

Aufgabenblatt 01

Quelle: <http://edu.cs.uni-magdeburg.de/EC/lehre/wintersemester-2015-2016/dokumentverarbeitung/uebungen/gruppe-2-fr-13-15-uhr/uebungsblaetter/aufgabenblatt-1> (21.10.2015)

Aufgabe 1: Dokumentklasse und Dokumentinstanz

Beschreiben Sie drei verschiedene Dokumentklassen. Geben Sie zu jeder Dokumentklasse mindestens ein Beispiel für eine Dokumentinstanz an!

Dokumentenklasse:

- spiegelt die Art des Objektes wieder, das dokumentiert werden soll
- nicht alle beziehen sich auf Objekte, sondern sie liegen dann als eigene Systemobjekte vor

Dokumenteninstanz:

- beinhaltet Elemente, Attribute, Entitäten und Zeichenfolgen
- Article:
 - z.B. wissenschaftlicher Artikel in einer Zeitschrift
 - Titel auf keiner eigenen Seite, also da wo der Artikel ist
 - Seitenzählung
 - einseitiges Layout
 - Abbildungen, Tabellen haben fortlaufende Nummerierung
- Book:
 - Titel ist auf einer eigenen Seite
 - Seitenzählung (z.B. römischen Ziffern)
 - in einem wissenschaftlichen Buch gibt es eine kapitelweise Nummerierung der Abbildung und Tabellen
- Letter:
 - hat keine bestimmte Ebenen, außer die Merkmale, wie Absender, Anschrift, Gruß
 - Bsp. Liebesbrief

Aufgabe 2: Dokumentanalyse

Sehen Sie sich das folgende Dokument an (Anhang 01). Beschreiben Sie die vorkommenden Strukturelemente im Dokument.

Strukturelement:

- Beschreibung (Anzahl in ein Bericht / in alle 3 Bereichen)
- Ueberschrift: Bezeichnung des Berichts ($1/n +$)
- Staerke: Staerke des Erbebens ($1/n +$)
- Datum/Zeit: Wann war das Erdbeben mit Untergliederung in verschiedenen Zeitzonen ($1/n +$)
- Koordinaten: wo war das Erdbeben in Laengen- und Breitengrade ($1/n +$)
- Tiefe: Daten zur Tiefe in Kilometer (Meilen) ($1/n +$)
- Region: in welchen Land war das Erdbeben ($1/n +$)
- Distanzen: Distanz zu anderen Staedten ($n + /n +$)
- Unsicherheiten: Bewegung des Erdbebens in horizontal und depth ($1/n +$)
- Parameter: naehere Daten zum Erdbeben (NST, Nph, Dmin, Rmss, Gp) ($1/n +$)
- M-type: (Mb, MI, Version) ($1/n +$)
- Quelle: wo gefunden ($1/n +$)
- Event ID: Nummer zur Identifikation ($1/n +$)
- Link: Verweis ($1/n +$)

Aufgabe 3: Elemente einer DTD

Beschreiben Sie, welche Elemente bei einer DTD verwendet werden dürfen. Erstellen Sie für das Dokument aus der vorangegangenen Aufgabe (Anhang 01) eine DTD.

Elemente:

- EMPTY - für keinen Inhalt
- ANY - für beliebigen Inhalt
- , - für Reihenfolgen
- | - für Alternativen (im Sinne „entweder...oder“)
- () - zum Gruppieren
- * - für beliebig oft
- + - für mindestens einmal
- ? - für keinmal oder genau einmal
- Wird kein Stern, Pluszeichen oder Fragezeichen angegeben, so muss das Element genau einmal vorkommen

Attribute:

- Attributliste enthaelt den Namen, den Typ und Vorgaben eines Attributes

- kann angegeben werden , ob ein Attribut vorkommen muss oder nicht oder sogar eines festen Wert enthaelt und welcher Wert als Standardwert benutzt wird

Entity:

- benannte Abkuerzung für eine Zeichenkette oder ein externes Dokument, innerhalb der DTD benutzt werden kann

PCDATA:

- ein Schluesselwort für einen Textblock, der auch weitere Anweisungen an den Parser enthalten kann

<!DOCTYPE text [<!ELEMENT text -- (überschrift, stärke, datum/zeit, koordinaten, tiefe, region, distanzen, unsicherheiten, parameter, m-type, quelle, event id, link)>

<!ELEMENT überschrift -- (#PCDATA)+>

<!ELEMENT stärke -- (#PCDATA)+>

<!ELEMENT datum/zeit -- (#PCDATA)+>

<!ELEMENT koordinaten -- (#PCDATA)+>

<!ELEMENT tiefe -- (#PCDATA)+>

<!ELEMENT region -- (#PCDATA)+>

<!ELEMENT distanzen -- (#PCDATA)+>

<!ELEMENT unsicherheiten -- (#PCDATA)+>

<!ELEMENT parameter -- (#PCDATA)+>

<!ELEMENT quelle -- (#PCDATA)+>

<!ELEMENT event id -- (#PCDATA)+>

<!ELEMENT link -- (#PCDATA)+>]>

Aufgabe 4: Minimierung

Kann man bei den folgenden Beispielen entscheiden, ob es sich um potentiell gültige SGML-Dokumente handelt, ohne dass SGML-Deklaration und DTD bekannt sind?

a) < a >< b >< a >< b >

b) < a >< /a >< b >< /b >

c) $\langle a \rangle \langle a \rangle /a \rangle /b \rangle$

d) $\langle a \rangle \langle b \rangle /a \rangle \langle b \rangle$

e) $\langle b \rangle /a \rangle \langle a \rangle \langle c \rangle /b \rangle \langle b \rangle$

Definieren Sie, falls möglich, für jede der Konstruktionen eine DTD, die diese zulässt. Die Elemente $\langle a \rangle$, $\langle b \rangle$ und $\langle c \rangle$ sollen dabei von einem Wurzelement $\langle test \rangle$ umschlossen werden, dessen Start- und Endtag vorhanden sein müssen.

a) nein, da es keine Endmarkierung gibt

b) ja, da es zu jeder Startmarkierung auch eine Endmarkierung gibt

$\langle test \rangle \langle a \rangle \dots \langle /a \rangle \langle b \rangle \dots \langle /b \rangle \langle /test \rangle$

c) nein, da zweimal das $\langle a \rangle$ kommt und dann die erst die Endmarkierung

d) ja, Parser nimmt Ende eines Elements an, wenn Endemarkierung eines Elements gefunden, zu dessen Inhalt betroffenes Element gehört

$\langle test \rangle \langle a \rangle \langle b \rangle \dots \langle /b \rangle \langle /a \rangle \langle b \rangle \dots \langle /b \rangle \langle /test \rangle$

e) nein, da die Endmarkierung zuerst kommt $\langle /a \rangle$, ohne eine Startmarkierung vorher

Aufgabe 5: Inhaltsmodell

Erstellen Sie drei Beispiele, die dem folgenden SGML-DTD-Fragment entsprechen (nehmen Sie an, dass die Inhaltsmodelle der in book enthaltenen Elemente jeweils (#PCDATA) sind):

element book - - ((editor|author)+, title+, subtitle, ed?, series?,place, publ, year, isbn?, issn?, sign,keyword, url, (comment? & note? & abstract?))>

Bsp.1:

- Gerald Teschl, "Mathematik für Informatiker: Band 1, Heidelberg, Springer-Verlag Berlin, 2013

Bsp.2:

- Gerald Teschl, "Mathematik für Informatiker: Band 1", "Diskrete Mathematik und Lineare Algebra ", 4. Auflage, Heidelberg, Springer-Verlag Berlin, 2013, ISBN-13: 978-3642379710, http://www.amazon.de/gp/product/3642379710?keywords=mathematik%20für%20informatiker&qid=1445522744&ref=sr_1_1&sr=8-1

Bsp.3:

- John Barwise, John Etchemendy, "Sprache, Beweis und Logik", "Aussagen- und Prädikatenlogik", Paderborn, mentis Verlag GmbH, 2005, "Must read"