# Zweite Vorlesung

Wohlgeformtheit: Grundlegende syntaktische Regeln

Siehe Einführung

XML: Instanzdokument

‚\*‘ kein, ein oder mehrere

Element kann mehrmals vorkommen, Attribut nicht

Entities

Allgemeine Entities

Parameter Entities

Allgemein: Vordefiniert, erlaubt Verwendung von Sondernzeichen

Zahlendefinierten: Zeichen, die nicht am Eingabegerät verfügbar sind; Dezimal oder Hexadezimal, wenn man sie im Text repräsentieren will

Benutzerdefiniert: Makro, Platzhalter für Textfragmente, die wieder vorkommen; Intern

Extern: XML-Fragment in eigenes File; kann man im Instanzdokument an bestimmten Stellen einsetzen

Fragment darf keine DTD enthalten

Am Ende muss das Gesamte wieder Wohlgeformt sein

Parameter: Makros in der DTD, kann man nur in der DTD verwenden

.mod als Namensende, um darauf hinzuweisen, dass es eine Parameter-Entitie ist

Man kann erneut definieren: Mehrere DTDs; Namen der Internen haben immer Vorrang vor Externen

DTD-Nachteile: M2-54

Namensräume:

Vermeidung von Namenskonflikten

URI: URL, beschreibt gewissen Aufbau, wie eine Adresse aussehen muss

Xmlns: Namensraum, denen die Elemente zugeordnet sind

Präfix kann definiert werden, der eine URL repräsentiert

Diesen Präfix klebt man nun vor jedes Element

Namensraum mit Default-Namensraum

Man verwendet keinen Präfix

Jedes Element, und dessen Kindelemente, werden dem Default-Namensraum zugeordnet

Man kann auch in der Hierarchie die Namensräume überschreiben

Bei Präfix wird das Element explizit zugeordnet

Wenn man eines nicht zuordnet, gehört es auch nicht dazu

Bei Default: Werden Elemente automatisch zugeordnet, Attribute aber nicht

Attribute müssen explizit zugeordnet werden per Präfix

XPATH 2.0

Abfragesprache

Hierarchische Dokumente abfragen

SQL: Mengenorientiert

XPATH: Pfadorientiert auf Hierarchien

Basis für XQUERY (Eigentliche Abfragesprache)

Man kriegt als Ergebnismenge immer eine Sequenz von Knoten

XQUERY: FLWOR-Ausdrücke wie SQL

XSLT 2.0: Stylesheet Sprache

XML Dokument abfragen, braucht Parser

Parser nimmt Dokument, macht es zu einem Baum

Knoten haben unterschiedliche Typen (Dokumentknoten, Wurzelknoten, Elementknoten, Attributknoten, Namensraumknoten, Kommentarknoten, Textknoten)

Jede Abfrage liefert eine Sequenz

Sequenz: Enthält Items (Knoten oder Atomare Werte)

Duplikate sind in Sequenzen erlaubt

Filter mit [ ] definiert