

Gedistribueerde toepassingen

Bert De Saffel

Master in de Industriële Wetenschappen: Informatica Academiejaar 2018–2019

Gecompileerd op 12 november 2018

Inhoudsopgave

I	Theorie	3
1	Extensible Stylesheet Language Family	4
1.1	Inleiding	4
1.2	XSLT	4
1.2.1	Structuur	4
1.2.2	Instructie-elementen	5
2	XML Path Language (XPath)	7
2.1	Paden	7
2.1.1	Basispad (Root Location Path)	7
2.1.2	Kinelement-stap (Child Element Location Steps)	7
2.1.3	Attribuut-stap (Attribute Location Steps)	8
2.1.4	Voorwaarden	8
2.1.5	Lange padnamen	8
2.2	Andere XPath-uitdrukkingen	9
2.2.1	Datatypes	9
2.3	Functies	10
3	Webservices	11
3.1	Componenten	11
3.1.1	EDI	11
3.1.2	RPC	12

3.2	Gedistribueerde objectsystemen	12
3.2.1	RMI	12
3.2.2	DCOM	12
3.2.3	CORBA	13
3.3	MOM	13
3.4	SOA	13
3.5	Webservices in Java	13
3.6	Webservices in C#	14
4	Netwerkprogrammatie in Java	15
5	Java NIO	17

Deel I

Theorie

Hoofdstuk 1

Extensible Stylesheet Language Family

1.1 Inleiding

XSL staat voor *Extensible Stylesheet Language Family* en is een standaard om XML-documenten te presenteren en te transformeren. De drie componenten van XSL zijn:

- **XSLT** (Extensible Stylesheet Language Transformations) : dit is een XML-taal dat XML-documenten kan omvormen naar opnieuw XML, maar kan ook HTML, L^AT_EX, JSON, enz... zijn.
- **XPath** : dit is een taal dat waarmee bepaalde stukken van een XML-document kunnen gemanipuleerd worden aan de hand van het opbouwen van paden.
- XSL-FO (XSL Formatting Objects)

1.2 XSLT

1.2.1 Structuur

Een XSLT-bestand moet altijd voldoen aan volgende structuur:

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet version="1.0"
  xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
  <xsl:output method="html"/>
  <xsl:template match="patroon">
    ...
  </xsl:template>
```

```

    ...
</xsl:stylesheet>

```

1. `xsl:stylesheet` is het basiselement
2. `xsl:output` is het element dat het type transformatie vastlegt in het attribuut *method*
3. `xsl:template` bevat wat er moet gebeuren voor elk element dat aan *patroon* voldoet.

1.2.2 Instructie-elementen

- **xsl:value-of** : Bepaalt de waarde van een XPath uitdrukking.

```
<xsl:value-of select='uitdrukking' />
```

- **apply-templates** : Zal de templates van de onmiddellijke kindelementen van het huidig element toepassen. Indien er geen templates zijn, zal de inhoud van het element naar de uitvoer uitgeschreven worden. Het optionele *select* attribuut zal enkel de templates uitvoeren voor elk element die aan de uitdrukking voldoet.

```
<xsl:apply-templates [select='uitdrukking'] />
```

- **xsl:for-each** : Voert een actie uit voor alle knopen die door het *select*-attribuut bepaalt wordt. Binnen de for-each kunnen meerdere instructie-elementen voorkomen.

```
<xsl:for-each select='uitdrukking'>...</xsl:for-each>
```

- **xsl:sort** : Dit is een kindelement van *xsl:apply-templates* of een *xsl:for-each* element. Dit element zal de knopen sorteren die aan de waarde van het select attribuut voldoen. Optionele attributen zijn

- *data-type* : legt het datatype vast (*text* of *number*). Dit heeft een invloed op de sorteervolgorde (bij *text* is 10 kleiner dan 2).
- *order* : bepaalt de sorteervolgorde (*ascending* of *descending*).

```
<xsl:sort select='uitdrukking' [data-type='text|number'] [order='ascending|descending'] />
```

- **xsl:if** : Dit is een typische selectiestructuur. Het bevat als enig attribuut *test* met als waarde een predicaat.

```
<xsl:if test='logische uitdrukking'> ... </xsl:if>
```

- **xsl:choose** : Een alternatieve selectiestructuur equivalent zoals een switch in programmeertalen zoals Java en C#. Dit element wordt gevolgd door één of meerdere **xsl:when** elementen. Een *xsl:when* element is equivalent met een *xsl:if* element en heeft ook als enig attribuut *test*.

```

<xsl:choose>
  <xsl:when test="logische uitdrukking 1">
    ...
  </xsl:when>
  <xsl:when test="logische uitdrukking 2">
    ...

```

```
        </xsl:when>
        ...
    </xsl:choose>
```

- **xsl:text** : Wanneer tekst spaties bevat is het aan te raden om `xsl:text` te gebruiken zodat deze ook opgenomen worden in het resultaat. Spaties die niet in een `xsl:text` element staan worden genegeerd.

```
<xsl:text>stukje tekst</xsl:text>
```

- **xsl:variable** : Legt een constante variabele vast. Het verplichte attribuut is *name*, wat de naam van de variabele vastlegt. Het invullen van deze waarden kan op twee manieren :
 1. via het select attribuut `<xsl:variabele name='varnaam' select='uitdrukking'/>`
 2. via de inhoud van het element: `<xsl:variabele name ="varnaam"> ... </xsl:variable>`
In dit geval kan de waarde ook opgebouwd worden uit meerdere instructies.
- **xsl:param** : Indien dit element gedeclareerd wordt in het begin van een `xsl:template` element, dan is het mogelijk om parameters mee te geven aan dit template. De declaratie kan op dezelfde manier gebeuren als bij `xsl:variabele`.
- **xsl:with-param** : Dit element is een kindobject van `xsl:apply-templates`. Dit element vult voor een bepaalde *name* de waarde die in *select* staat.

```
<xsl:apply-templates ... >
    <xsl:with-param name='varnaam' select='uitdrukking'>
</xsl:apply-templates>
```

Hoofdstuk 2

XML Path Language (XPath)

XPath is een taal om knopen te selecteren van een XML-document. Het resultaat van een XPath-uitdrukking kan een getal, string of logische waarde zijn.

2.1 Paden

Belangrijke XPath uitdrukkingen zijn paden. Een pad identificeert nul, één of meerdere knopen in een XML-document. Een pad bestaat uit *location steps* die verbonden zijn met een /. Bij elk pad zijn volgende wildcards mogelijk:

- *: selecteert elk elementknoop in de huidige context.
- node(): selecteert alle knopen, dus niet enkel elementen
- @*: selecteert alle attributen in de huidige context.

2.1.1 Basispad (Root Location Path)

Het eenvoudigste pad is / en selecteert het root-element van het XML-document. Verder is dit een absoluut pad, dus deze uitdrukking zal **altijd** het root-element teruggeven.

2.1.2 Kinelement-stap (Child Element Location Steps)

Dit pad bestaat uit de naam van een element en selecteert alle elementen met deze naam in de huidige context. De context is afhankelijk van de ouder.

```
<xsl:template match="author">  
  <xsl:apply-templates select="name" />  
</xsl:template>
```



```
</xsl:template>
```

In dit voorbeeld wordt de XPath uitdrukking van het selectattribuut `/author/name/`.

2.1.3 Attribuut-stap (Attribute Location Steps)

Een attribuut selecteren begint met een `@` en wordt gevolgd door de naam van het attribuut.

```
<xsl:template match ="query">
  <xsl:value-of select="@isbn"/>
</xsl:template>
```

2.1.4 Voorwaarden

```
<xsl:apply-templates select="//book[title='XML']/name[.='Ongenaë']/>
```

```
<xsl:apply-templates select="//person[@born<=1976]"/>
```

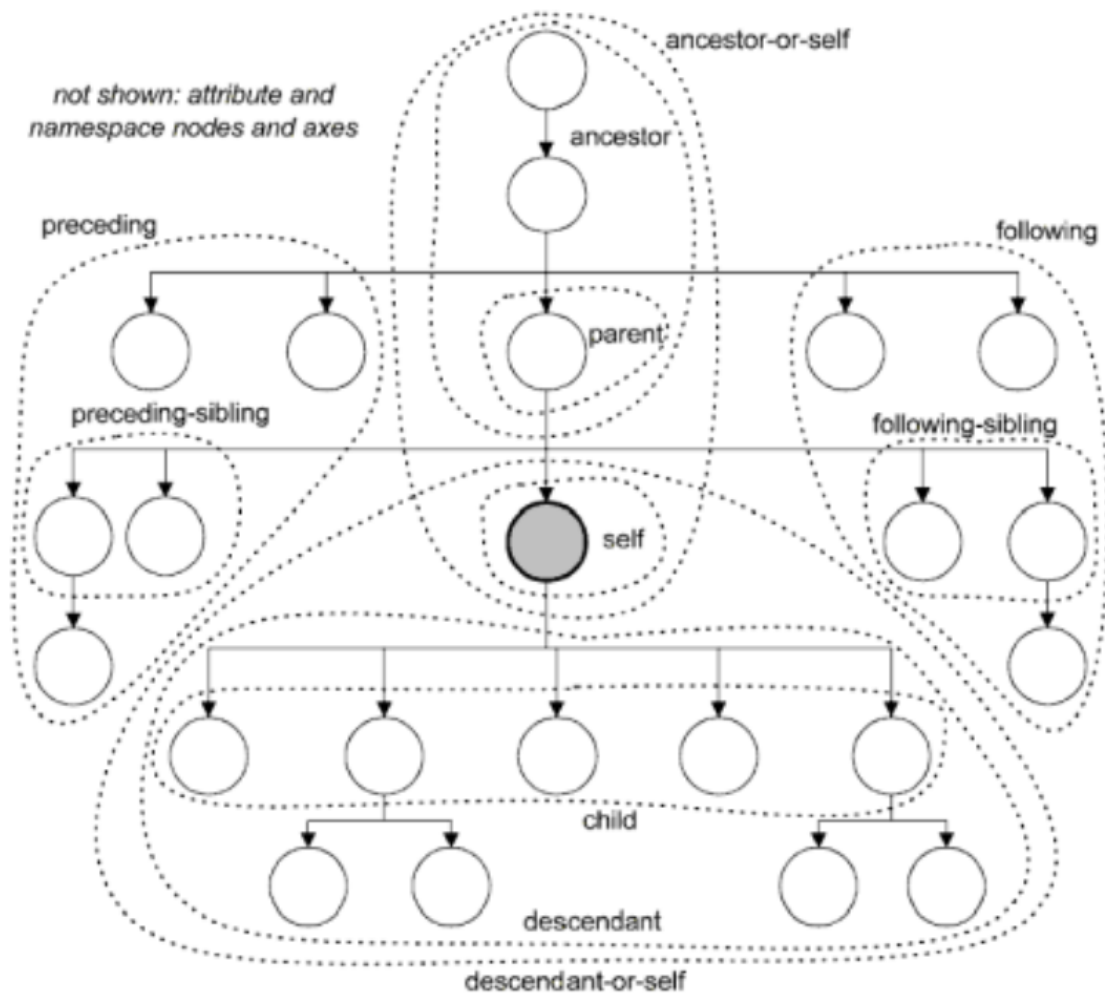
2.1.5 Lange padnamen

In elke XPath uitdrukking wordt elke stap voorafgegaan door de `child::` as.

```
books/book/isbn => child::books/child::book/child::isbn
```

andere assen:

- parent (of `..`)
- self (of `.`)
- descendant-or-self (of `//`)
- ancestor
- ancestor-or-self
- namespace
- descendant
- following-sibling
- preceding-sibling
- following
- preceding



2.2 Andere XPath-uitdrukkingen

XPath kan ook getallen, logische waarden en strings manipuleren.

2.2.1 Datatypes

- **Getallen** : Stel dat *born* het geboortjaar van een persoon is, dan zal de volgende uitdrukking de eeuw uitschrijven van dit jaar. `<xls:value of select="(@born - (@born mod 100)) div 100 + 1"`
- **Strings** : Strings staan tussen enkele of dubbele aanhalingstekens. Strings vergelijken kan met `=` en `!=` operatoren.

- **Logische waarden** : De sleutelwoorden `true` en `false` bestaan niet in XPath, wel worden ze respectievelijk voorgesteld door de functies `true()` en `false()`. Op logische waarden kunnen de operatoren *and* en *or* toegepast worden. Met `not(...)` wordt de negatie toegepast.

2.3 Functies

XPath kent een aantal functies.

- **Knopenfuncties** zoals *position()*, *last()*, *count(...)* en *id(...)* bepalen respectievelijk
 - het volgnummer van de huidige knoop in de context,
 - het aantal knopen in de huidige context (volgnummer van de laatste knoop),
 - telt het aantal knopen van het argument, dat een verzameling knopen is en
 - bepaalt een verzameling knopen. Deze verzameling bevat alle elementen van een XML-document met één van de gespecificeerde ID's. Het argument is een string bestaande uit ID's gescheiden door spaties.
- **Stringfuncties** zoals *concat(...)*, *contains(...)*, *starts-with(...)*, *string(...)* (conversie naar string) en *string-length(...)*.
- **Logische functies** zoals *true()*, *false()*, *not(...)*, *boolean(...)* (conversie naar bool waarde).
- **Numerieke functies** zoals *ceiling(...)*, *floor(...)*, *number(...)*, *round(...)*.

Hoofdstuk 3

Webservices

Een gedistribueerd systeem zal over verschillende servers API's beschikbaar hebben. Tabel 3 geeft de voornaamste verschillen tussen een monolitische architectuur en een gedistribueerde architectuur. Er zijn twee belangrijke componenten bij gedistribueerde systemen, **EDI (Electronic Data**

	Monolitisch	Gedistribueerd
Communicatie	Tussen processen, gedeeld geheugen	Over een netwerk
Globale toestand	Mogelijk	Niet mogelijk
Globale tijd	Mogelijk via lokaal besturingssysteem	Niet mogelijk
Fouten	Eenvoudig te ontdekken	Gedeeltelijk falen moeilijk te ontdekken
Locatie	Alle componenten op één machine	Variabel
Beveiliging	Taak van het besturingssysteem	Inherent kwetsbaar door networkcommunicatie.

Interchange) en **RPC (Remote Procedure Call)**.

3.1 Componenten

3.1.1 EDI

Dit soort toepassingen digitaliseren het papierwerk dat vaak plaatsvindt in eender welke organisatie. Dit werkt in drie stappen:

1. Verzamel de informatie die verzonden moet worden. Deze data komt vaak uit interne bronnen, die een willekeurige structuur kunnen hebben?
2. Zet deze informatie om naar het EDI formaat door gebruik te maken van vertaler die de interne informatie kan omzetten naar het EDI formaat.
3. Verstuur deze informatie naar de bestemming. Een zender en ontvanger zijn direct aangesloten met elkaar via software.

3.1.2 RPC

RPC is het aanroepen van een procedure (functie, methode, ...) op een ander toestel dan waarop een programma fysiek uitgevoerd wordt. Dit proces bevat volgende stappen:

1. De client maakt een RPC aan en verstuurt deze naar een server die deze call kan behandelen.
2. De server verwerkt de procedure en stuurt het resultaat terug naar de client.
3. De client kan gebruik maken van dit resultaat op lokaal niveau.

3.2 Gedistribueerde objectsystemen

Gedistribueerde systemen maken gebruik van EDI en RPC om een systeem te implementeren. Er worden enkele besproken zoals **RMI (Remote Method Invocation)**, **DCOM (Distributed Component Object Model)** en **CORBA (Common Object Request Broker Architecture)**

3.2.1 RMI

Deze technologie is ontwikkelt door Oracle en bestaat typisch uit een **Server** en **Client** applicatie. Een server zal methoden beschikbaar stellen en zal wachten tot dat een client deze methode uitvoert (invocation). RMI is het mechanisme dat de verbinding en informatieuitwisseling tussen een client en een server behandelt. Een gedistribueerd systeem gebaseerd op RMI kan de volgende 3 zaken uitvoeren.

1. *Lokalisatie van remote objecten.* Remote objecten moeten opgeslagen worden in een zogenaamde **RMI registry**.
2. *Communicatie met remote objecten.* Het aanroepen van methoden van een remote object wordt volledig door RMI behandelt. Dit heeft als gevolg dat een programmeur niet eens moet weten dat hij met een remote object bezig is.
3. *Klassedefinities laden van remote objecten.* Aangezien dat RMI objecten kan terugsturen/versturen, voorziet RMI de mogelijkheid om klassedefinities van deze objecten op te halen.

3.2.2 DCOM

DCOM is een uitbreiding op COM om communicatie tussen verschillende toestellen, hetzij op een LAN, hetzij op een WAN of zelfs op het internet aan te spreken. DCOM behandelt het low-level gedeelte van de netwerkcommunicatie, zodat een programmeur hier zich niet meer mee bezig hoeft te houden.

3.2.3 CORBA

CORBA is zoals de anderen, ook een manier om communicatie te hebben tussen verschillende systemen. CORBA baseert zich op 4 punten:

1. Applicatieobjecten die specifiek zijn voor de applicatie. Een CORBA object is een encapsulatie van zulke objecten.
2. Een **ORB (Object Request Broker)** dat de requests van verschillende services behandelt.
3. Een verzameling van services die waarschijnlijk gebruikt zullen worden door veel applicaties.
4. Een verzameling van componenten die bovenop deze services gebouwd zijn.

3.3 MOM

MOM (Message Oriented Middleware) is een software systeem dat instaat voor het aanmaken, versturen, ontvangen en lezen van berichten op zowel synchrone of asynchrone manier. Een client applicatie moet dus enkel deze middleware aanspreken om berichten op te halen of te versturen.

3.4 SOA

SOA (Service Oriented Architecture) is een architectuurstijl dat gebruik maakt van services die onafhankelijk zijn van andere services. Een service kan individueel aangepast worden zonder dat andere services hiervan last hebben. Deze architectuurstijl kent een aantal uitdagingen:

- Gelijktijdige toegang tot bronnen
- Wat indien een deel van de opdracht mislukt?
- Wat als één partner incompatibel wordt?
- Geen gedeeld geheugen voor requester en provider
- Traagheid en onbetrouwbaarheid gebruikte transport

Er blijkt ook een gelijkheid te zijn met SOA en microservices. Het verschil tussen microservices en SOA is **ToDo: saai**

3.5 Webservices in Java

Voor webservices te implementeren in java maken we gebruik van SOAP. Een SOAP aanvraag ziet er als volgt uit :

```
<?xml version="1.0" encoding="UTF-8"?>
<S:Envelope
  xmlns:S="http://schemas.xmlsoap.org/soap/envelope/">
  <S:Header/>
  <S:Body>
    <ns2:geefBoek xmlns:ns2="http://web.boeken.iii.be/">
      <isbn>isbn2</isbn>
    </ns2:geefBoek>
  </S:Body>
</S:Envelope>
```

Een eenvoudige webserivce dat een boek opvraagd met een bepaald isbn ziet er dan zo uit:

```
@WebService(serviceName = "Catalogus")
public class Catalogus {
    @WebMethod(operationName = "geefBoek")
    public Book geefBoek(@WebParam(name = "isbn") String isbn){
        try {
            BoekenLijst boeken = new BoekenLijstImpl();
            return boeken.geefBoek(isbn);
        } catch (Exception e) {
            return null;
        }
    }
}
```

3.6 Webservices in C#

.ToDo: al de rest lel

Hoofdstuk 4

Netwerkprogrammatie in Java

Belangrijke netwerkklassen in java, in de package java.net:

- **TCP**
 - URL
 - URLConnection
 - Socket
 - ServerSocket
- **UDP**
 - DatagramPacket
 - DatagramSocket
 - MulticastSocket

De cliënt maakt een **Socket** object aan. De constructor heeft de locatie van een bepaalde server, en een poortnummer, die de server openzet. Communicatie met de socket gebeurt met de **PrintWriter** klasse, die naar de server zal sturen, en de **BufferedReader** klasse, die berichten van de server zal ontvangen. Gebruik best de `try` met `resources` om instanties van deze objecten te maken, zodat ze automatisch verdwijnen indien de code uit het `try` block komt. Vanaf dan kan er met de server gecommuniceert worden door weg te schrijven via de **PrintWriter**. Wachten op een bericht van de server kan door de `readLine()` methode van de **BufferedReader**klasse op te roepen. Dit is een synchrone methode dus de applicatie zal blokkeren.

De server zal gebruik maken van een **ServerSocket** klasse. Deze werkt op identiek dezelfde manier als de **Socket** klasse die de cliënt gebruik, behalve dat een **ServerSocket** klasse meerdere sockets kan accepteren. Tot slot is er nog een protocol nodig, die de communicatie tussen de server en communicatie beschrijft. Een eenvoudig protocol vereist bijvoorbeeld dat een bericht start met een bepaald sleutelwoord.


```
// CLIENT
String hostName = "localhost";
int portNumber = 4444;
try(
    Socket socket = new Socket(hostName, portNumber);
    PrintWriter out = new PrintWriter(socket.getOutputStream(), true);
    BufferedReader in = new BufferedReader(
        new InputStreamReader(socket.getInputStream())
    );
){
    String input = ... // input ophalen van gebruiker
    while(!input.equals("exit")){
        out.println(input); // naar server
        System.out.println("echo: ") + in.readLine()); // van server
        input = ... // input ophalen van gebruiker
    }
} catch (IOException ex) {
    System.err.println(ex.getMessage());
}

// SERVER
try {
    ServerSocket socket = new ServerSocket(4444);
    while(true) {
        try(
            PrintWriter out = new PrintWriter(
                socket.getOutputStream(), true
            );
            BufferedReader in = new BufferedReader(
                new InputStreamReader(socket.getInputStream())
            );
        ){
            String input, output;
            while((input = in.readLine()) != null){
                output = protocol.processInput(input);
                out.println(output);
                if(output.equals("Bye")){
                    break;
                }
            }
        }
    }
}
}
```

Een socket wordt best geïmplementeerd met behulp van threads, zodat verschillende cliënten op hetzelfde moment behandeld kunnen worden.

Hoofdstuk 5

Java NIO

Java NIO (New IO) biedt een alternatieve manier om met IO te werken.

- **Channels en Buffers.** Data wordt altijd gelezen van een bepaald kanaal naar een buffer, of geschreven van een buffer naar een kanaal.
- **Non-blocking IO.** Een thread kan aan een kanaal vragen om data van een buffer in te lezen. Terwijl het kanaal dit doet, kan de thread andere zaken uitvoeren en wachten tot het kanaal klaar is.
- **Selectors.** Een selector is een object dat meerdere kanalen kan monitoren. Het reageert op events (connectie opened, data arrived, ...). Een enkele thread kan meerdere kanalen monitoren met behulp van selectors. Een selector registreren gebeurt door een kanaal mee te geven, en de `select()` methode op te roepen van de `Selector` klasse.

De kern van Java NIO zijn de klassen `Channel`, `Buffer` en `Selector`. Andere klassen zoals `Pipe` en `FileLock` dienen gebruikt te worden in samenwerking met de drie kernklassen. Elke klasse kent meerdere implementaties. Zo kent de `Channel` klasse bijvoorbeeld: `FileChannel`, `DatagramChannel`, `SocketChannel` en `ServerSocketChannel`. De klasse `Buffer` kent meerdere implementies afhankelijk van het datatype: `ByteBuffer`, `CharBuffer`, `DoubleBuffer`, `FloatBuffer`, `IntBuffer`, `LongBuffer`, `ShortBuffer`. De verschillen tussen de `Channel` en `Stream` klasse zijn:

- Lezen en schrijven is mogelijk met één kanaal. Een stream is ofwel leesmode of schrijfmode.
- De lees en schrijfoperatie kan asynchroon gebeuren.
- Een kanaal kan lezen van, of schrijven naar een `Buffer`

Het gebruik van een eenvoudig `Channel` object:

```
RandomAccessFile file = new RandomAccessFile("data/file.txt", "rw");
FileChannel inChannel = file.getChannel();
ByteBuffer buffer = ByteBuffer.allocate(48);
int bytesRead = inChannel.read(buf);
```

```
while (bytesRead != -1){
    System.out.println("Read " + bytesRead);
    buf.flip();
    while (buf.hasRemaining()){
        System.out.print((char) buf.get());
    }
    buf.clear();
    bytesRead = inChannel.read(buf);
}
file.close();
```