

avans
hogeschool

Module XML

Week4: DOM, SAX, XML-programmeren

XML Parsers: definitie

Een XML Parser is code die een XML document:

- In het geheugen leest (van schijf of via TCP/IP)
- Analyseert (structuur achterhaalt)
- Valideert (optioneel)
- Presenteert aan programmeurs via API (optioneel)

Elke XML applicatie heeft een XML Parser nodig

XML Parsers: overzicht

Categorieën XML Parsers:

- Validerende en niet-validerende parsers
- Parsers met support voor SAX
- Parsers met support voor DOM
- Parsers geschreven in bepaalde taal (Java, C++, etc.)

XML Parsers: validerend

- Alle parsers checken op “well-formed”.
- Validerende parsers checken de grammatica aan de hand van de DTD of het XML-Schema
- Het valideren van XML-documenten is ‘niet eenvoudig’.

XML Parsers: DOM support

DOM = **D**ocument **O**bject **M**odel

DOM Parsers:

- Creëren een Document Object in de vorm van een boomstructuur.
- Definiëren een API waarmee applicaties de stijl, structuur en inhoud van een XML-doc kunnen lezen en schrijven.
 - *(API is verzameling methoden)*
- Zetten het **hele** document object in het actieve geheugen.

XML Parsers: SAX support

SAX = **S**imple **A**PI for **X**ML

SAX Parsers:

- Gaan element-voor-element van boven naar beneden door een XML-doc en genereren events naar een XML-applicatie (op streaming manier)
- Zelf data-structuur opbouwen

XML Parsers: wanneer DOM?

- Documenten zijn (relatief) klein tov beschikbaar geheugen
- Snelheid is minder belangrijk
- Verwerking vraagt 'heen en weer', dus random access tot gegevens
- Ontwikkelaars werken i.h.a. liever met tree structuur dan met events-sequence
- Betere portabiliteit (??): DOM is W3C standaard

Hoe meer van deze eigenschappen, hoe beter DOM past

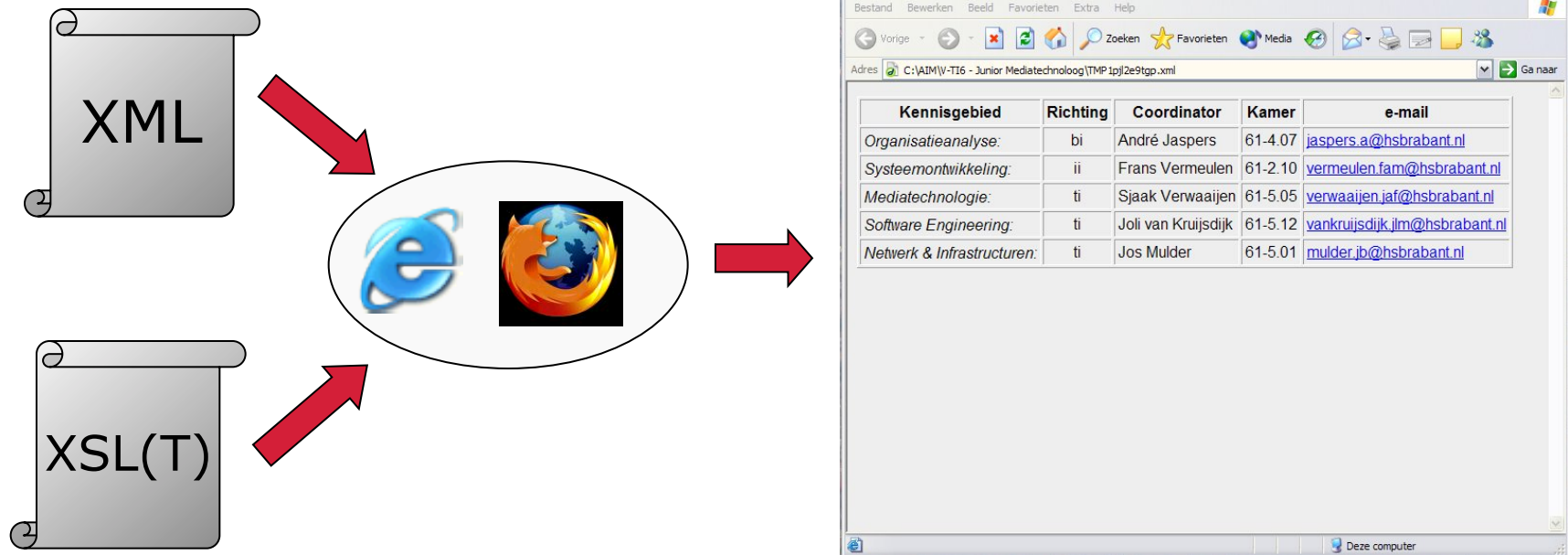
XML Parsers: wanneer SAX?

- Snelheid is belangrijk
- Beperkt geheugen
- Als je kleine, aaneengesloten 'chunk's per stuk wilt verwerken
- Interne data-structuur van programma is anders dan het XML document zelf
- Wel lezen, maar niet (terug)schrijven

Hoe meer van deze eigenschappen, hoe beter SAX past

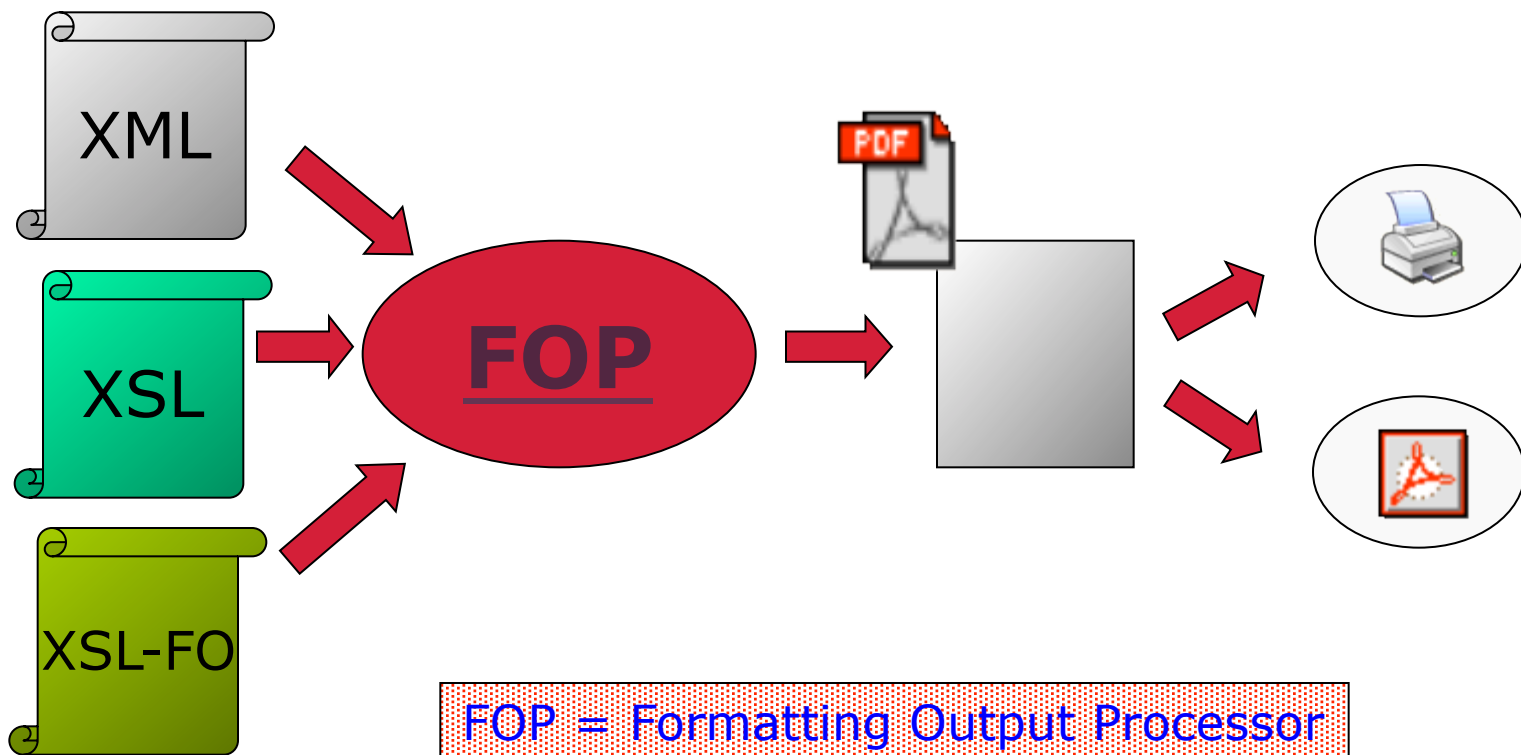
XML Parsers: voorbeelden

Non-validating parser



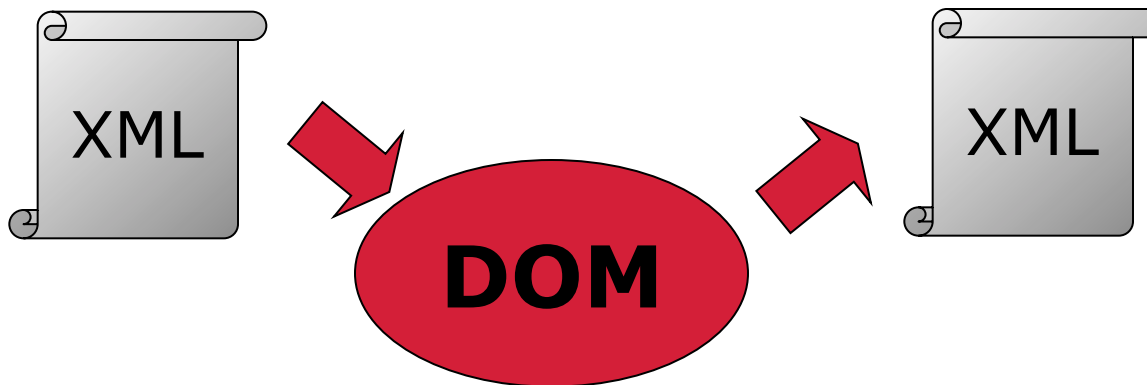
XML Parsers: voorbeelden

DOM gebaseerde formatter



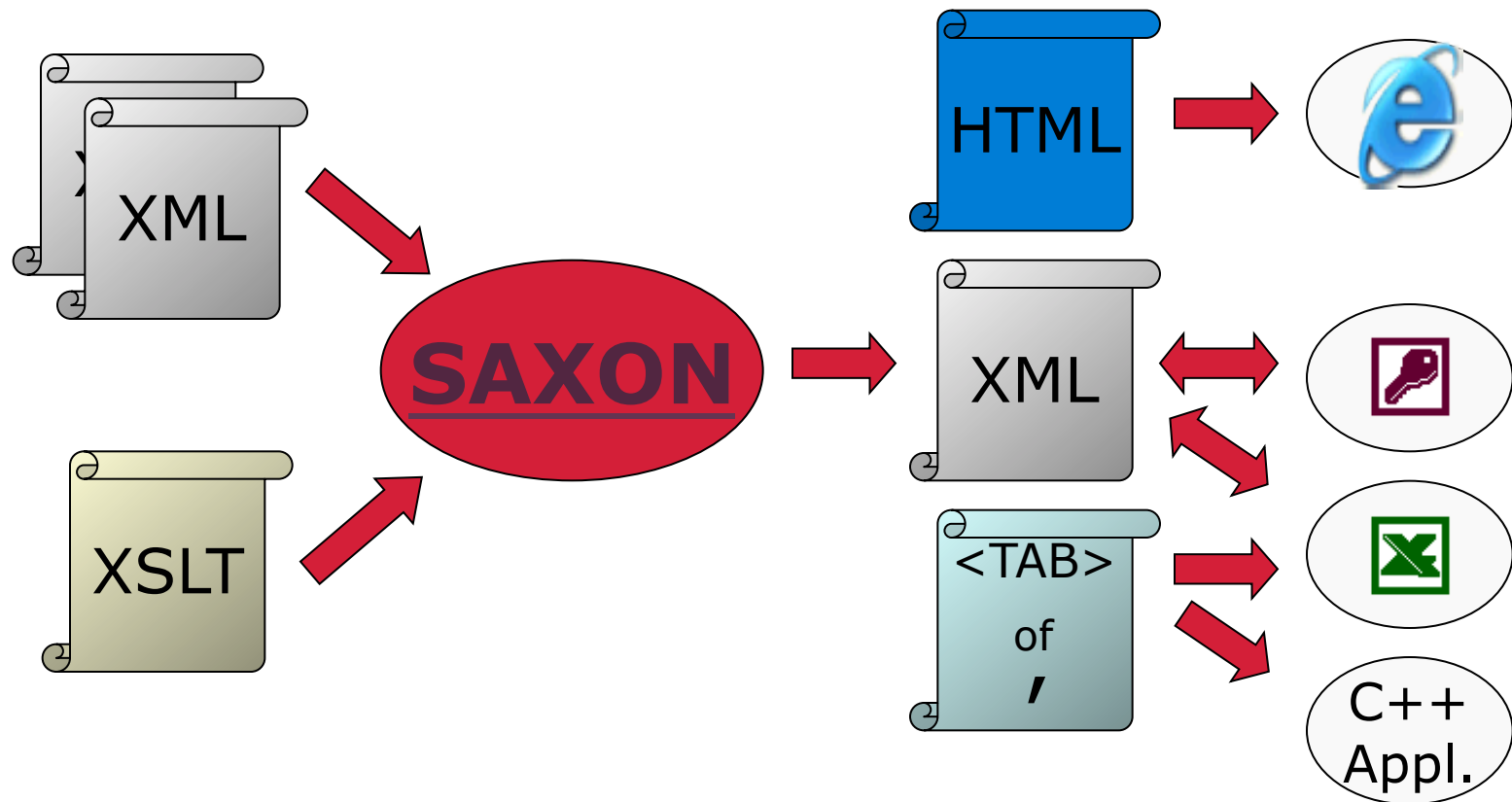
XML Parsers: voorbeelden

DOM gebaseerde tree-bewerking: zie samples



XML Parsers: voorbeelden

SAX parser



XML Parsers: DOM – lezen (PHP)

```
$dom = new DOMDocument();
// Please don't bother with whitespace
$dom->preserveWhiteSpace = false;

// read from a file
if (!$dom->load($file))
{ //error }

//vraag de root op
$root = $dom->documentElement;
//en vervolgens de domTree vanaf de root
domTree($root, 0);
...
```

XML Parsers: DOM - lezen

```
function domTree($node, $depth) {
...
if ($node->nodeType == XML_ELEMENT_NODE)
{ // element, dan alle children doorlopen
  // eerst attributen
  foreach ($node->attributes as $child)
    {domTree($child,$depth);}
  // dan alle children
  foreach ($node->childNodes as $child)
    {domTree($child,$depth+1);}
}
...
}
```

Demo & Tussenoefening

- *Demo's JavaScript*
- *Demo's C#*
- *Tussenoefening*

XML Parsers: SAX-1 (PHP)

```
$parser = xml_parser_create();
xml_set_element_handler($parser, 'start_element',
    'end_element');
xml_set_character_data_handler($parser, 'character_data');

if (!($fp = fopen($filename, 'r'))) {
    $blockSize = 4 * 1024;          // read in 4 KB chunks
    while($data = fread($fp, $blockSize)) {
        if (!xml_parse($parser, $data, feof($fp))) {
            // an error occurred; tell the user where
            echo 'Parse error:'.xml_error_string($parser).
                "at line".xml_get_current_line_number($parser);
        }
    }
    fclose($fp);
}
xml_parser_free($parser);
```

XML Parsers: SAX-2

```
function start_element($parser, $name, &$attributes) {
    ... doe wat je aan het begin van een element moet doen
}
function end_element($parser, $name) {
    ... doe wat je aan het einde van een element moet doen
}
function character_data($parser, $data) {
    ... doe wat je met character data moet doen
}
function processing_instruction($inParser, $inTarget, $inCode)
{
    ... doe wat je met processing instructions moet doen
}
function external_entity_reference($inParser, $inNames,
    $inBase, $inSystemID, $inPublicID) {
    ... doe wat je met external entity references moet doen
}
```

XML en PHP 4 / 5

- XML ondersteuning in PHP4 'gebrekkig'
- In PHP5 opnieuw geschreven:
 - *Stabieler*
 - *Consistente modules (gebaseerd op zelfde library)*
 - *Diverse uitbreidingen*
 - *Bevat:*
 - *DOM*
 - *SAX*
 - *SimpleXML*
 - *XSLT*
 - *Validatie*

XML genereren

PHP kan ook XML genereren ipv HTML:

- Via header()-functie MIME-type omzetten naar "text/xml" ipv (standaard) "text/html"
- Vervolgens XML-tags uitspugen mbv PHP
- B.v. te gebruiken voor RSS
(Really Simple Syndication of Rich Site Summary)

XML tonen met XSLT

- on-the-fly gegenereerde XML tonen

```
echo "<?xml-stylesheet href='aai.xml' type='text/xsl' ?>";
```

- XML-file of XML-stream tonen (PHP5)

```
// load XML-source
$xml = new DOMDocument;
$xml->load('aai.xml');
// load XSL-source
$xsl = new DOMDocument;
$xsl->load('aai.xsl');
// create XSLT-processor and attach XSL
$proc = new XSLTProcessor;
$proc->importStyleSheet($xsl);
// and do it!
echo $proc->transformToXML($xml);
```

- Demo's

PHP5 & XML

- SimpleXML
 - *Behandelt XML-DOM als PHP-OO*
- Nuttige artikelen (zie BB)
 - *XML in PHP 5 - What's New? (ZEND)*
 - *PHP 5 Meets XML and the DOM (Trachtenberg)*
- RSS:
 - *Nuttige artikelen (zie BB-Websites)*

Demo's

- Zie wk4_samples.zip
 - *PHP:*
 - XML genereren vanuit PHP
 - XML tonen mbv XSLT
 - DOM/SAX
 - *Javascript:*
 - XML tonen mbv XSLT
 - IE <-> FF <-> Sarissa
 - *C#*
 - XML schrijven
 - XML lezen mbv DOM
 - XML valideren
 - XML tonen mbv XSLT

Praktijktoets

- Praktijktoets in lesweek 8 of 9:
 - *toetsstof = lesstof van eerste 3 workshops*
 - DTD, Schema, XSL(T)
 - *'open boek', maar GEEN generatoren e.d. !!!*
 - Big Brother is watching you ...
 - NotePad++ is maximum aan editors!
 - werken met XML Tools Plugin voor NotePad++
 - *opgaven zullen vergelijkbaar zijn met huiswerk*
 - *proeftentamen in course aanwezig*
 - *oefenen ... oefenen ... oefenen*
 - *levert je 2 EC op (code WEBS4-PR onder cursus IIIN-WEBS4)*

Vragen ???