XML – DOM
Document Object Model

Markus Durzinsky *
Student der Otto-von-Guericke-Universität Magdeburg

23. April 2004

Inhaltsverzeichnis

*eMail: Mardur@gmx.net

1 Einführung

Dies ist das ausgeschriebene Script zum Vortrag *XML – Document Object Model*, gehalten am 20. Januar 2004. Der Vortrag entstand im Rahmen des Proseminars *XML - Grundlagen und Anwendungen* im Wintersemester 2003 an der Otto-von-Guericke-Universität Magdeburg. Das Proseminar wurde betreut von Prof. Dr. Dietmar Rösner und Dr. Sylke Krötzsch.

Die Vorträge des Proseminars, die diesem vorausgegangen sind, befassten sich inhaltlich mit den Themen

- allgemeine Struktur von wohlgeformten XML-Dokumenten
- Vorgabe der Struktur durch eine Document Type Definition, welche die Verschachtelung der Elemente und die Attribute festlegt
- XML Schema zur genaueren Vorgabe der Struktur und des Inhalt einer XML Datei als mit DTDs
- Umwandlung von XML-Dokumenten mit XSLT und XPath

In diesem Vortrag geht es um die Möglichkeit XML-Dokumente ganz allgemein einzulesen und zu bearbeiten. Dabei stehen nicht die Interna oder die Implementierung eines Parser[1] im Vordergrund, sondern dessen Verwendung.

2 Modelle zum Einlesen von XML Dateien

Da XML-Dokumente reine Textdateien sind, ist es auch möglich sie einfach zeilenweise als Text einzulesen. Den Text könnte man dann zum Beispiel mit Hilfe von regulären Ausdrücken durchsuchen.

Ein kleines XML Fragment, dass immer als Beispiel herhalten muss ist in Abbildung 1 angegeben.

```
<person alter="25">
  <name>Peter Schmidt</name>
  <beruf>Student</beruf>
</person>
```

Abbildung 1: Ein kurzes XML Fragment

Für eine strukturiertere Methode des Einlesens, wird man aber auf einen Parser zurückgreifen, der die Eigenschaften eines XML-Dokuments berücksichtigt. Bei solchen Parsern gibt es im Wesentlichen zwei verschiedene Modelle, nach denen die Daten aufbereitet werden.

2.1 ereignisorientiertes Modell

Zum einen können die Elemente, Attribute und Textsegmente einfach in der Reihenfolge weitergegeben werden, in der sie im Dokument auftauchen. Ausser

[1] mit Parser ist ein Programm gemeint, dass nicht nur die Wohlgeformtheit eines XML-Dokuments überprüft, sondern auch den Inhalt in geeigneter Form weitergibt

zur Überprüfung der Wohlgeformtheit, muss ein solcher Parser keine weiteren Daten speichern. Damit ist der Speicherbedarf des Parsers minimal. Des weiteren erfolgt die Weitergabe der Daten vom Parser an die Anwendung ohne Wartezeit sofort nach dem Öffnen der Datei.

Eine Spezifikation für ein solches Modell bildet das *Simple API for XML*, oder kurz SAX. Wie dass am Beispiel aussieht, zeigt Abbildung 2.

```
startDocument()
  attribute("alter", "25")
  startElement("person")
    startElement("name")
      charData("Peter Schmidt")
    endElement("name")
    startElement("beruf")
      charData("Student")
    endElement("beruf")
  endElement("person")
endDocument()
```

Abbildung 2: SAX Ausgabe

Dabei entspricht jede Zeile einem Aufruf der entsprechenden Methode. Die Anwendung nimmt diese Daten entgegen und muss die Speicherung selbst übernehmen.

2.2 Baum-Modell

Zum anderen bietet sich die Variante an, das Dokument in einer Baumstruktur abzulegen, da wohlgeformte XML-Dokumente von sich aus eine solche Struktur besitzen. Die Spezifikation von DOM verwendet dieses Baum-Modell. Den Baum des Beispiels sieht man in Abbildung 3.

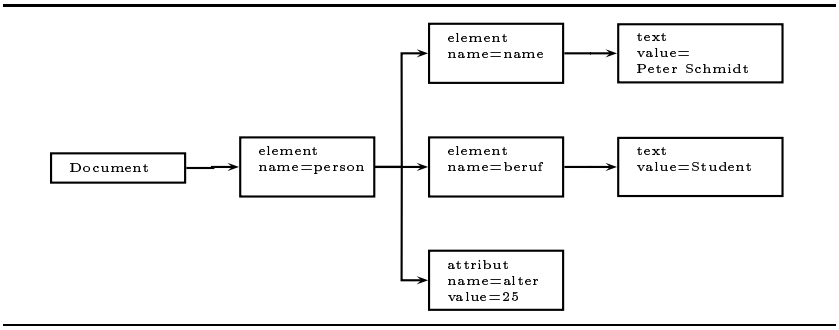

Abbildung 3: DOM Baum

Im Gegensatz zu SAX muss ein DOM Parser das gesamte Dokument im Speicher halten. Dies macht das Modell für sehr grosse Dokumente unbrauchbar. Auch muss die Anwendung warten bis das Dokument vollständig verarbeitet wurde. Nach dem Einlesen hat man dann aber vollen und direkten Zugriff auf

die einzelnen Elemente. Des weiteren besteht die Möglichkeit das Dokument beliebig zu verändern und wieder zu speichern.

Welches dieser Modelle man für eine eigene Anwendung nutzen sollte, hängt deshalb davon ab, wie gross die XML Datei ist und was man damit machen möchte. Eine Auflistung der Vor- und Nachteile enthält Tabelle 1.

Simple API for XML		**Document Object Model**	
$\oplus$	Abarbeitung beginnt sofort	$\ominus$	lange Wartezeit beim Einlesen
$\oplus$	kaum Speicherbedarf	$\ominus$	hoher Speicherbedarf
$\ominus$	Dokument einmal durchlaufen	$\oplus$	beliebiger Zugriff auf Elemente
$\ominus$	nur Lesen	$\oplus$	Änderungen möglich

Tabelle 1: Vor- und Nachteile

3 Document Object Model

Das Document Object Model ist eine Spezifikation zur allgemeinen Beschreibung von Dokumenten in einer Baumstruktur, nicht nur aber insbesondere XML-Dokumente. Dazu wird eine Reihe von Schnittstellen[2] vorgegeben, welche jeweils bestimmte Teile dieses Dokuments representieren. Es wird vorgeschrieben, welche Daten eines XML-Dokuments in welche Instanzen der Schnittstellen überführt werden. Es wird aber nicht festgelegt, wie dies geschieht.

Level-weise entwickelt ...

4 Schnittstellen

hier kommt eine Übersicht rein.

Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein.

4.1 Hilfsschnittstellen in DOM

ein bisschen `NodeList` und `NamedNodeMap`

Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt

[2] Eine Schnittstelle (engl. interface) in der objektorientierten Programmierung ist eine Klasse, die noch implementiert werden muss. Es sind nur schon die Eigenschaften und Methoden vorgegeben, sowie deren Semantik.

später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein.

4.2 die Schnittstelle `Node`

Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein.

4.3 spezialisierte Schnittstellen

Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein.

5 Anwendungsbeispiel mit JavaScript

Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein.

6 Anhang – Spezifikation der Schnittstellen

6.1 `Node`

Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein.

6.2 NodeList

Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein.

6.3 NamedNodeMap

Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein.

6.4 Document

Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein.

6.5 DocumentType

Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein.

6.6 Attr

Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier

kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein.

6.7 Notation

Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein.

6.8 Entity

Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein.

6.9 EntityReference

Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein.

6.10 ProcessingInstruction

Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein.

6.11 `CharacterData`

Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein.

6.12 `Comment`

Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein.

6.13 `Text`

Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein.

6.14 `CDATASection`

Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein. Hier kommt später noch Text rein.

Literatur

[1] Elliot Rusty Harold: *XML Bible*, IDG Books Wolrdwird, Inc., 1999

[2] Elliot Rusty Harold, W. Scott Means: *XML in a nutshell*, Deutsche Übersetzung der 2. Auflage 2003, O'Reilly Verlag GmbH & CO. KG