

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. ein bestimmter 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
 - z.B. "Die kürzeste Geschichte der Zeit"
- Letter:
 - hat keine bestimmte Ebenen, außer die Merkmale, wie Absender, Anschrift, Gruß
 - z.B. Meine Kündigung des Mietvertrags

Quelle: http://www.students.uni-mainz.de/julschue/LaTeX/03_Dokumentenklassen.pdf

(22.10.2015)

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)>  
>
```

```
<!DOCTYPE earthquakes [  
  <!ELEMENT earthquake -- (Titel, Magnitude, Date-Times, Location,  
    Depth, Region, Distances, Uncertainty, Parameterlist, M-Type,  
    Version, Sourcelist, EventID, URL)+>  
  <!ELEMENT Titel - - ("Earthquake Details")>  
  <!ELEMENT Magnitude - - (#PCDATA)>  
  <!ELEMENT Date-Time - - (UTC, LT, (TimezoneTime)+)>  
  <!ELEMENT UTC - - (#PCDATA)>  
  <!ELEMENT LT - - (#PCDATA)>  
  <!ELEMENT TimezoneTime - - (#PCDATA)>  
  <!ELEMENT Location - - (#PCDATA)>  
  <!ELEMENT Depth - - (#PCDATA)>  
  <!ELEMENT Region - - (#PCDATA)>  
  <!ELEMENT Distances - - (DistanceFrom)+>  
  <!ELEMENT DistanceFrom - - (#PCDATA)>  
  <!ELEMENT Uncertainty - - (horr, depth)>  
  <!ELEMENT horr - - (#PCDATA)>  
  <!ELEMENT depth - - (#PCDATA)>  
  <!ELEMENT Parameterlist - - (NST?, param, param, param, param, M-  
    Type, Version)>  
  <!ELEMENT NST - - (#PCDATA)>  
  <!ELEMENT param - - (#PCDATA)>  
  <!ELEMENT M-Type - - (#PCDATA)>  
  <!ELEMENT Version - - (#PCDATA)>  
  <!ELEMENT Sourcelist - - (Source)+>  
  <!ELEMENT Source - - (#PCDATA)>  
  <!ELEMENT EventID - - (#PCDATA)+>  
  <!ELEMENT URL - - (#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) `< a >< a >< /a >< /b >`

d) `< a >< b >< /a >< b >`

e) `< b >< /a >< a >< c >< /b >< b >`

Definieren Sie, falls möglich, für jede der Konstruktionen eine DTD, die diese zulässt. Die Elemente `< a >`, `< b >` und `< c >` sollen dabei von einem Wurzelement `< test >` umschlossen werden, dessen Start- und Endetag vorhanden sein müssen.

Vorbemerkung: FÜR ALLE BEISPIELE WÜRDEN AUCH GELTEN: -> dann aber nichtmehr eindeutig

```
<!DOCTYPE test [  
  <!ELEMENT test - - (a|b|c)+ >  
  <!ELEMENT a o o (#PCDATA) >  
  <!ELEMENT b o o (#PCDATA) >  
  <!ELEMENT c o o (#PCDATA) >  

```

a) ja

```
<!DOCTYPE test [  
  <!ELEMENT test - - (a|b)+ >  
  <!ELEMENT a - o (#PCDATA) >  
  <!ELEMENT b - o (#PCDATA) >  

```

b) ja

```
<!DOCTYPE test [  
  <!ELEMENT test - - (a,b) >  
  <!ELEMENT a - - (#PCDATA) >  
  <!ELEMENT b - - (#PCDATA) >  

```

c) ja, dann aber nichtmehr eindeutig zuzuordnen. z.B. `<a>Hallo Welt`

```
<!DOCTYPE test [
  <!ELEMENT test - - (a|b)+ >
  <!ELEMENT a - o (a|PCDATA) >
  <!ELEMENT b o - (#PCDATA) >
]>
```

d) ja gültig. Eindeutig auflösbar auch ohne ``:

```
<!DOCTYPE test [
  <!ELEMENT test - - (a,b) >
  <!ELEMENT a - - (b) >
  <!ELEMENT b - O (#PCDATA) >
]>
```

e) sh. Universal DTD am Dokumentenanfang

oder:

```
<!DOCTYPE test [
  <!ELEMENT test - - (b)+ >
  <!ELEMENT a o o (c|PCDATA)* >
  <!ELEMENT b - o (a|PCDATA)* >
  <!ELEMENT c - o (#PCDATA) >
]>
```

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?))>
```

```
<book>
  <author> Andy Weir
  <title> The Martian </title>
  <subtitle> A Novel
  <place> USA
  <publ> BOWY
  <year> 2015
  <isbn> 978-1101905555
  <keyword> Fiction
  <keyword> Science Fiction
  <keyword> Action & Adventure
  <comment> "Brilliant..."--Wall Street Journal
</book>
```

```
<book>
  <author> Dr. Anton Hammer
  <author> Dr. Hildegard Hammer
  <author> Dr. Karl Hammer
  <title> Physikalische Formeln und Tabellen </title>
  <ed> 8
  <place> München
  <publ> Lindauer Verlag
  <year> 2002
  <isbn> 3-87488-182-2
  <keyword> Physik
  <keyword> Gymnasium
  <url>http://www.amazon.de/Physikalische-Formeln-Tabellen-Hammer/dp/3874881822
</book>
```

```
<book>
  <author> Figgen
  <author> Plath
  <author> Rothfischer </title>
  <title> Verkehrsflugzeuge
  <subtitle> Das neue Typen-Taschenbuch
  <place> München
  <publ> GeraMond Verlag
  <year> 2008
  <isbn> 978-3-7654-7011-0
  <url> http://geramon.de
  <note> Aktualisierte Neuauflage!
</book>
```