für Referenzen:

- In XML- Dokumenten
 - In Entity- Attributwerten
- In einer DTD
 - In Entity- Defaultattributwerten
 - Vorwärts- Referenzen sind erlaubt

Beispiel

```
<!DOCTYPE Buch [  
<!ELEMENT Buch (#PCDATA)>  
<!NOTATION GIF SYSTEM "gif.doc">  
<!ENTITY picA SYSTEM "coverA.gif" NDATA GIF>  
<!ENTITY picB SYSTEM "coverB.gif" NDATA GIF>  
<!ATTLIST Buch pic ENTITY "picA">  
]>  
<Buch pic="picB">...</Buch>
```

Gliederung

1. Einleitung
2. Interne allgemeine Entities
3. Externe geparste allgemeine Entities
4. Nicht geparste Entities
- 5. Interne und externe Parameterentities**
6. Grenzen des Markup
7. Externe Bezeichner

Interne & externe Parameterentities

- Syntax: wie bei den allgemeinen geparsten Entities, mit folgendem Unterschied
 - Deklaration: Prozentzeichen und ein Leerzeichen vor dem Entitynamen
 - Referenz: Prozentzeichen statt des Ampersands
- Parameter-Entities sind immer geparst

Interne & externe Parameterentities

Erlaubte Vorkommnisse für Referenzen:

- In XML- Dokumenten
 - gar nicht, da Parameter- Entity
- In einer DTD
 - Vorwärts- Referenzen sind nicht erlaubt

Beispiel: Internes Parameterentity

```
<!DOCTYPE Buch [  
<!ENTITY % buch-entity "<!ELEMENT Buch  
(#PCDATA)>">  
%buch-entity;  
>  
<Buch>blabla</Buch>
```

↓ entspricht

```
<!DOCTYPE Buch [  
<!ELEMENT Buch (#PCDATA)>  
>  
<Buch>blabla</Buch>
```

Beispiel: Internes Parameterentity

```
<!DOCTYPE Buch [  
<!-- Vorwärtsreferenz nicht erlaubt -->  
%buch-entity;  
<!ENTITY % buch-entity "<!ELEMENT Buch  
                                (#PCDATA)>">  
]  
<Buch>blabla</Buch>
```

Beispiel: Externes Parameterentity

```
<!DOCTYPE Buch [  
<!ENTITY % buch-entity  
SYSTEM "content.dtd">  
%buch-entity;  

```

Datei content.dtd

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

↓ entspricht

```
<!DOCTYPE Buch [  
<!ELEMENT Buch  
(#PCDATA)>  
<Buch>blabla</Buch>
```

Übung 1 Aufgabe 3

Gliederung

1. Einleitung
2. Interne allgemeine Entities
3. Externe geparste allgemeine Entities
4. Nicht geparste Entities
5. Interne und externe Parameterentities
- 6. Grenzen des Markup**
7. Externe Bezeichner

Grenzen des Markup

- Markup darf sich nicht über Entity- Grenzen hinaus erstrecken – sonst ist ein Dokument nicht wohlgeformt
- So darf z.B. ein Element nicht in einem Entity beginnen und in einem anderen enden
- Zum Markup gehören:
 - Elemente, Deklarationen, Kommentare, Verarbeitungsanweisungen und Entity-Referenzen

Beispiel

```
<!DOCTYPE Test[
<!ELEMENT Test (#PCDATA|em)*>
<!ELEMENT em (#PCDATA)>
<!ENTITY illegal1 "Gleich beginnt ein Tag <em>, jedoch">
<!ENTITY illegal2 "endet es erst hier</em>">
<!--bisher ist noch nichts fehlerhaft -->
]><Test>
<!-- folgende Referenzen sind fehlerhaft -->
&illegal1;    <!--Start-Tag ohne Ende-Tag -->
&illegal2;    <!--Ende-Tag ohne Start-Tag -->
</Test>
```

Gliederung

1. Einleitung
2. Interne allgemeine Entities
3. Externe geparste allgemeine Entities
4. Nicht geparste Entities
5. Interne und externe Parameterentities
6. Grenzen des Markup
- 7. Externe Bezeichner**

Externe Bezeichner

- Externe Bezeichner verweisen auf Informationen außerhalb des Entities, in dem sie auftreten
- Arten
 - Systembezeichner
 - Öffentliche Bezeichner

Systembezeichner

- Sie verwenden URIs, um auf die Position der Informationsobjektes zu verweisen
 - Absolute URI:
`<!ENTITY javaPage SYSTEM "http://java.sun.com/index.html">`
 - Relative URI:
`<!ENTITY logoPic SYSTEM "logo.gif">`

Öffentliche Bezeichner

- Sie verwenden einen öffentlich deklarierten Namen, der von einer XML-Software in eine URI übersetzt wird
- Kennt die XML-Software diesen Namen nicht, so verwendet sie den Systembezeichner

Öffentliche Bezeichner

Zweck

- In XML- Kreisen beliebte Entities können so von der XML- Software bereitgestellt werden
 - und müssen nicht jedes Mal von einem Server heruntergeladen werden

Öffentliche Bezeichner

Beispiel

```
<!DOCTYPE MEMO PUBLIC  
    "-//SGMLSOURCE//DTD MEMO//EN"  
    "http://www.sgmlsource.com/dtds/memo.dtd">
```