

XML

Alle tekortkomingen op een rijtje! (en een paar pluspunten...)

De geschiedenis van XML

- **SGML**
 - **Standard Generalized Markup Language**
 - **IBM: back to the sixties...**
 - **Markup: structuur, niet processing**
- **HTML: SGML spinoff**
- **1998: XML**
 - **SGML voor het Web, zonder ballast**
- **2002: Hype**
- **2009: Waar staan we?**

Waarom XML?

- XML als documentformaat
 - uitgevers: een bron, veel weergaven
 - SGML is te complex
- XML voor gegevensuitwisseling
 - ASCII, CSV is te simpel en foutgevoelig
 - EDI is te duur
- XML voor het Web
 - HTML kan geen semantiek uitdrukken
 - XML volgende generatie ('Semantic Web')

XML en Unicode

XML

- XML = SGML – ballast + Unicode
 - Een XML document is een boom
 - Een root node
 - Attributen, elementen, tekst

de Volkskrant

ECONOMIE
Helias, u bent ontslagen.
Slokje water?

V.K.NL / ECONOMIE



KUNST
Uitdaging voor choreograaf.
blootrevue in Crazy Horse

V.K.NL / KUNST



LEVEN
Het snob-appeal
van het ebook

TWEEVOLKSKRANTAL



DINSDAG 22 SEPTEMBER 2009
Prijs € 1,30 Jaargang 88 Nummer 2947 32 pagina's

Nieuwe functie Máxima binnen VV

De Nederlandse regering heeft de Vlaamse regering uitgenodigd om de komende maanden te werken aan een plan voor de komende jaren. Het kabinet zal hiermee bezig zijn met de komende maanden. Het kabinet zal hiermee bezig zijn met de komende maanden.

CPB kraakt plannen besparing zorg

Het Centraal Planbureau (CPB) kraakt plannen voor besparingen in de zorg. Het CPB vindt de plannen te hard. Het CPB vindt de plannen te hard.

Deed na botsing pompstation A 27

Een pompstation op de A27 is na een botsing met een vrachtwagen vernield. De pompstation is vernield. De pompstation is vernield.

'Ook pensioenleeftijd moet naar 67'

De regering wil de pensioenleeftijd opheffen naar 67 jaar. De regering wil de pensioenleeftijd opheffen naar 67 jaar.

Film Wit Licht blijft Oscar-zending

De film 'Wit Licht' blijft de Oscar-zending. De film 'Wit Licht' blijft de Oscar-zending.

VK.NL / VK-TV

Popmuziek, Nederlandse muziek, het is er allemaal. Popmuziek, Nederlandse muziek, het is er allemaal.

Waarom niet de 010?

Waarom niet de 010? Waarom niet de 010? Waarom niet de 010?

WEEK LINK WAT ZIJN

WEEK LINK WAT ZIJN. WEEK LINK WAT ZIJN. WEEK LINK WAT ZIJN.

ADRES PRUIZEN

ADRES	PRUIZEN
DE VOLKSKRANT BY	1999-2009
DE VOLKSKRANT BY	1999-2009
DE VOLKSKRANT BY	1999-2009

DE VOLKSKRANT 4 WEKEN OP PROEF VOOR € 10

DE VOLKSKRANT 4 WEKEN OP PROEF VOOR € 10. DE VOLKSKRANT 4 WEKEN OP PROEF VOOR € 10.

NU IN DE VOLKSKRANT WEBWINKEL

NU IN DE VOLKSKRANT WEBWINKEL. NU IN DE VOLKSKRANT WEBWINKEL.

Verpakken en doorverkopen van hypotheeklen is 'cruciaal' voor economisch herstel

IMF: leg bank niet te veel aan banden

ANALYSE
LEVEN
VERPAKKEN
Om de kredietverlening te stimuleren moeten banken hun hypotheeklen en andere leningen weer kunnen doorverkopen, vindt het IMF. Sinds de crisis zal momenteel te weinig leningen worden verkocht op de markt.

AMSTERDAM - Volgens een nieuw rapport van de International Monetary Fund (IMF) is het doorverkopen van leningen een van de belangrijkste manieren om de kredietverlening te stimuleren. Het IMF vindt dat banken hun hypotheeklen en andere leningen weer kunnen doorverkopen op de markt.

Uitstoot van CO₂ gedaald door recessie

De uitstoot van CO₂ is gedaald door de recessie. De uitstoot van CO₂ is gedaald door de recessie.

ROME ROUWT ITALIANEN BEWIJZEN DE LAATSTE EER
Een groep van 100.000 mensen heeft in Rome de laatste eer bewezen aan de voormalige leider van de Italiaanse Communistische Partij.

De laatste eer is bewezen aan de voormalige leider van de Italiaanse Communistische Partij. De laatste eer is bewezen aan de voormalige leider van de Italiaanse Communistische Partij.

Ouders kunnen 'witte vlucht' onderwijs keren

Ouders kunnen de 'witte vlucht' uit het onderwijs keren. Ouders kunnen de 'witte vlucht' uit het onderwijs keren.



De 'witte vlucht' uit het onderwijs kan worden gestopt. De 'witte vlucht' uit het onderwijs kan worden gestopt.

'Wie worden na deze recessie de nieuwe helden?'

CONCERTGEBOUWKEST

1988-1998

PIANISTEN SEIZOEN 2009-10

1988-1998

Ouders kunnen 'witte vlucht' onderwijs keren

Ouders kunnen de 'witte vlucht' uit het onderwijs keren. Ouders kunnen de 'witte vlucht' uit het onderwijs keren.

CONQUESTOR

PIANISTEN SEIZOEN 2009-10

Ouders kunnen de 'witte vlucht' uit het onderwijs keren. Ouders kunnen de 'witte vlucht' uit het onderwijs keren.

CONQUESTOR

PIANISTEN SEIZOEN 2009-10

UITKLAAPMENU



UITKLAAPMENU

Documenten

- opmaak
- kleur
- plaatjes in de tekst
- **vetgedrukte** en *schuingedrukte* tekst
- titel / paragraaf / lijst

	wo	do	vr	za	zo
Zonneschijn (%)	20	40	50	40	40
Neerslagkans (%)	20	10	10	10	20
Neerslaghoeveelheid (mm)	2	0	0	0	0-1
Minimumtemperatuur (°C)	15	9/11	8/11	9/11	9/11
Middagtemperatuur (°C)	18	16/19	17/20	17/20	16/19
Windrichting	W	NW	W	W	W
Windkracht (bft)	4	3	3	2	3

Data

- tabellen, rijen, kolommen
- cijfers
- korte teksten
- datatypes: %, mm, °C, bft
- datatypes en structuur van een XML document vastleggen in een schema

Document Object Model (DOM)

- XML document wordt ingeladen in een object
 - de ‘DOM tree’
- methoden
 - `getElementsByTagName(...)`
 - `getAttribute(...)`
 - `createElement(...)`
- dezelfde methoden in iedere taal: Java, C, .NET...

Namespaces

- een document kan meerdere vocabulaires herbergen
- wat is dan de betekenis van <name>, <id>...
- namespace voorbeelden:
 - `<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">`
 - `<schema xmlns="http://www.w3.org/2001/XMLSchema">`
 - `<element name="purchaseOrder" type="po:PurchaseOrderType"/>`
 - `<faultcode>SOAP-ENV:MustUnderstand</faultcode>`

Pro XML

- combineert documenten met data
- software breed beschikbaar
- geen problemen met
 - encoding
 - regeleinden
 - afspraken

Contra XML

- overhead door start- en eindtags
- DOM is complex
- DOM sluit niet goed aan op programmeertalen
- Namespaces zijn complex
- Redundantie
- `<mdg:auteur naam="Marc" xmlns:mdg="http://www.marcdegrauw.com/voorbeeld/" />`
- `<auteur xmlns="http://www.marcdegrauw.com/voorbeeld/" naam="Marc"></auteur>`

XPath en XSLT

XPath

- XPath: taal om nodes uit de boom te trekken

/	document root
/book	alle <book> elementen onder root
/book[@id]	alle <book> elementen onder root met een 'id' attribuut



```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform" version="1.0">
3   <xsl:template match="message">
4     <xsl:copy>
5       <xsl:apply-templates select="@*" />
6       <xsl:copy-of select="child::require" />
7       <xsl:apply-templates select="child::prescription" />
8     </xsl:copy>
9   </xsl:template>
10
11  <xsl:template match="prescription">
12    <xsl:copy>
13      <xsl:copy-of select="medication | amount" />
14    </xsl:copy>
15  </xsl:template>
16
17  <xsl:template match="@*">
18    <xsl:copy />
19  </xsl:template>
20 </xsl:stylesheet>
21
```

XSLT stylesheet

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <message version="2">
3   <require>
4     <version>1</version>
5     <version>2</version>
6   </require>
7   <prescription>
8     <medication>aspirin</medication>
9     <amount>24</amount>
10    <packaging>box</packaging>
11  </prescription>
12 </message>
13
```

XML bronbestand

```
1 <?xml version="1.0" encoding="utf-8"?>
2 <message version="2">
3   <require>
4     <version>1</version>
5     <version>2</version>
6   </require>
7   <prescription>
8     <medication>aspirin</medication>
9     <amount>24</amount>
10  </prescription>
11 </message>
12
```

XML resultaat

XSLT

- demo (stylesheet – bron – resultaat)
- voordelen
 - snel
 - veel implementaties op veel platforms
- nadelen
 - verbose
 - namespaces zijn foutgevoelig

Schemataalen

Schemataalen

- 1998: XML, DOM, XPath
- 1999: Namespaces, XSLT
- 2001: XML Schema
- W3C versus ISO
- 1999: Schematron
- 2001: RelaxNG
- 2003: ISO RelaxNG
- 2006: ISO Schematron
- demo L1, L2

RelaxNG

```
element addressBook {
  element card { attribute contact_type { "business" | "personal" } ,
    (
      ( element name { text } , element email { text } )
      |
      ( element name { text } , element phone { text } ) ) ,
    element adres { text } ? ,
    element * { text } ?
  }*
}
```

XML Schema

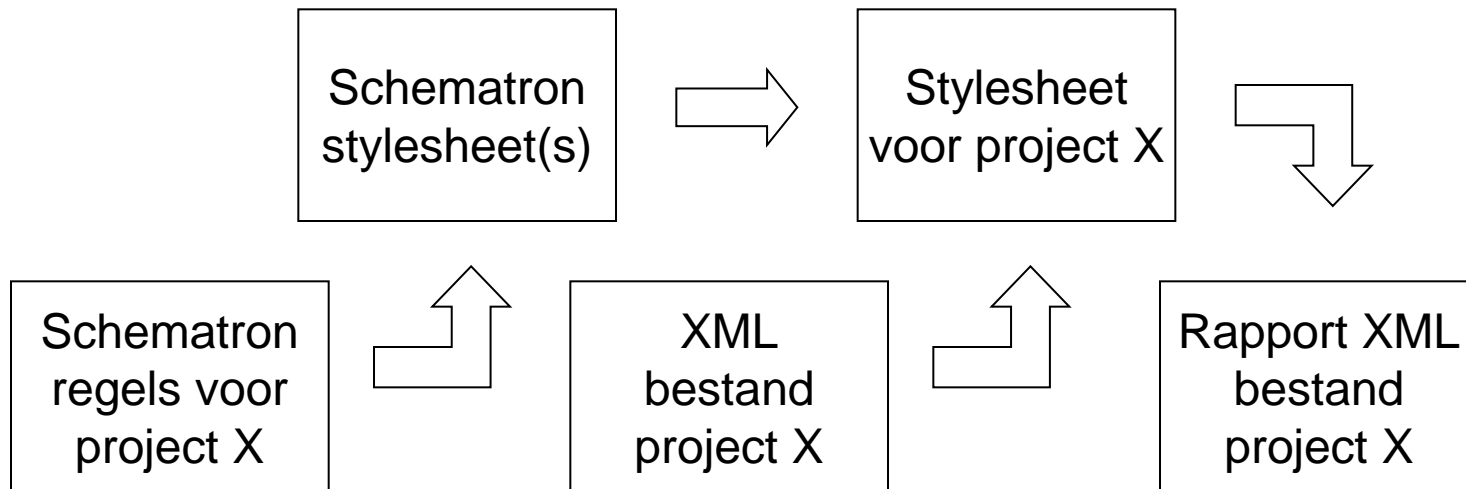
```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
  elementFormDefault="qualified">
  <xs:element name="addressBook">
    <xs:complexType>
      <xs:sequence>
        <xs:element minOccurs="0" maxOccurs="unbounded" name="card">
          <xs:complexType>
            <xs:sequence>
              <xs:choice>
                <xs:sequence>
                  <xs:element name="name" type="xs:string"/>
                  <xs:element name="email" type="xs:string"/>
                </xs:sequence>
                <xs:sequence>
                  <xs:element name="name" type="xs:string"/>
                  <xs:element name="phone" type="xs:string"/>
                </xs:sequence>
              </xs:choice>
              <xs:element minOccurs="0" name="adres" type="xs:string"/>
              <xs:any processContents="skip"/>
            </xs:sequence>
            <xs:attribute name="contact_type" use="required">
              <xs:simpleType>
                <xs:restriction base="xs:token">
                  <xs:enumeration value="business"/>
                  <xs:enumeration value="personal"/>
                </xs:restriction>
              </xs:simpleType>
            </xs:attribute>
          </xs:complexType>
        </xs:element>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

ongeldige constructie!

ongeldige constructie!

Schematron

- rules
- in XML
- alleen een XSLT processor nodig



Schematron

AuthenticationToken.sch [D:\aorta\oorta6\XML\schematron\AuthenticationToken.sch] - <oxygen/> XML Editor

File Edit Find Project Perspective Options Tools Document Window Help

XPath 1.0

L1.xsd x L2.xsd x AuthenticationToken.sch x IU1.xsl x

```
9   author: Marc de Graauw
10  -->
11  <schema xmlns="http://purl.oclc.org/dsdl/schematron">
12    <title>HL7 Authentication Token Validation</title>
13    <ns uri="urn:hl7-org:v3" prefix="hl7"/>
14    <ns uri="http://www.aortarelease.nl/805/" prefix="ao"/>
15    <ns uri="http://schemas.xmlsoap.org/soap/envelope/" prefix="soap"/>
16    <pattern>
17      <title>Authentication Token</title>
18      <rule context="/soap:Envelope[1]/soap:Header[1]/ao:authenticationTokens[1]/ao:signedData[1]">
19        <!-- Transmission wrapper elements are located by looking for interactionId siblings... -->
20        <assert
21          test="ao:authenticationData/ao:messageId/ao:root = //hl7:interactionId/..hl7:id/@root"
22          >Root in messageId is niet gelijk aan message.id root</assert>
23        <assert
24          test="ao:authenticationData/ao:messageId/ao:extension = //hl7:interactionId/..hl7:id/@extension"
25          >Extension in messageId is niet gelijk aan message.id root</assert>
26        <assert
27          test="ao:authenticationData/ao:notBefore &lt; ao:authenticationData/ao:notAfter "
28          >Datum notBefore moet voor notAfter liggen</assert>
29        <assert
30          test="ao:authenticationData/ao:addressedParty/ao:root = '2.16.840.1.113883.2.4.6.6'"
31          >Root in addressedParty is niet gelijk aan 2.16.840.1.113883.2.4.6.6</assert>
32        <assert
33          test="ao:authenticationData/ao:addressedParty/ao:extension = '1'"
34          >Extension in addressedParty is niet gelijk aan 1</assert>

```

Text Grid Author

U+0020 37:1

Schema pro en con

- Schema's
 - zonder schema: eindeloze discussies
- XML Schema
 - is complex
 - kan veel simpele constraints niet uitdrukken
 - is overal aanwezig, m.n. in Microsoft en Web Services
- RelaxNG
 - simpeler
 - minder software
- Schematron
 - regelgebaseerd
 - XSLT processor is de hele toolkit

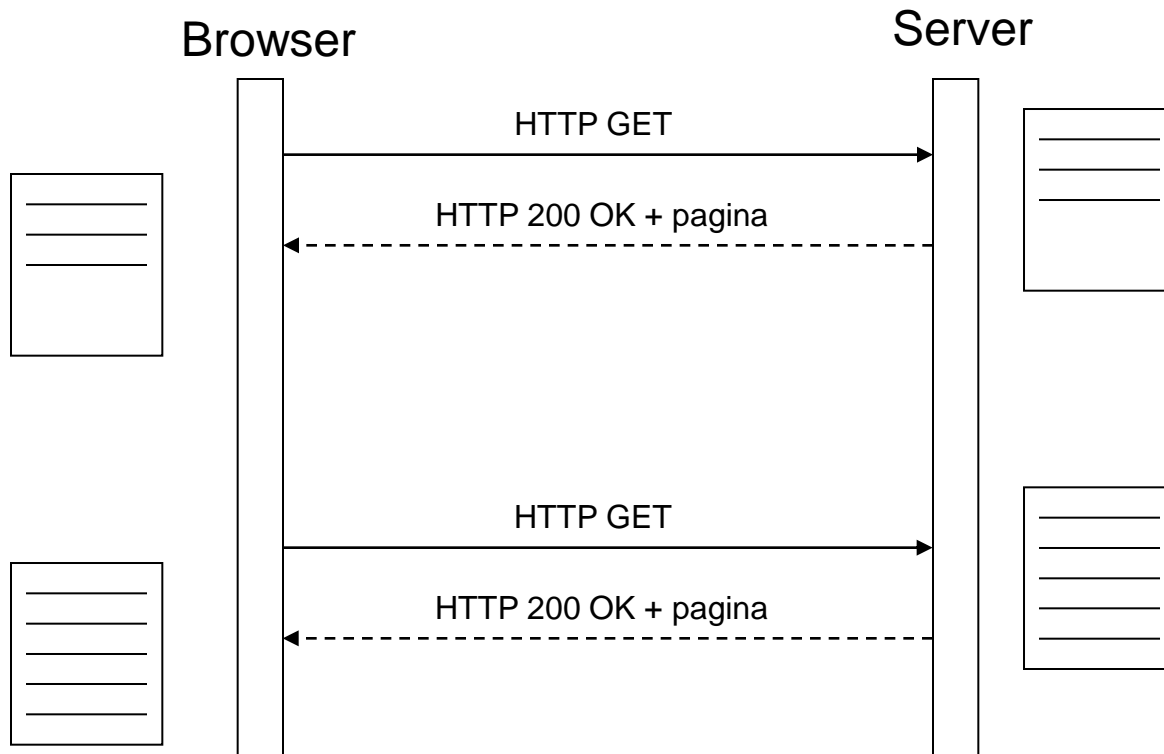
JSON en Ajax

JSON

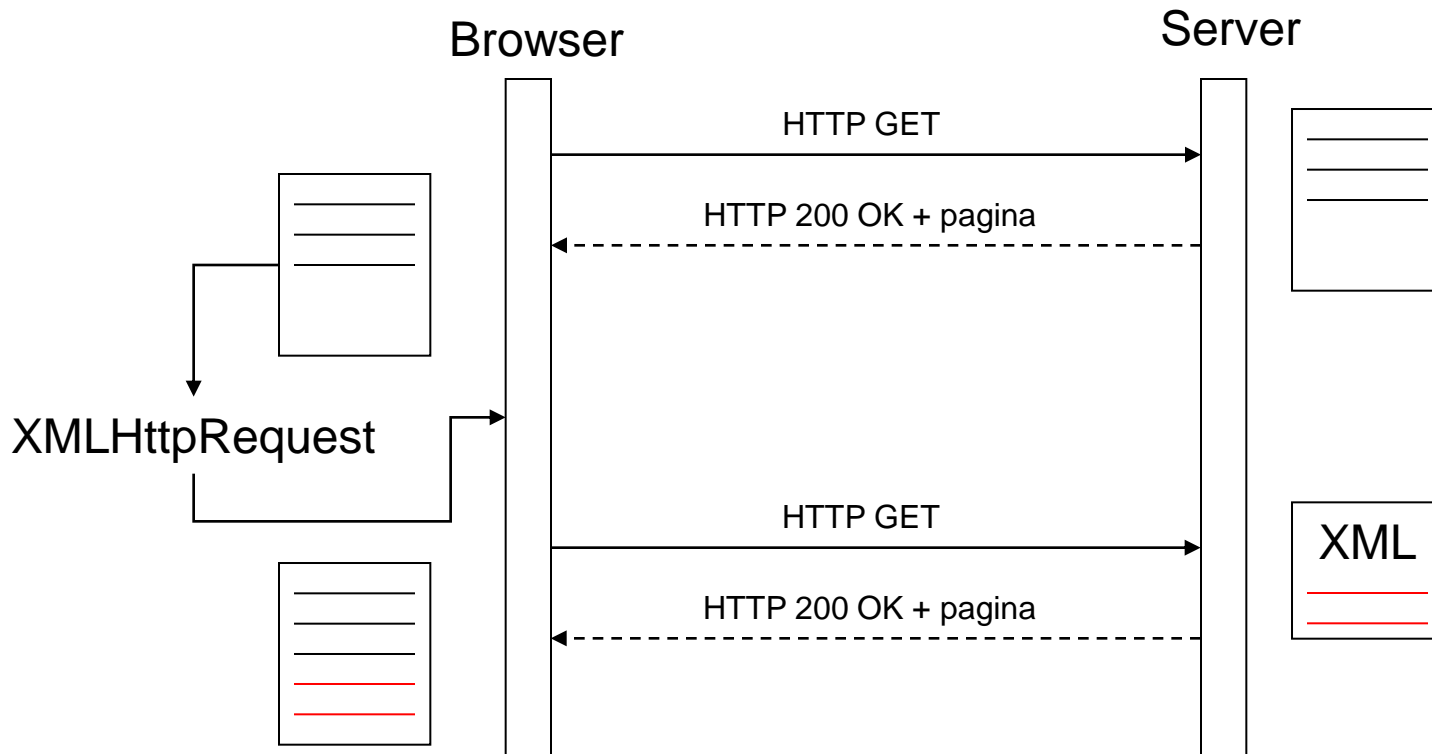
```
person = {  
  "name": "Simon Willison",  
  "age": 25,  
  "height": 1.68,  
  "urls": [  
    "http://simonwillison.net/",  
    "http://www.flickr.com/photos/simon/",  
    "http://simon.incutio.com/"  
  ]  
}
```

- JavaScript Object Notation
- compact
- sluit goed aan op programmeertalen
- geen schemataal
- altijd Unicode; voorkeur UTF-8

HTML



XMLHttpRequest



AJAX

- Asynchronous JavaScript and XML
 - XMLHttpRequest
 - ‘XML’ is hier vaak JSON!
 - pulldowns, refresh, drag & drop
 - Javascript libraries
 - Prototype, jQuery e.v.a.
- AJAX
 - volledige applicatie in de browser
 - ASP

Web Services

Web Services basis

- XML voor de gegevens
- Schema voor
 - contract
 - design
 - codegeneratie
 - validatie
- Transport: SOAP en HTTP
- Servicebeschrijving: WSDL

Web Services stack (WS-*)

- Security
 - XML Signature, XML Encryption
 - XML Canonicalization (2x)
 - WS-Security, WS-SecureConversation, WS-Trust
- Transport
 - WS-Addressing
 - WS-ReliableMessaging
 - MTOM
- Metadata
 - WS-Policy
- Totaal: complex!

Waarom XML?

- XML als documentformaat
 - uitgevers: het werkt
 - SGML is grotendeels vervangen
- XML voor gegevensuitwisseling
 - veelgebruikt: EDI en CSV zijn geen alternatief
 - Web Services: complex...
- XML voor het Web
 - grotendeels mislukt
 - JSON vult een deel van deze niche

Vragen?