**Schemasprachen für XML-Dokumente: XML DTD**

**Was ist eine Schemasprache für XML?**

Definition: Festlegung von erlaubten Strukturen für eine Klasse von Dokumenten (**Instanzen**) mit formalen Mitteln

Wozu: Gegengewicht zu "X"ML, Freiheit des Vokabulars

* Kommunikation: Eingrenzen der Freiheit des Vokabulars in einer Community / Domäne
* Überprüfung / Validierung von Vereinbarungen
* syntaxgestützte Eingabeunterstützung
* Kenntnis der möglichen Strukturen kann Suchen, Filtern erleichtern

Was wird festgelegt?

* Vorrat an Strukturelementen und Attributen
* Vorschriften zu ihrer Kombination, auch in Bezug auf Text
* Definition von Integritätsbedingungen (z.B. Eindeutigkeit von IDs und Referenzen)

Beispiele:

* XML Document Type Definition (DTD)
* XML Schema
* Relax NG
* Schematron

**XML Document Type Definition (DTD)**

Vokabular für **Elementnamen** und **Attributnamen**, Beziehungen von Elementen und Attributen einschließlich **Identity Constraints** werden formal festgelegt, mit Hilfe von

* Elementdeklarationen
* Attributdeklarationen

XML DTD macht XML zu **Metasprache** für **Markup-Sprachen**

Aufbau von XML DTD als **Katalog** (Pattern **Salami-Slice**)

Die Abkürzung DTD steht für **Document Type Definition**. Eine DTD besteht aus **Elementdeklarationen** und **Attributdeklarationen**. Die **Typdefinition** für ein Element mit Namen XXX in einer DTD setzt sich zusammen aus **Elementdeklaration** und **Attributdeklaration** für das Element XXX. Eine Elementdeklaration besteht aus **Elementname** und **content model** (**Inhaltsmodell**). Ein Inhaltsmodell (content model) wird gebildet aus **Elementnamen** und den folgenden Operatoren: **+ \* | ?** **,**. Dazu kommen auf der Syntax-Ebene **(** und **)** zur Komposition von Inhaltsmodellen. Eine Attributdeklaration besteht aus **Elementnamen**, **Attributnamen**, **Datentyp** und **Defaultregelung**. Hinzukommen kann ein optionaler Defaultwert und die Vorgabe, dass dieser Defaultwert nicht geändert werden kann (#fixed). Ein XML-Dokument, das einer DTD entspricht, ist **valide**. Software, die diese Eigenschaft überprüft, heißt **Parser** (**XML-Prozessor**).

**Wohlgeformtheit und Validität**

XML-Dokumente müssen **wohlgeformt** sein (Überprüfung durch alle Parser)

* DTD optional
* nur ein Wurzelelement
* korrekte Klammerung
* keine mehrfachen Attribute in Elementen

Zusätzlich: **Validität** (Überprüfung durch validierende Parser)

* Referenzen definiert
* Instanz erfüllt Deklarationen von Elementen und Attributen (Typen, Defaults)
* validierender Parser setzt Defaultwerte ein und erkennt / validiert Attributtypen wie ID, IDREF und ENTITY

Beispiel:

