**3 Säulen des Document Engineering**: Elektronisches Publizieren, Document Management, Knowledge Management

**Model** **der** **strukturierten** **Element**: Trennung von Daten und Verwendungszweck

**Inhaltstext, logische Struktur**: Markierung mit dem Grund direkt im Text

**Rolle von XML**: Sprache zur Kodierung von strukturierten Dokumenten, Zentrales Format (Drehscheibencharakter)

**Unicode**: Ziel ist es den weltweiten Datenaustausch zu vereinfachen. Universaler Zeichensatz für alle Schriften der Welt. 17 Ebenen zu 256 Kacheln mit 256 Positionen

**Metasprache** **für** **Markupsprachen**: Sprache zum Definieren von Sprachen. XML stellt Regeln zur Verfügung, mit denen man zb HTML definieren kann.

Ein XML Dokument besteht aus Element, Attribut und Text

**eingebettetes Markup** Textstrom in Unicode/UTF-8 aus Inhaltstext und eingebetteten Tags, die die logische Struktur repräsentieren

**Markupsprache**: Markiert Textbereiche an Anfang und Ende mit eingebetteten Tags

**Tag**: Syntax wie bei HTML

**Wohlgeformtheit**: Tags müssen immer geschlossen werden (auch ohne Inhalt) und können hierarchisch genestelt werden. Einhaltung der XML-Syntax, DTD optional, nur ein Wurzelelement

**Validität**: Überprüfung durch validierenden Parser, setzt Defaultwerte ein, Referenzen definiert

**Entity**-**Referenz**: Entities sind definierte Kürzel (&lt = Kleiner-als, &amp = &, &gt = größer-als, &quot = „, &apos = ‚)

**Aufgaben** **eines** **Parsers**: Syntaxüberprüfung, Abstraktion von "syntaktischem Zucker", Überprüfen der Wohlgeformtheit, Auswertung von DTD

**Namensraum**-**Name**: Namen in XML-Dokumenten mit einer URI zu verknüpfen

**qualifizierter** **Name** Präfix + lokaler Name

**erweiterter** **Name** Namensraum-Name + lokaler Name

**Datenmodelle** **für** **XML**-**Dokumente**: XML Information Set, XDM

**Lokalisierung** **von** **XML** **Fragmenten**: XPath

**XML** **Information** **Set**: Definiert ein Datenmodell für XML in Form von Komponenten, mit Atrributen

**Datenmodell**: Legt fest, welche Bestandteile eines XML Dokuments wesentlich sind, definiert durch Information Items und Attribute, W3C: XQuery, XPath, InfoSet

**XDM**: Strukturen von Knoten/Information Items. XML Dokument als Baumstuktur, definiert Datenmodell auf denen XPath, XQuery, XSLT operieren

**XPath**: Teilsprache von XQuery, navigation, basiert auf XDM

**Location** **Path**: Teil von XPath, schrittweises navigieren durch den Baum

**Aufbau** **von** **Step** (3 Teile): Achse, Knotentest, Prädikate (Optional)

**Syntax** **für** **Step**: **Achse::Knotentest [ Prädikat ] [ Prädikat ] [ Prädikat ]…**

**Schemasprachen**: Festlegung von erlaubten Strukturen für eine Klasse von Dokumenten(Instanz) mit formalen Mitteln. Vorrat an Strukturelementen und Attributen, Vorschrift zur Kombination, DTD, XML Schema, Schematron, Relax NG

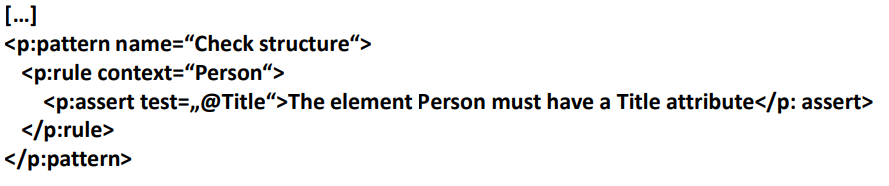
**Document** **Type** **Definition** (**DTD**): Vokabular für **Elementnamen** und **Attributnamen**, Beziehungen von Elementen und Attributen einschließlich **Identity Constraints,** Element und Attributdeklarationen, einfache Konzepte, eigene Syntax

**Element**: Elementname und Inhaltsmodell inklusive Operatoren

**Attribut**: **Elementnamen**, **Attributnamen**, **Datentyp** und **Defaultregelung**

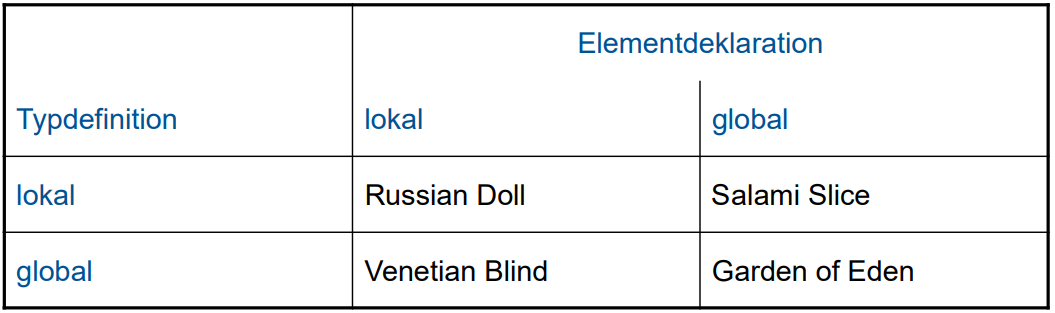
**XML** **Schema** : weit verbreitet, komplex, mächtiger als DTD, adressiert Datentypen und OO Konzepte

**Schematron**: regelbasiert, deklarativ nicht generativ, Beschreibt Regeln für Komponenten eines XML-Dokuments



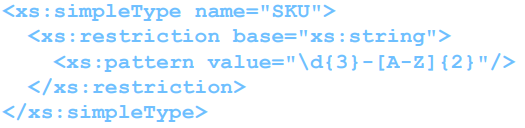
**Datentyp**: Werte eines Datentyps sind bestimmte Zeichenketten: Strings mit bestimmten Bildungsgesetzen und Gleichheitsregelungen, anwendbar auf Attribute, Textinhalt

Strukturtyp (Complex Typ): anwendbar auf Elemente, hierarchische Struktur



Möglichkeiten von selbstdefinierenden Datentypen: Restrictions, Liste und Union

**Facetten**: eigene Datentypen definieren, Facetten schränken Wertebereich ein, pattern beispiel für Facetten



**Identity** **Constraints**: Identifizierung von Objekten, referenzierung von Objekten, Datentyp ID und IDRef und Konzepte key (beliebig viele Attribute, mächtiger) und keyref

**XQuery**: Abfragesprache für XML

**XSLT**: Transformieren von XML Dokumente in andere Formate, Sprache ist deklarativ, XSLT Prozessor als Interpreter

**XSLT** **push** vs **pull**: apply template vs for each and value of