

XML

Veerle Ongenae

Overzicht Onderwerpen XML

- XML

- XML en namespaces

- XML Technologieën

- Structureren en datatypes: DTD en XML Schema
 - Presentatie: CSS en XSL
 - Manipulatie (o.a. transformatie): XPath, DOM en SAX
 - Diverse

XML en namespaces

- Wat is XML?
- Structuur XML-document
- Welgevormd XML-document
- Namespaces

Wat is XML?

- eXtensible Markup Language
 - Standaard van World WideWeb Consortium
 - <http://www.w3c.org>
- Meta-taal
 - Taal om “markup”-talen te definiëren
- Markup-taal
 - Tekst opmaken m.b.v. tags (labels)

Doel?

- XML geeft
 - Een logische structuur
 - En betekenis
 - Aan tekstuele gegevens
- Syntax lijkt op HTML

Wat is XML?

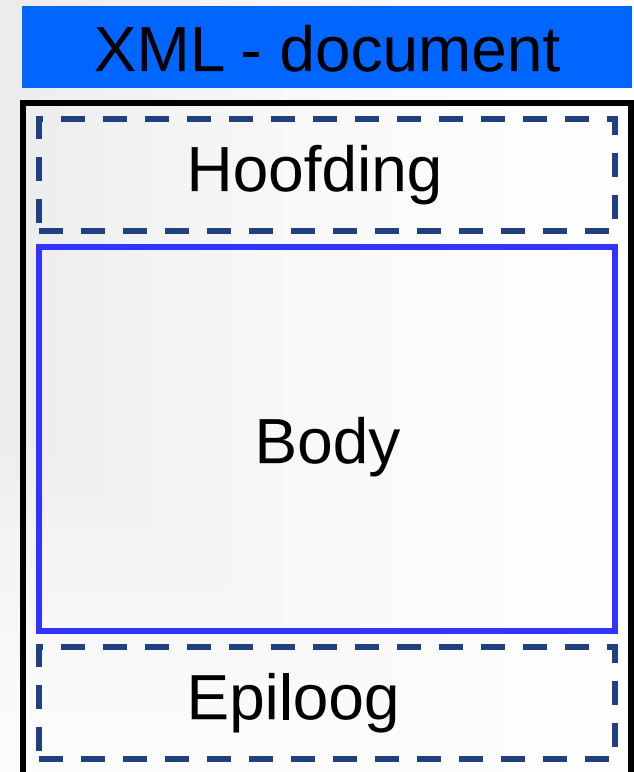
- XML ↔ HTML
 - Geen opmaak
 - Enkel structuur van data
 - “Eigen” tags
- Voorbeeld
 - boeken.xml
 - boeken.html

XML en namespaces

- Wat is XML?
- Structuur XML-document
- Welgevormd XML-document
- Namespaces

Structuur XML-document

- XML-document
 - Tekst
 - Bestaat uit
 - Hoofding (optioneel)
 - Body of corpus
 - Epiloog (optioneel)



Body XML-document

- Body bestaat uit één element
 - Basiselement genoemd (root element)

```
<?xml version="1.0"?>
```

```
<books>  
  <book>  
    <isbn>0-596-00058-8</isbn>  
    ...  
  </book>  
</books>
```

Hoofding

Body

Basiselement

Elementen

- Een element kan bevatten
 - Tekst
 - Elementen
 - Entiteiten
 - CDATA-stukken
 - Commentaar
 - Processing Instructions
- Een element kan attributen hebben

Elementen

- Beschrijven inhoud
- Een aantal voorbeelden

- Eenvoudige inhoud

`<title>XML in a Nutshell</title>`

- Container voor andere elementen

`<book>`

`<isbn>0-596-00058-8</isbn>`

`<title>XML in a Nutshell</title>`

`<author>Harold, Elliotte Rusty</author>`

`</book>`

Elementen

- Leeg element, container voor attributen

`<query isbn="0-596-00058-8" zipcode="9040"/>`

- `<reply></reply>` kan verkort worden tot `<reply />`

- Container voor andere elementen en attributen

`<figuur use="optioneel">`

`<bestand>tekening.gif</bestand>`

`<tooltip>Een tekening</tooltip>`

`</figuur>`

Structuur van een element

- Begin:
<naamElement lijstAttributen>
- Lijst van attributen (0 of meerdere)
naamAttribuut="waardeAttribuut"
- Inhoud van het element
- Einde:
</naamElement>
- Lege elementen:
<naamElement lijstAttributen />

Elementen: Naamgeving

- Naam van een element bestaat uit letters, cijfers en andere leettertekens
- Eerste karakter is geen cijfer of leesteken
- Naam begint niet met “xml” (noch kleine noch grote letters)
- Geen spaties in naam
- Geen “:” in naam

Body XML-document

- Body bestaat uit één element
 - Basiselement genoemd (root element)
- Een element kan bevatten
 - Tekst
 - Elementen
 - Entiteiten
 - CDATA-stukken
 - Commentaar
 - Processing Instructions
- Een element kan attributen hebben

Attributen

- Extra informatie over een element
- Volgorde is onbelangrijk
- Uniek in element
- Tussen enkele of dubbele aanhalingstekens
- Informatie die geen deel uit maakt van de data van een element

<file use="optional">computer.gif</file>

Elementen Versus Attributen

- Geen officiële regels
- Algemene consensus
 - Elementen voor data
 - Attributen voor informatie over data
- Redenen om geen attributen te gebruiken
 - Geen meervoudige waarden
 - Niet eenvoudig uitbreidbaar, kunnen niet opgedeeld worden in onderdelen
 - Moeilijker te manipuleren in programma's

Body XML-document

- Body bestaat uit één element
 - Basiselement genoemd (root element)
- Een element kan bevatten
 - Tekst
 - Elementen
 - Entiteiten
 - CDATA-stukken
 - Commentaar
 - Processing Instructions
- Een element kan attributen hebben

Entiteiten

- De tekens <, &, >, “, ‘ kunnen niet voorkomen in tekst (als inhoud van een element of attribuut)
 - Indien nodig, gebruik entiteiten
- Voorgedefinieerde:
 - < (<)
 - & (&)
 - > (>)
 - " (“)
 - ' (‘)

Entiteiten

- Zelfgedeclareerde (in DTD)

```
<!ENTITY public "<literal  
role='keyword'>public</literal>">
```

```
<!ENTITY servlets SYSTEM  
"webtech\docbook\servlets.docbook">
```

- Gebruik

```
&public;  
&servlets;
```

Entiteiten: structuur

- Syntax

<!ENTITY naamEntiteit “waardeEntiteit”>

Of

<!ENTITY naamEntiteit SYSTEM “PAD”>

- Het bestand PAD bevat dan de declaratie van de entiteit
- Ook mogelijk
 - Publiek declareren (PUBLIC ipv SYSTEM)
 - URL op het web

Body XML-document

- Body bestaat uit één element
 - Basiselement genoemd (root element)
- Een element kan bevatten
 - Tekst
 - Elementen
 - Entiteiten
 - CDATA-stukken
 - Commentaar
 - Processing Instructions
- Een element kan attributen hebben

CDATA-stukken

- De inhoud van een CDATA-stuk wordt niet beschouwd als “markup”
- Inhoud van een CDATA-stuk wordt NIET behandeld bij het parsen, ... van een XML-bestand
- In een CDATA-stuk kan je dus gebruik maken van <, >, &, “, ‘
- Begin: <![**CDATA**[
- Einde:]>

CDATA-stukken

- Voorbeeld
 - Stukken code
 - XML-voorbeelden

```
<![CDATA[  
type naamVariabele;  
naamVariabele = waarde;  
]]>  
<![CDATA[  
<book>  
  <title>De title</title>  
</book>  
]]>
```

Body XML-document

- Body bestaat uit één element
 - Basiselement genoemd (root element)
- Een element kan bevatten
 - Tekst
 - Elementen
 - Entiteiten
 - CDATA-stukken
 - Commentaar
 - Processing Instructions
- Een element kan attributen hebben

Commentaar

- Commentaar
 - <!-- Je commentaar -->

Body XML-document

- Body bestaat uit één element
 - Basiselement genoemd (root element)
- Een element kan bevatten
 - Tekst
 - Elementen
 - Entiteiten
 - CDATA-stukken
 - Commentaar
 - Processing Instructions
- Een element kan attributen hebben

Processing Instructions

- Extra informatie voor externe applicaties

- Syntax

`<?target pseudoattributen?>`

`target:` applicatie of identificatie

`pseudoattributen:` extra informatie

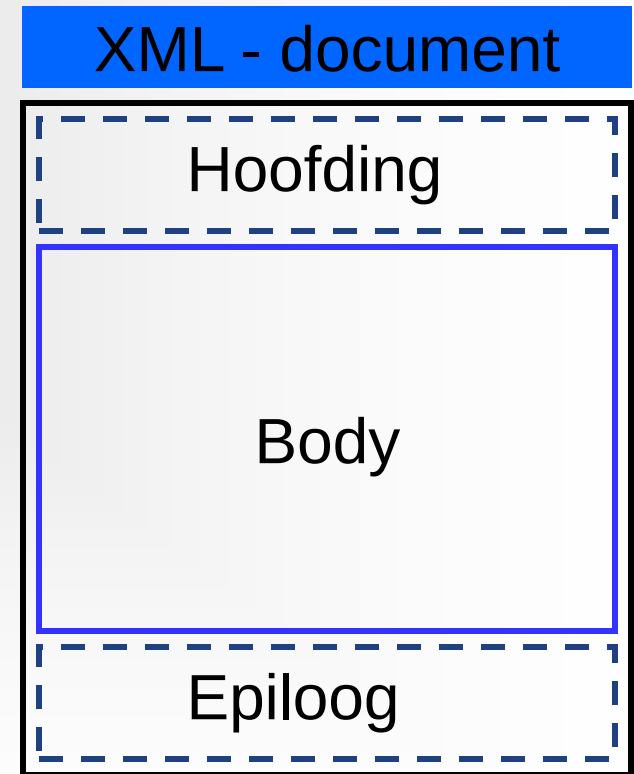
- Voorbeeld

`<?robots index="yes" follow="no"?>`

`<?xml-stylesheet href="boeken.css"
type="text/css"?>`

Structuur XML-document

- XML-document
 - Tekst
 - Bestaat uit
 - Hoofding (optioneel)
 - Body of corpus
 - Epiloog (optioneel)



Hoofding van een XML-document

- Bestaat uit
 - XML-declaratie
 - Commentaar
 - Processing instructies

```
<?xml version="1.0"?>
```

```
<books>
```

```
  <book>
```

```
    ...
```

```
  </book>
```

```
</books>
```

XML-declaratie

Hoofding

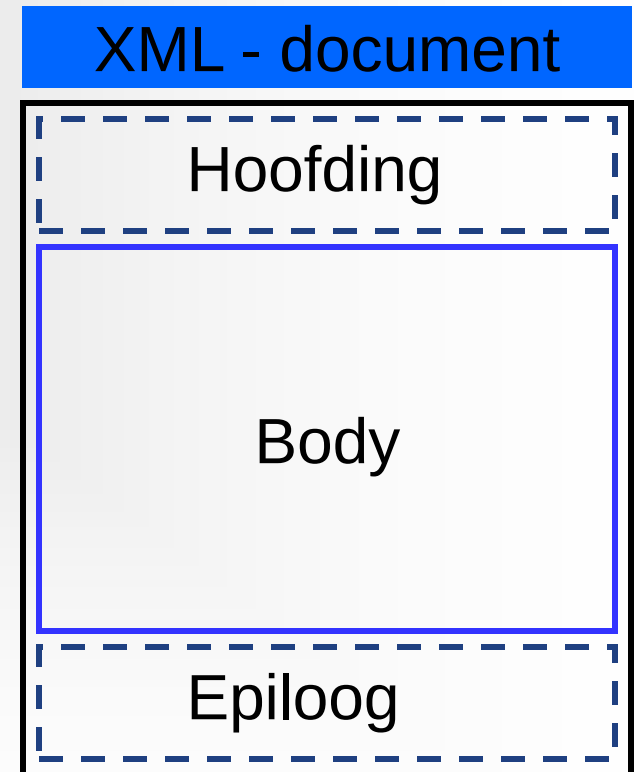
Body

XML Declaratie

- Eerste lijn XML-document
 - `<?xml version="1.0" encoding="ISO-8859-1" standalone="yes"?>`
 - encoding: codering tekst
 - Standaard: UTF-8 (Unicode 8 bits)
 - Optioneel
 - ISO-8859-1 (Latin-1)
 - standalone:
 - Standaard: no
 - Is er geen externe structuur (bv. in DTD) vastgelegd?
 - Optioneel

Structuur XML-document

- XML-document
 - Tekst
 - Bestaat uit
 - Hoofding (optioneel)
 - Body of corpus
 - Epiloog (optioneel)



Epiloog

- Bestaat uit
 - Processing instructies
 - Commentaar

XML en namespaces

- Wat is XML?
- Structuur XML-document
- Welgevormd XML-document
- Namespaces

Welgevormd XML-document

- Voldoet aan XML-syntax
 - Elke element heeft begin-tag en een eind-tag (eventueel in verkorte notatie)
 - De namen van de elementen bevatten geen spaties, beginnen niet met een leesteken of cijfers, ...
 - De naam van een element begint niet met xml
 - Elementen vormen strikte boomstructuur met één basiselement

Welgevormd XML-document

- Voldoet aan XML-syntax
 - De waarde van attributen staan tussen “” of ‘’
 - De attributen van één element hebben verschillende namen
 - Geen commentaar of processing instructions in tags
 - Geen speciale karakters als inhoud van een element of attribuut (gebruik de voorgedefinieerde entiteiten)
 - ...

XML en namespaces

- Wat is XML?
- Structuur XML-document
- Welgevormd XML-document
- Namespaces

Namespaces

- Doel
 - Onderscheid maken tussen elementen met dezelfde naam van verschillende XML-talen
 - Verschillende XML-talen in één document gebruiken
- Namespace bepaalt de context van een element (de taal waarvan het element deel uitmaakt)
- Gekenmerkt door een URI (Uniform Resource Identifier)
 - Hoeft geen bestaande URI te zijn

Namespaces

- Structuur element (bij gebruik namespaces)
 - Naam namespace + “:” + naam element
- Voorbeelden
 - Gebruik lange naam voor elementen
<http://www.zwiftbooks.com:title>
...</http://www.zwiftbooks.com:title>
<http://www.music.com:title>
...</http://www.music.com:title>

Namespaces

- Aangeven namespace met attribuut `xmlns` (standaarddeclaratie)

<book

`xmlns="http://www.zwiftbooks.com">`

`...</book>`

- Namespace geldt voor het element en al zijn kinderen (attributen of elementen) tenzij expliciet een andere namespace wordt opgegeven

Namespaces

■ Afkortingen

```
<zbooks:book  
  xmlns:zbooks="http://www.zwiftbooks.com">  
  <zbooks:isbn>0-596-0058-8</zbooks:isbn>  
</zbooks:book>
```

- Elementen zonder afkorting behoren tot de standaard-namespace van het XML-document

Namespaces

- Lange elementnamen gebruiken
- Aangeven namespace van een element en zijn kinderen met attribuut xmlns
- Afkortingen gebruiken

Informatie

- <http://www.w3.org/XML/>
- <http://www.w3schools.com/>
- <http://www.w3schools.com/xml/default.asp>
- <http://www.brics.dk/~amoeller/XML/index.html>
- <http://www.xml.com/pub/a/98/10/guide0.html>

Boeken

- “XML in a nutshell, a desktop quick reference”, Elliotte Rusty Harold & W. Scotte Means, O'Reilly
- “XML, Web Services and the Data Revolution”, Frank P. Coyle, Addison Wesley