

XML Vorlesung ETHZ, Sommersemester 2006

Organisatorisches, Überblick und Einführung

Erik Wilde

4.4.2006

<http://dret.net/lectures/xml-ss06/>

4.4.2006

XML Vorlesung ETHZ SS 2006

1

Übersicht

- Organisatorisches
- Inhalt und Ziel der Vorlesung
 - Vorstellung der behandelten Themen
- Herkunft und Motivation für XML
 - wofür sollte XML ursprünglich benutzt werden
 - wofür wird XML heute grösstenteils benutzt

4.4.2006

XML Vorlesung ETHZ SS 2006

2

Organisatorisches

- elektronische Einschreibung
 - Pflicht für Email-Versand und Testat-Vergabe!
 - <https://www.einschreibung.ethz.ch/>
- Übungen
 - tardis Account (D-ITET Suns) ist notwendig
 - Betreuung Dienstags 15⁰⁰-17⁰⁰ im ETL E11
 - andernfalls per Email an die Betreuer (Sai & Petra)
- Abgabe der Übungen
 - i.A. bis 24⁰⁰ am Montag vor der folgenden Vorlesung
 - generell keine Übungsabgabe nach Veröffentlichung der Musterlösung

4.4.2006

XML Vorlesung ETHZ SS 2006

3

Inhalt und Ziel der Vorlesung

- Inhalt: XML als Format für strukturierte Daten
 - XML als hierarchisches Informationsmodell
 - XML Schema als XML-Standard der Zukunft
 - XSLT als Transformationssprache für XML
 - Speicherung von XML in XML Datenbanken
- Ziel: Umgang mit XML lernen
 - XML primär als Informationsmodell begreifen
 - Tools für den Umgang mit XML kennenlernen
 - XML als Grundlage für viele Anwendungen sehen
- "XML is ASCII for the 21st century"

4.4.2006

XML Vorlesung ETHZ SS 2006

4

Herkunft und Motivation für XML

- XML im Vergleich zu HTML
- Motivation für die Einführung von XML
 - Geschichte des Web
 - HTML als präsentationsorientierte Sprache
 - SGML als erprobte Grundlage von HTML
- Aufbau einer Architektur mit XML
 - XML als Basis für viele andere Komponenten
 - XML ist nur der Anfang der Entwicklung
 - XML wird in verschiedenen Bereichen verwendet
 - Web, B2B, XML Protocol, Prototyping

4.4.2006

XML Vorlesung ETHZ SS 2006

5

XML und HTML

- sieht "so ungefähr" aus wie HTML
 - gleiche Basis (SGML)
 - *proven success* (SGML und HTML sind Erfolge)
 - geringere Hemmschwelle für Umsteiger
- funktioniert ähnlich wie HTML
 - gleiche Strukturierungsverfahren (Grammatiken)
 - rein textorientiertes Format (keine Binärdaten!)
- andere Zielgruppe als HTML
 - weiterverarbeitbare Information (B2B)
 - anwendungsabhängige Datenstrukturen

4.4.2006

XML Vorlesung ETHZ SS 2006

6

Hypertext Markup Language (HTML)

- Anwendung (*Application*) von SGML
- HTML SGML *Declaration*
 - bestimmt die konkrete Syntax von HTML
 - bestimmt die SGML Features von HTML
- HTML SGML *Document Type Definition (DTD)*
 - Definition der Elemente, Attribute und Grammatik
- zusätzliche Definitionen (kein SGML!)
 - Einschränkungen von Attributwerten (z.B. Zahlen)
 - Bedeutungen von Elementen und Attributen
 - alles in Prosa beschrieben (nicht formal definiert)

4.4.2006

XML Vorlesung ETHZ SS 2006

7

SGML

- *Standard Generalized Markup Language*
 - definiert in ISO Standard 8879 (1986)
- Markup Language trennt Struktur und Text
 - Struktur ist syntaktisch identifizierbar
- Vorteile einer Markup Language
 - einfache Repräsentation (Erzeugung "von Hand")
 - Austauschbarkeit als rein textbasierte Dokumente
- Definition beliebiger Dokumenttypen
 - anwendungsspezifische Dokumenttypen
 - gemeinsamer zugrundeliegender Mechanismus
 - gemeinsam verwendbare Software

4.4.2006

XML Vorlesung ETHZ SS 2006

8

Markup Beispiel (XML \Rightarrow SGML)

```
<?xml version="1.0" ?>
<!DOCTYPE kurs SYSTEM "kurs.dtd">

<kurs>
<titel kurz="XML">XML - Grundlagen und Umfeld</titel>

<referent email="xml@dret.net"
           homepage="http://dret.net/">
  <vorname>Erik</vorname>
  <name>Wilde</name>
  <organisation homepage="http://www.tik.ee.ethz.ch/">ETH
  Zürich</organisation>
</referent>

<referent> ... </referent>
<inhalt> ... </inhalt> </kurs>
```

4.4.2006

XML Vorlesung ETHZ SS 2006

9

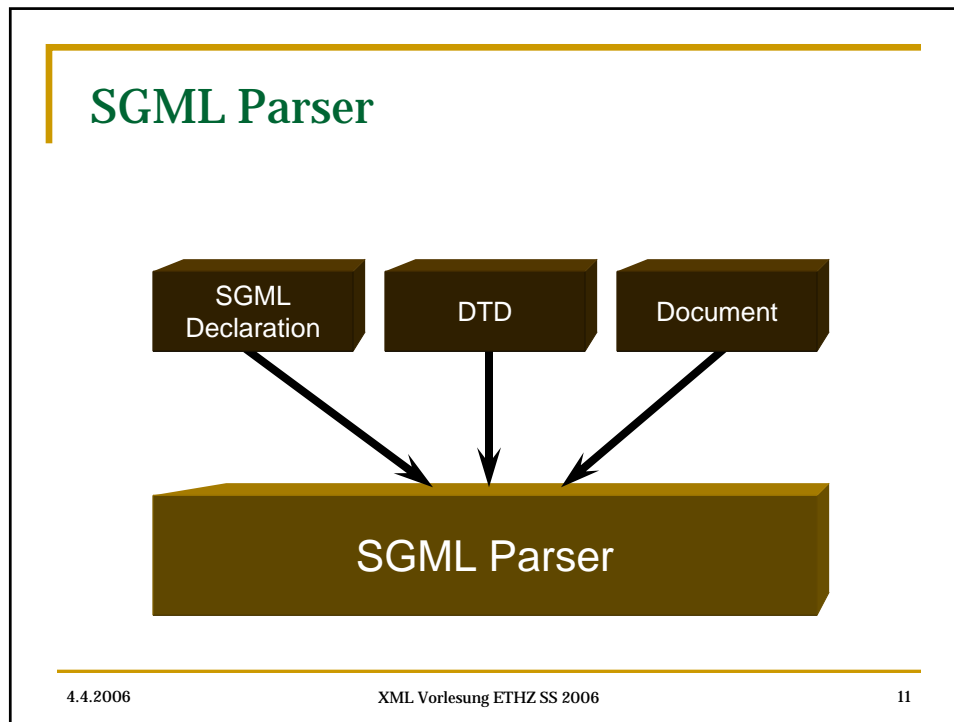
Aufbau von SGML

- die *SGML Declaration*
 - konkrete Syntax (Zeichen mit spezieller Bedeutung)
 - SGML Features (wie *Markup Minimization*)
- die *Document Type Definition (DTD)*
 - definiert erlaubte Elemente
 - definiert Kombination der erlaubten Elemente
 - definiert Attribute der erlaubten Elemente
 - definiert *Entities*, z.B. für Sonderzeichen
- das *Document* selber
 - Struktur des Inhaltes gemäss der DTD
 - Inhalt der Struktur (Text)

4.4.2006

XML Vorlesung ETHZ SS 2006

10



SGML Declaration

- bezieht sich auf mehrere Dokumente
- Festlegung der konkreten Syntax
 - SGML selber verwendet abstrakte Syntax
 - Zeichen mit Sonderbedeutung
 - HTML: < > </ <! <? = " & ;
- Festlegung von Zeichensätzen
- Festlegung von Kapazitäten
 - Länge von Namen, Schachtelungstiefen, ...
- SGML Features (von HTML bzw. XML verwendet)
 - *tag omission* (Weglassen von Tags)
 - *short tags* (Abkürzen von Tags)

4.4.2006 XML Vorlesung ETHZ SS 2006 12

SGML Document Type Definition

- Festlegung einer Grammatik
 - bestimmt die *Wörter* einer Sprache
 - bestimmt die *Regeln zur Satzbildung*
- Definition der Elemente
 - Elementnamen (frei wählbar)
 - Attributnamen, -typen und -werte
- Definition zur Kombination der Elemente
 - Vorkommen der Elemente in einem Dokument
 - *Model Groups*: `<! ELEMENT UL (LI)+ >`
 - nur HTML: *Exceptions (Inclusions/Exclusions)*

4.4.2006

XML Vorlesung ETHZ SS 2006

13

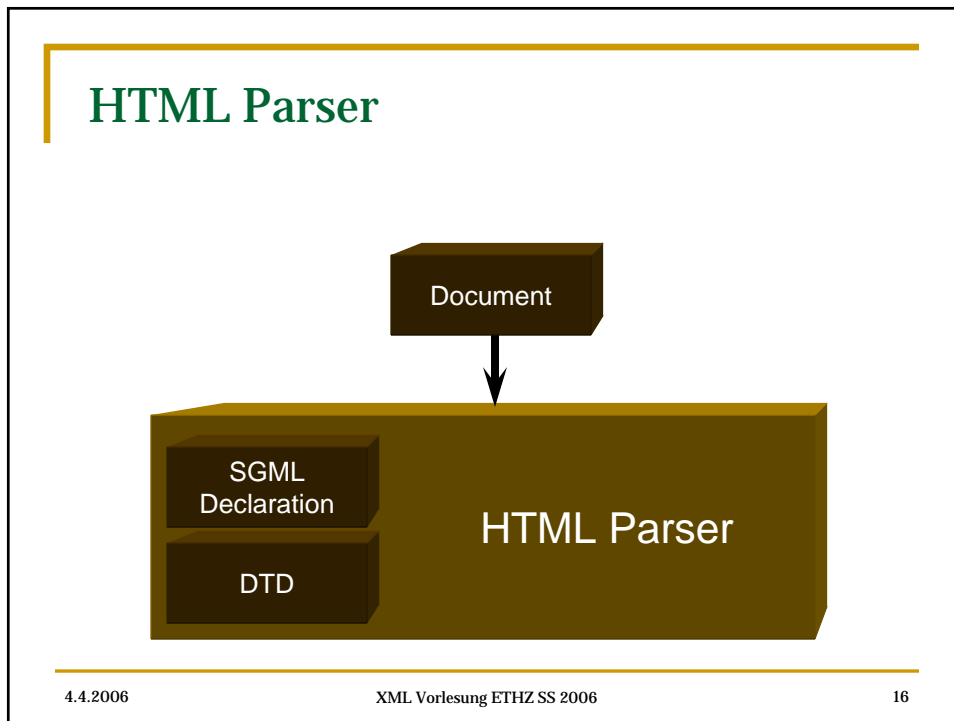
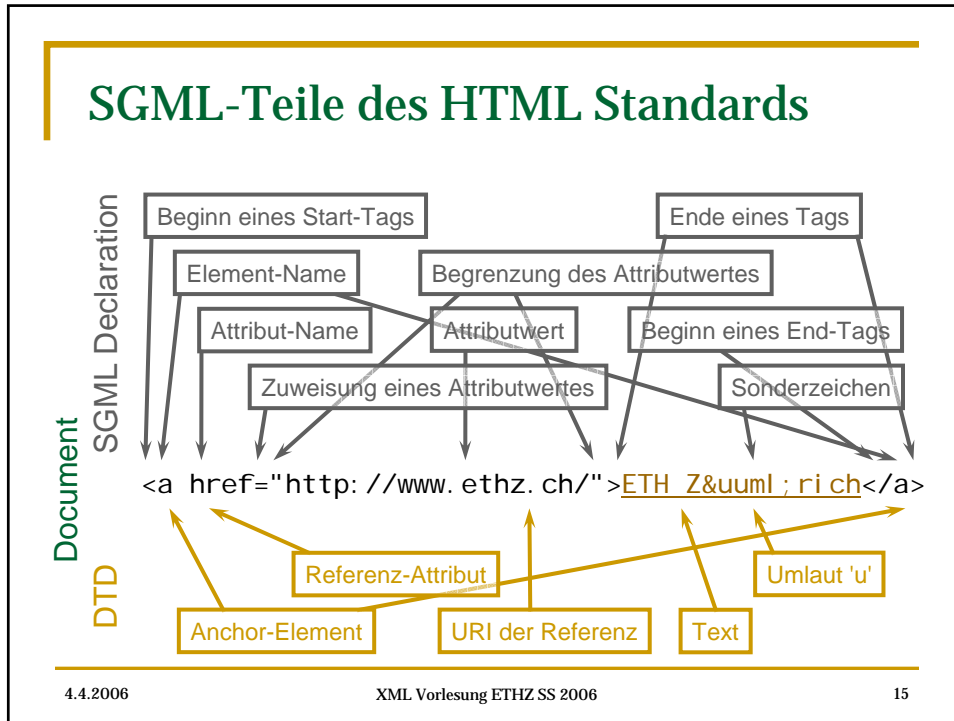
Document

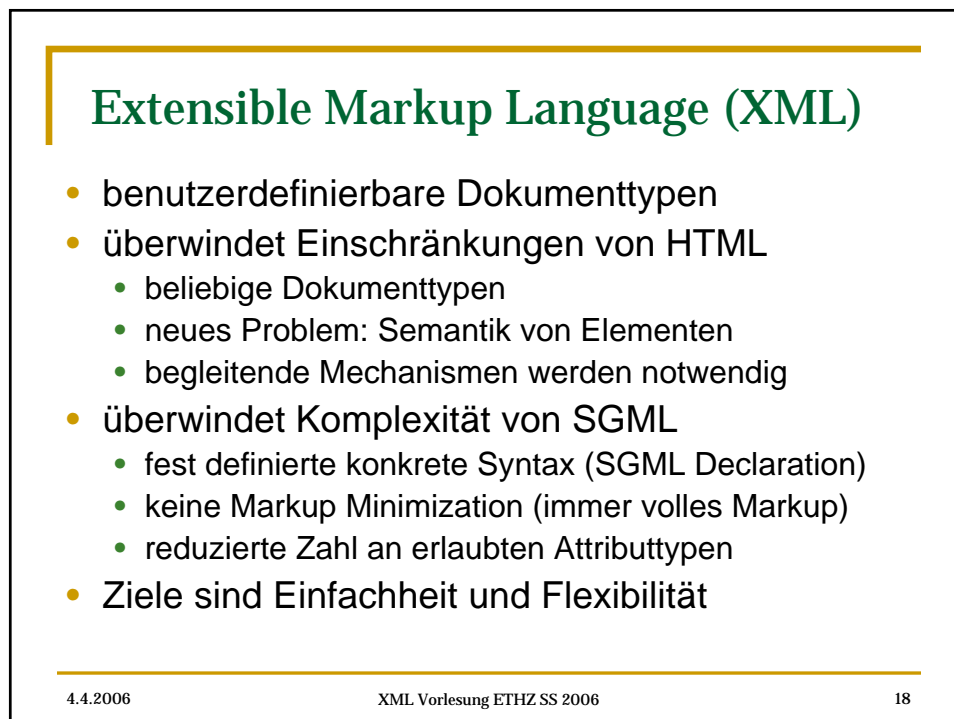
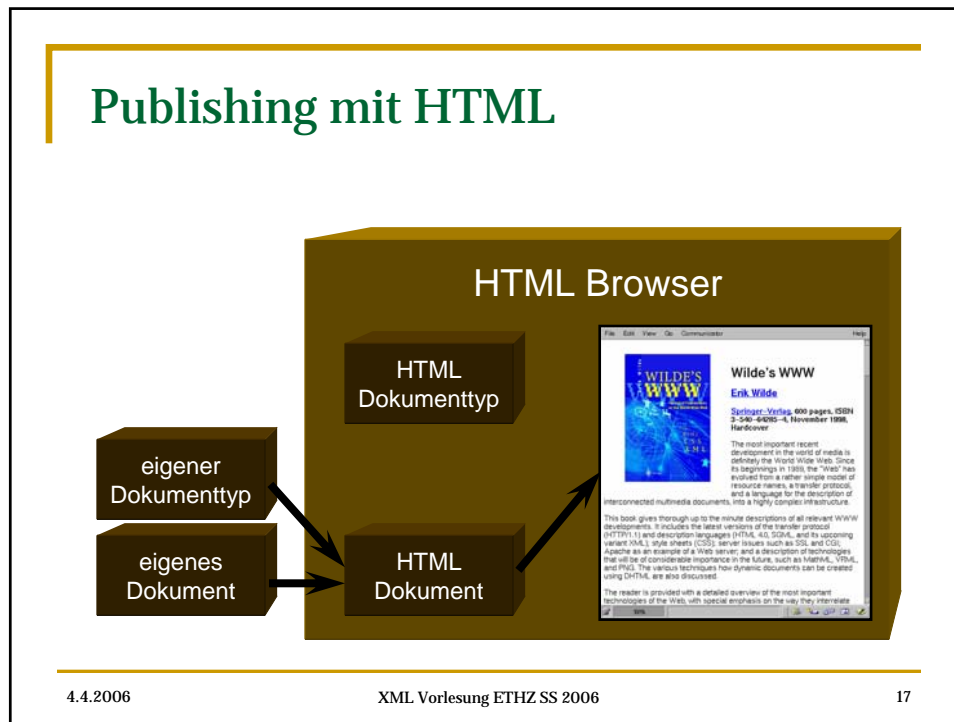
- Instanz eines bestimmten Dokumententyps
 - Kennzeichnung zu Beginn des Dokuments
 - kann nur mit Hilfe der DTD interpretiert werden
- Einhaltung der Regeln der SGML Declaration
- Einhaltung der Regeln der DTD
- stellt einen *document tree* dar
- Erstellung eines SGML Dokumentes
 - als Textdatei (ursprüngliches Modell)
 - mit SGML-Tools (zunehmend verbreitetes Modell)

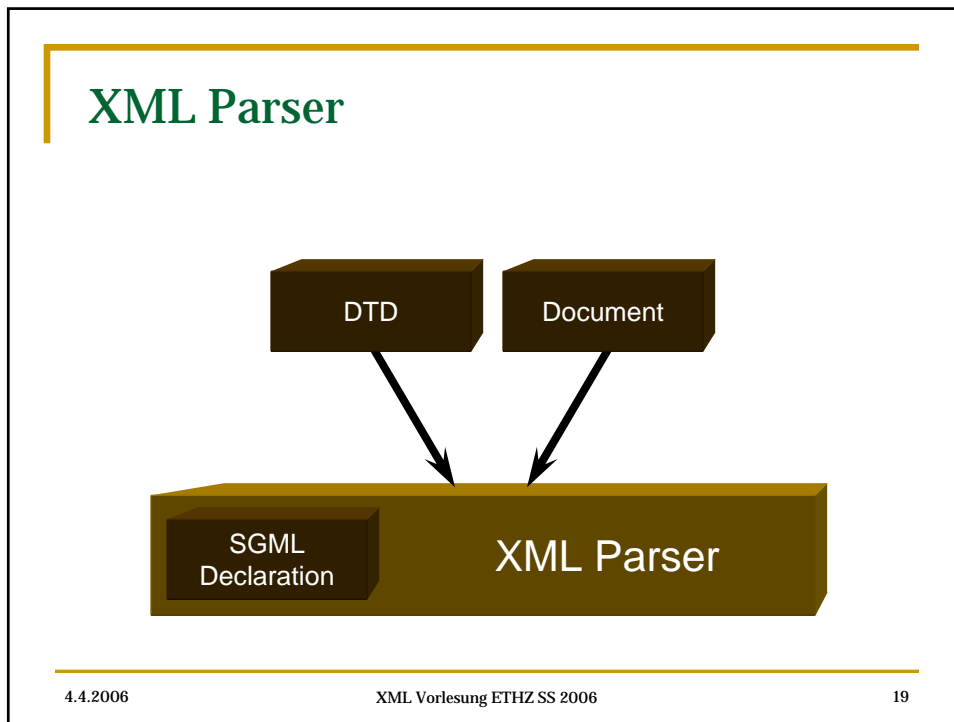
4.4.2006

XML Vorlesung ETHZ SS 2006

14



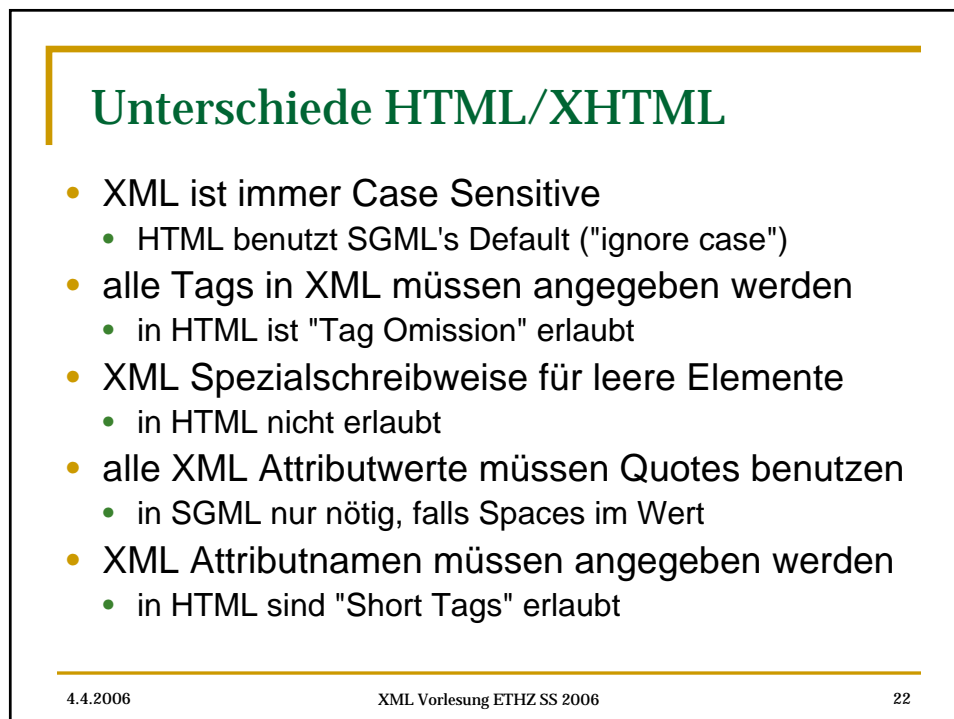
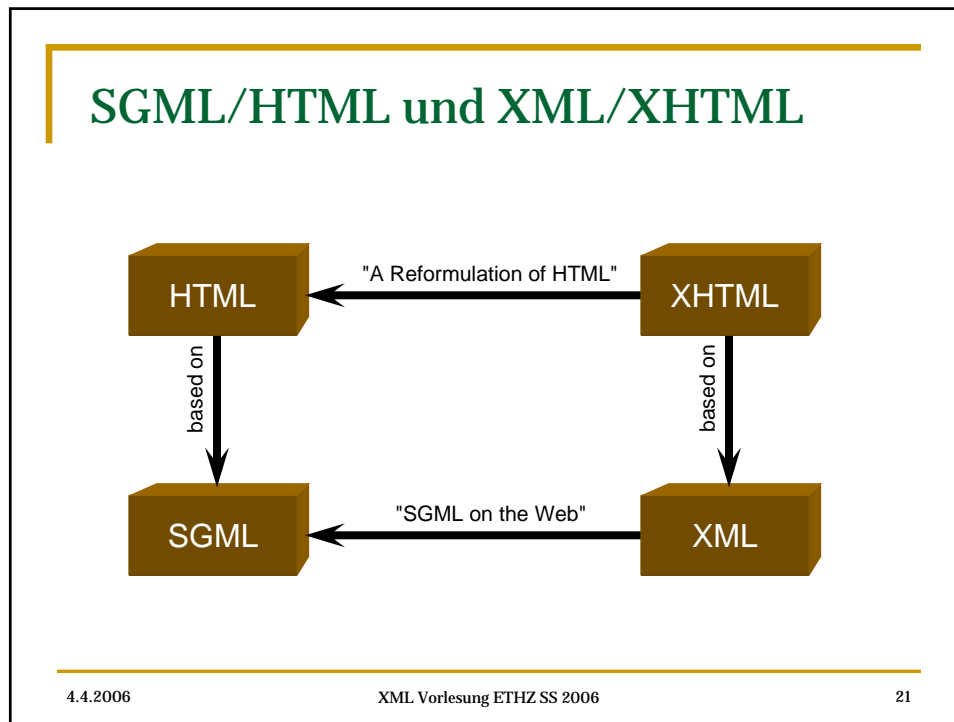




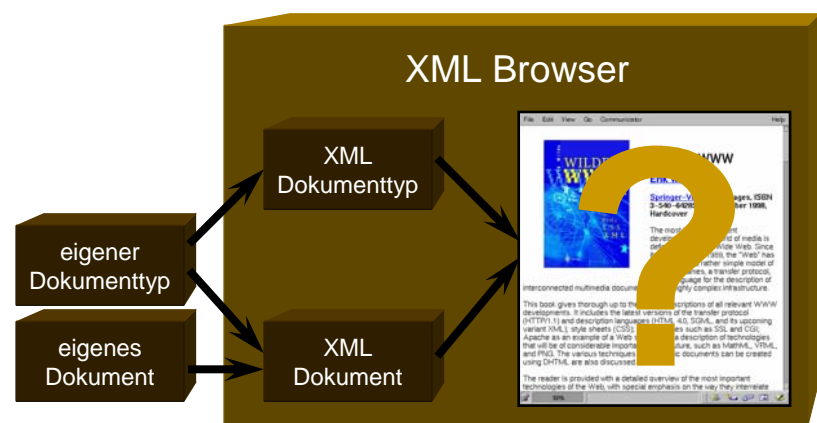
Vergleich SGML/XML/HTML

	SGML	XML	HTML
SGML Declaration	frei	fix	fix
DTD	frei	frei	fix
Document	frei	frei	frei

4.4.2006 XML Vorlesung ETHZ SS 2006 20



Publishing mit XML

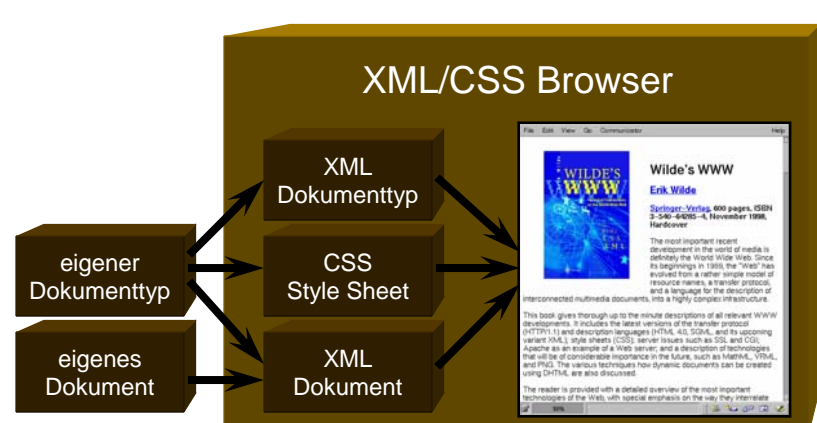


The diagram illustrates the workflow of an XML browser. On the left, two boxes labeled 'eigener Dokumenttyp' and 'eigenes Dokument' have arrows pointing to two central boxes: 'XML Dokumenttyp' and 'XML Dokument'. Arrows from these two central boxes point to a screenshot of a browser window titled 'XML Browser'. The browser window shows a document with a large yellow question mark overlaid on it, indicating a rendering issue or a missing component.

XML Browser

4.4.2006 XML Vorlesung ETHZ SS 2006 23

Publishing mit XML/CSS



The diagram illustrates the workflow of an XML/CSS browser. On the left, two boxes labeled 'eigener Dokumenttyp' and 'eigenes Dokument' have arrows pointing to three central boxes: 'XML Dokumenttyp', 'CSS Style Sheet', and 'XML Dokument'. Arrows from these three central boxes point to a screenshot of a browser window titled 'XML/CSS Browser'. The browser window shows a document with a clean, styled layout, indicating successful rendering.

XML/CSS Browser

4.4.2006 XML Vorlesung ETHZ SS 2006 24

XML/XSL Publishing (Client-Side)

The diagram illustrates the client-side publishing process. On the left, three boxes represent the input: 'eigener Dokumenttyp' (own document type), 'eigenes Dokument' (own document), and 'XML Dokument' (XML document). Arrows from these boxes point to a central column of three boxes: 'XML Dokumenttyp' (XML document type), 'XSL Style Sheet', and 'XML Dokument'. From this central column, arrows point to a window titled 'XML/XSL Browser'. The browser window displays a page titled 'Wilde's WWW' with a book cover image and text, representing the rendered output.

XML/XSL Browser

4.4.2006 XML Vorlesung ETHZ SS 2006 25

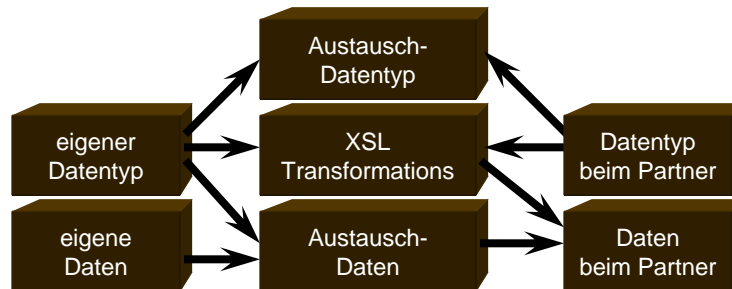
XML/XSL Publishing (Server-Side)

The diagram illustrates the server-side publishing process. On the left, three boxes represent the input: 'eigener Dokumenttyp' (own document type), 'eigenes Dokument' (own document), and 'XML Dokument'. Arrows from these boxes point to a central column of three boxes: 'XML Dokumenttyp' (XML document type), 'XSL Style Sheet', and 'XML Dokument'. From this central column, arrows point to a window titled 'HTML Browser'. The browser window displays a page titled 'Wilde's WWW' with a book cover image and text, representing the rendered output.

HTML Browser

4.4.2006 XML Vorlesung ETHZ SS 2006 26

B2B und XML/XSLT (Server-Side)



4.4.2006

XML Vorlesung ETHZ SS 2006

27

XML im B2B-Bereich

- mit Abstand grösster Erfolg von XML
 - XML als politisch erfolgreiche Lösung
 - keine Bindung an Hardware oder Betriebssystem
 - keine Bindung an Hersteller oder Konsortien
 - XML kam zur richtigen Zeit
 - Anfang 1998: Internet-Boom in voller Fahrt
 - Investitionswille praktisch unbeschränkt
- Anwendungen auf dem Web noch am Anfang
 - bisher nur möglich in Form von XHTML
 - fehlende Bausteine und fehlende Software

4.4.2006

XML Vorlesung ETHZ SS 2006

28

XML als Revolution des Web?

- Bedarf nach benutzerdefinierbaren Strukturen
- XML als logische Weiterentwicklung des Web
 - HTML wurde als zu limitiert erkannt
 - SGML wurde als zu komplex erkannt
 - XML als Kompromisslösung
- XML als Kompromiss aus altem und neuem
 - SGML funktioniert seit langem
 - das Web (mit HTML/HTTP) funktioniert auch
 - EDIFACT funktioniert ebenfalls seit langem
 - XML als "SGML on the Web"

4.4.2006

XML Vorlesung ETHZ SS 2006

29

XML und verwandte Standards

- XML selber ist ein einfacher Standard
 - Syntax für baumstrukturierte Dokumente
 - Syntax für eine einfache Schema-Sprache
- XML wird von vielen Standards begleitet
 - Hauptziel der Vorlesung: Orientierung bieten
 - Standards kennenlernen und einschätzen können
- W3C entwickelt viele XML-Standards
 - z.T. von Konsortiumsmitgliedern eingebracht
- viele Entwicklungen in anderen Organisationen

4.4.2006

XML Vorlesung ETHZ SS 2006

30

Zusammenfassung

- Grundlagenvorlesung zum Thema XML
 - Basis für Anwendungen in vielen Bereichen
 - praktische Erfahrung mit XML-Technologien
- XML als Web-Standard
 - entwickelt als "SGML on the Web"
 - Anwendungen in verschiedensten Bereichen
- XML lebt vom Umfeld
 - orientieren und einschätzen lernen
- XML wird es noch eine ganze Weile geben...