XML-Standards für das E-Business

© Klaus Schild, 2003

Organisatorisches

© Klaus Schild, 2003

Unterlagen

- Webseite der Lehrveranstaltung:
  http://www.inf.fu-berlin.de/inst/ag-nbi/lehre/0304/V_EBXML/
- hier finden sich:
  - Termine
  - Folien
  - Übungsblätter
  - Hinweise auf Literatur

© Klaus Schild, 2003

Literatur

- Hunter et al., Beginning XML (2nd Edition), Wrox Press, 2001: S. 29-147.
- http://www.w3.org/TR/xmlschema-0/ oder
  S. 217-288 aus Hunter et al. (2001).
- http://www.w3.org/TR/soap12-part0/
- http://www.w3.org/TR/wsdl
- http://java.sun.com/webservices/docs/1.2/tutorial/doc/: Kap.1, Unterkapitel The "SAX API" und "The DOM API".
- Buch kann zum kopieren im Sekretariat in der Fabekstr. 15 ausgeliehen werden.
- Bei der Rücksprache setze ich voraus, dass diese Literatur auch gelesen wurde.

© Klaus Schild, 2003

Formalitäten

- Für BSc- und MSc-Studierende ist eine *verbindliche* Anmeldung zur Veranstaltung notwendig.
- Ohne diese Anmeldung dürfen *keine* Leistungen erbracht werden.
- verbindliche Anmeldung mit Unterschrift in der nächsten Woche

© Klaus Schild, 2003

Übungsbetrieb

- Übungen alle *zwei Wochen* Do 10:15-11:45 im SR 053
- dort:
  - weitere Fragen zur Vorlesung
  - Fragen zum aktuellen Übungsblatt
  - Rückgabe des letzten Übungsblattes
- ca. 6 Übungsblätter
  - enthalten praktische Aufgaben
  - werden bewertet
- Es sollten Gruppen von 3 Studierenden gebildet werden.
- 20 Lizenzen des XML-Editors XML Spy stehen in den PC-Pools zur Verfügung.
- Es können aber auch andere XML-Editoren benutzt werden.

© Klaus Schild, 2003

Scheinkriterien

- alle Übungsaufgaben erfolgreich bestanden
- die individuelle Rücksprache erfolgreich bestanden
- Gesamtnote: Mittel aus Übungsnote und Note der Rücksprache

© Klaus Schild, 2003

Termine

- Übungen *alle zwei Wochen* Do 10:15-11:45 im SR 053
- erste Übung am 6.11.
- verpflichtende Anmeldung für BSc- und MSc-Studierende am 5.11.
- individuelle Rücksprachen bis Ende Februar 2004

© Klaus Schild, 2003

Kommunikation

- Veranstalter: schild@inf.fu-berlin.de
- Sprechstunde ohne Anmeldung: immer abwechselnd zur Übung Do 10:15-11:45 in der Fabekstr. 15.
- Forum: nbi_v_ebxml@lists.spline.inf.fu-berlin.de

© Klaus Schild, 2003

Inhalt der Veranstaltung

© Klaus Schild, 2003

Vorlesungsinhalt

Die wichtigsten XML-Technologien

- Im Vordergrund stehen dabei maschinenlesbare Dokumente (Maschine-Maschine-Kommunikation)
- *nicht* behandelt werden: XML-Technologien zur Präsentation von Dokumenten: XHTML oder WML (Mensch-Maschine-Kommuniaktion)

Web Services

- plattformunabhängige XML-Protokolle für verteilte Systeme

E-Business-Anforderungen

- Zuverlässigkeit, Sicherheit, Transaktionen

© Klaus Schild, 2003

XML-Technologien

Was ist XML?

© Klaus Schild, 2003

Binärdateien

- enthalten reichhaltige Informationen, wie Daten zu interpretieren sind (Metadaten)
- Z.B. kann ein Word-Dokument Informationen enthalten, dass ein bestimmtes Textfragment fett darzustellen ist.
- Nachteil: werden nur von einem bestimmten Anwendungsprogramm verstanden

© Klaus Schild, 2003

Textdateien

- enthalten wie Binärdateien nur Bits
- Bits sind so angeordnet, dass sie Zahlen repräsentieren, die wiederum bestimmte Zeichen darstellen.
- werden von jedem Anwendungsprogramm verstanden, die entsprechende Zeichenrepräsentation kennt (z.B. Texteditoren und Browser)
- können deshalb als anwendungsunabhängig bezeichnet werden
- Nachteil: Alle Zeichenketten werden gleich behandelt, es gibt keine Informationen, wie Daten zu interpretieren sind (Metadaten).

© Klaus Schild, 2003

Auszeichnungssprachen

- kombinieren Vorteile von Binärdateien mit denjenigen von Textdateien
- anwendungsunabhängige Dateiformate, die reichhaltige Metadaten enthalten können
- Eine *Auszeichnungssprache* (engl. *markup language*) ist eine textbasierte Sprache, die sog. *Tags* („Markierung“) benutzt, um Daten mit zusätzlicher Information (Metadaten) zu versehen:

  <tag-name>ausgezeichneter Text</tag-name>

- Beispiel: *Hypertext Markup Language* (HTML)

© Klaus Schild, 2003

Verallgemeinerte Auszeichnungssprachen

- In HTML gibt es eine *vorgegebene* Auswahl von Tags, keine anderen dürfen verwendet werden.
- Verallgemeinerte Auszeichnungssprachen (engl. *generalized markup languages*) erlauben es, beliebige Tags zu verwenden.
- Vorteil: beliebige Metainformationen darstellbar
- Nachteil: Bedeutung der Metainformationen (Tags) ist offen und muss durch die Anwendung festgelegt werden
- Beispiele: SGML und XML

© Klaus Schild, 2003

SGML

- *SGML = Standard Generalized Markup Language*
- 1969 von Charles Goldfarb und zwei seiner Kollegen bei IBM entwickelt.
- seit 1986 ein internationaler Standard für das Dokumentenmanagement
- hat *keine* vorgegebenen Tags, auch keine für das Layout von Dokumenten

© Klaus Schild, 2003

Beispiel

- SGML erlaubt das Strukturieren von Dokumenten:

```
<book>
  <title>Beginning XML</title>
  <edition>2nd</edition>
  <authors>
    <author>David Hunter</author>
    <author>Curt Cagle</author>
    <author>Chris Dix</author>
  </authors>
  <date>2001</date>
  <publisher>Wrox Press</publisher>
  <abstract>...</abstract>
  <chapters>...</chapters>
</book>
```

© Klaus Schild, 2003

SGML

- SGML gibt *keine* konkreten Tags vor.
- SGML stellt aber eine Möglichkeit zu Verfügung, spezielle Auszeichnungssprachen mit konkreten Tags zu definieren.
- Solche Untermengen von SGML werden auch *Anwendungen* genannt.
- Diese beschreiben, an welchen Stellen im Dokument, welche Tags benutzt werden dürfen. Sie definieren spezielle *Dokument-Typen*.
- Bekannteste Anwendung von SGML ist HTML.
- Weil SGML spezielle Auszeichnungssprachen definieren kann, wird SGML auch als *Meta-Sprache* bezeichnet.

➔ http://userpage.fu-berlin.de/~corff/SGML/

© Klaus Schild, 2003

Vor- und Nachteile von SGML

+ als Auszeichnungssprache kombiniert SGML die Vorteile von Binärdateien mit denjenigen von Textdateien
+ als verallgemeinerte Auszeichnungssprache ist SGML beliebig erweiterbar
+ als Meta-Sprache erlaubt SGML die Deklaration von konkreten Auszeichnungssprachen wie HTML
- SGML ist sehr komplex (Spezifikation über 500 Seiten lang)
- SGML-Parser sind schwierig zu implementieren

© Klaus Schild, 2003

Extensible Markup Language (XML)

- Für die Präsentation von Dokumenten hat sich HTML bewährt.
- Das Internet wird in Zukunft immer mehr auch für die Maschine-Maschine-Kommunikation genutzt werden.
- Für den Austausch von Dokumenten und Daten zwischen Computern ist HTML ungeeignet, hierfür wird eine verallgemeinerte Auszeichnungssprache (wie SGML) gebraucht.
- Für das Web ist SGML viel zu komplex.

XML ist eine konsequente Vereinfachung von SGML, die für Web-Anwendungen hinreichend allgemein ist.

© Klaus Schild, 2003

Eine kurze Geschichte von XML

1969 Charles Goldfarb entwickelt bei IBM die Generalized Markup Language (GML).
1980 ANSI veröffentlicht ersten Entwurf von SGML.
1986 ISO verabschiedet SGML.
1989 Bernes-Lee schlägt SGML-basiertes Hypertext-System vor.
1990 Beners-Lee entwickelt HTML, HTTP und URL. World Wide Web nimmt seinen Betrieb mit zwei Maschinen am CERN auf.
1994 Gründung des World Wide Web Consortiums (W3C)
1995 HTML 2.0
1998 XML 1.0
2000 XHTML 1.0 (Reformulierung von HTML in XML)
XML 1.0 (2nd Edition)

© Klaus Schild, 2003

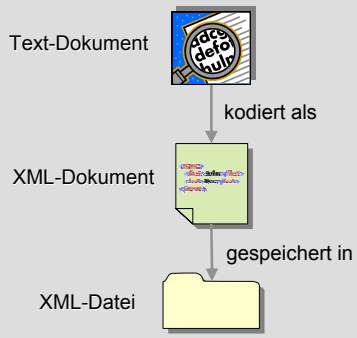
Extensible Markup Language (XML)

- Von SGML hat XML das grundlegende Konzept der verallgemeinerten Auszeichnungssprache übernommen:

- XML-Dokumente werden von den meisten modernen Browsern angezeigt.

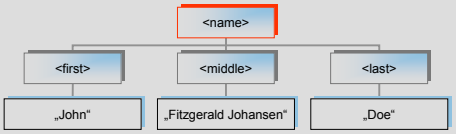
© Klaus Schild, 2003

XML-Dokumente



© Klaus Schild, 2003

Baumstruktur von XML-Dokumenten



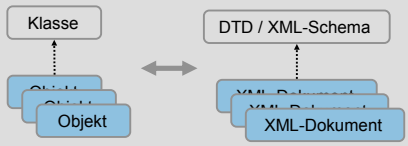
- Jedes XML-Dokument hat *genau ein* Wurzelelement.
- Dieses Wurzelelement entspricht dem Dokument-Typ.
- Kind-Elemente sind immer *geordnet*.

Dokument-Typen

- Beschreiben den *prinzipiellen Aufbau* von Dokumenten eines bestimmten Typs, z.B.:
  1) Das Wurzelelement ist <name>.
  2) <first>, <middle> und <last> sind Kind-Elemente von <name>.
  3) <first> und <last> sind obligatorisch, <middle> optional.
  4) <first>, <middle> und <last> erscheinen in dieser Reihenfolge.

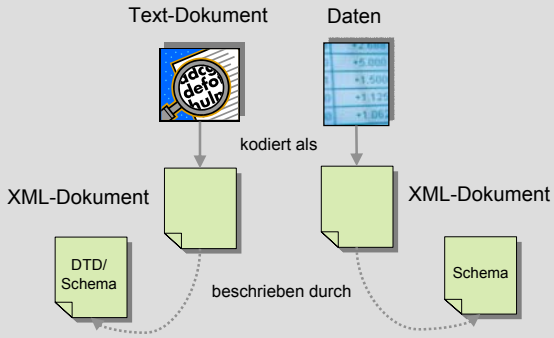
Dokument-Typen

- Dokument-Typen können entweder mit DTDs (*Document Typ Definitions*) oder *XML-Schemata* spezifiziert werden.



- DTDs für Spezifikation von Text-Dokumenten ausreichend, XML-Schemata zur Spezifikation von Daten besser geeignet.

Text-Dokumente vs. Daten



DTDs vs. XML-Schemata

```
<location>
  <latitude>32.904237</latitude>
  <longitude>73.620290</longitude>
  <uncertainty units="meters">2</uncertainty>
</location>
```

XML-Schema

DTD

- Eine Ortsangabe besteht aus dem Breitengrad, dem Längengrad und einem Maß für die Unsicherheit der beiden Angaben.
- Ein Breitengrad ist eine Dezimalzahl zwischen -90 und +90.
- Ein Längengrad ist eine Dezimalzahl zwischen -180 und +180.
- Das Maß für die Unsicherheit ist eine nicht-negative Zahl.
- Die Unsicherheit wird entweder in Meter oder in Fuß angeben.

Erweiterbarkeit

- Das X in XML steht für erweiterbar (engl. *extensible*).
- Um zu verstehen, was Erweiterbarkeit konkret bedeutet, vergleichen wir XML mit HTML:

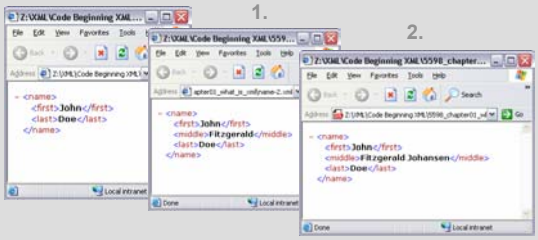
HTML:

- vorgegebene Auswahl an Sprach-Elementen
- Neues Sprach-Element kann nur eingeführt werden, wenn sich das W3C auf eine neue HTML-Version einigt!

XML:

- beliebige Elemente können benutzt werden
- Nur die Anwender des entsprechenden Elementes müssen sich auf eine gemeinsame Interpretation einigen.

Beispiel für Erweiterbarkeit



1. Neue Elemente können jederzeit eingeführt werden (verallgemeinerte Auszeichnungssprache).
2. Ein Element kann *unabhängig* von anderen Elementen erweitert werden (strikte Element-Struktur).

Die XML-Familie: Der Kern

- XML 1.0: Syntax wohlgeformter XML-Dokumente und Definition einfacher Dokument-Typen (DTD)
- Name Spaces: gleichzeitige Verwendung unterschiedlicher Vokabularien in einem XML-Dokument, z.B. Titel einer Person und eines Buches
- XML-Schema: Definition komplexen Datentypen wie sie von Programmiersprachen bekannt sind

Die XML-Familie

- Extensible Stylesheet Language (XSLT): Transformation von XML-Dokumenten in beliebige Text-Formate, einschl. HTML und andere XML-Formate
- XPath: Zugriff auf beliebige Teile eines XML-Dokumentes, wie z.B. die Nachnamen aller Autoren
- XLink: Verknüpfung von XML-Dokumenten, ähnlich wie Hypertext-Links in HTML.
- Document Object Model (DOM): Zugriff, Modifikation und Erstellung von XML-Parse-Bäumen

Exkurs: Das W3C

W3C WORLD WIDE WEB consortium

„lead the World Wide Web to its full potential by developing common protocols that promote its evolution and its interoperability.“

- 1994 als Projekt am MIT gegründet.
- keine Normierungsorganisation im Sinn klassischer staatlicher Standardisierung
- kann Einhaltung von Normen *nicht* auf rechtlichem Wege einklagen
- definiert deshalb lediglich Empfehlungen (*recommendations*)
- Alle W3C-Recommendations sind lizenzfrei.

Standardisierungsprozess des W3C

Recommendation (REC)	offizieller W3C-Standard
Proposed Recommendation (PR)	Stellt offiziellen Konsens der betreffenden AG dar. Wird dem Advisory Committee übergeben.
Candidate Recommendation (CR)	Von entsprechender Community bereits eingehend begutachtet. Aufruf zur Begutachtung auch an alle anderen.
Last Call WD	Letztes WD. Aufforderung an alle betroffenen AGs zur Begutachtung.
Working Draft (WD)	Aktueller Diskussionsstand einer AG. Nicht unbedingt Konsens der AG.

Anwendungen von XML

zwei unterschiedliche Kategorien von Anwendungen

XML als Meta-Sprache

- XML wird als Meta-Sprache zur Definition von konkreten Auszeichnungsprachen für spezielle Anwendungen eingesetzt.

Lösung konkreter Probleme

- XML-Technologien werden zur Lösung konkreter Probleme eingesetzt.

XML als Meta-Sprache

- XHTML: Reformulierung von HTML in XML
- WML: Wireless Markup Language
  Präsentation von Inhalten auf mobilen Endgeräten
- DocBook: strukturierte Darstellung von Bücher/Artikel
- MathML: Mathematical Markup Language
  Standard für mathematische Ausdrücke
- SVG: Scalable Vector Graphics
  Standard für Vektorgraphiken
- SMIL: Synchronized Multimedia Integration Language
  Standard für Multi-Media-Anwendungen
- VoiceXML: Voice Extensible Markup Language
  Standard für interaktive Sprachanwendungen

unterschiedliche Anwendungen, aber einheitliche Syntax!

© Klaus Schild, 2003

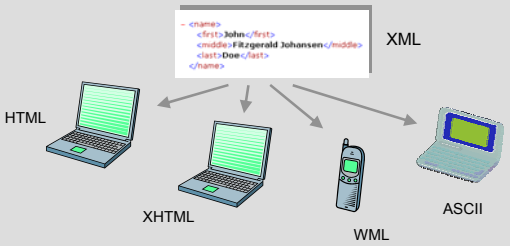
XML als Lösung konkreter Probleme

- Trennung von Inhalt und Präsentation
- Semantic Web
- Web Services

© Klaus Schild, 2003

Trennung von Inhalt und Präsentation

- Ein und derselbe Inhalt kann auf verschiedenen Endgeräten unterschiedlich dargestellt werden.



➔ auch Multi-Delivery genannt

© Klaus Schild, 2003

Semantic Web

- Statt einfacher Suche nach Schlagworten, Suche in strukturierten Dokumenten.
- Beispiel: Wenn ein Artikel mit DocBook strukturiert wird, kann gezielt nach Stichwörter in der Kurzfassung gesucht werden.

© Klaus Schild, 2003

Web Services

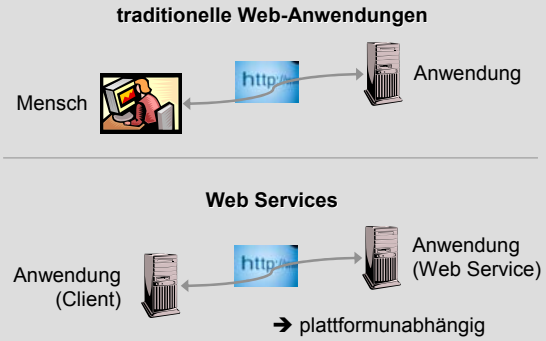
- entfernte Prozeduraufrufe über gängige Internet-Protokolle wie HTTP
- Integration der Microsoft und der Unix-Welt
- Zum ersten mal in der Geschichte konnte sich die IT-Industrie auf einen gemeinsamen Standard einigen.

© Klaus Schild, 2003

Was sind Web Services?

© Klaus Schild, 2003

Web Services



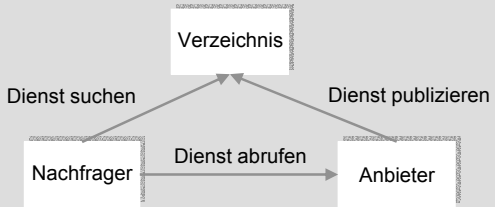
Web Services

- integrieren *existierende* Systeme
- Fassade für ähnliche Systeme
- zwei verschiedene Erscheinungsformen: entfernte Prozeduraufrufe (synchron) oder Messaging (asynchron)
- Nachrichten-Protokoll: XML-Standard SOAP
- SOAP-Nachrichten meist über HTTP(S)
- Schnittstellen können mit XML-Standard WSDL (Web Services Description Language) beschrieben
- einfacher Zugriff auf ansonsten verdeckte Systeme

Beispiel Google

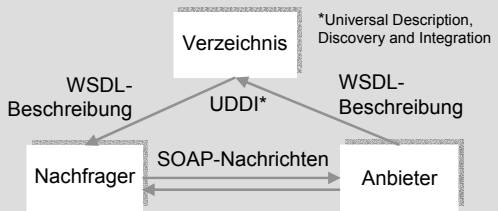
- Google-Dienst wird auch als Web Service angeboten.
- Google-Dienst kann damit in beliebige Anwendungen integriert werden, um z.B.:
  - zu einem bestimmten Themen in periodischen Abständen nach neuen Informationen zu suchen
  - automatisch neue Trends im WWW zu identifizieren
  - die Rechtschreibkorrektur von Google zu nutzen

Service-Oriented Architecture



- Sowohl die Suche als auch der Abruf des Dienstes kann automatisch oder manuell geschehen.
- Nur einfache Dienste wie Aktienkurs(WKN,Boersenplatz) können vollautomatisch gesucht und aufgerufen werden.

Web Service Standards

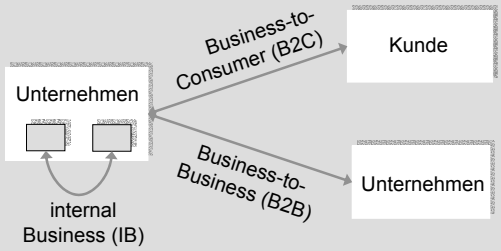


- Nutzung von SOAP zur Übertragung von Web-Service-Nachrichten wird als zwingend angesehen.
- Schnittstellenbeschreibung werden eher als optional angesehen.
- Standards für Register umstritten

E-Business-Anforderungen

E-Business

- E-Business: geschäftliche Beziehungen, die über ein elektronisches Medium abgewickelt werden

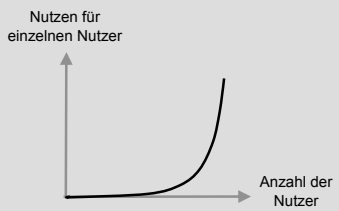


Typische E-Business-Anforderungen

- Nachricht nur vom Empfänger lesbar?
- Absender der Nachricht eindeutig identifizierbar?
- Nachricht unversehrt (unverändert)?
- Nachricht tatsächlich genau einmal übermittelt?
- Antwort auch noch nach längerer Zeitverzögerung möglich?
- Passt die Nachricht in den vorgesehenen Geschäftsprozess (Interaktionsmuster)?

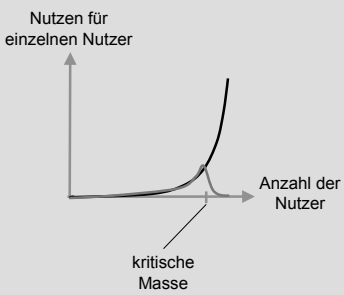
- keine dieser Anforderungen wird von SOAP oder WSDL direkt unterstützt
- Entsprechende Erweiterungen von SOAP und WSDL werden zur Zeit entwickelt.

Exkurs: Netzwerkeffekte



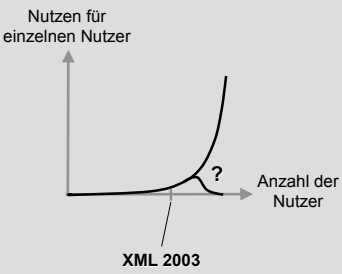
- Beispiele: Telefon, E-Mail, HTML, XML
- Gibt es nur einen Nutzer, ist der Nutzen für diesen Nutzer gleich Null.
- Je mehr Nutzer, desto höher der Nutzen für die einzelnen Nutzer.

Kritische Masse



- Unterhalb einer kritischen Masse ist der Nutzen sehr gering.
- Oberhalb einer kritischen Masse steigt der Nutzen rapide an.
- Bei konkurrierenden Netzwerken *nicht* vorhersagbar, welches Netzwerk sich durchsetzen wird.

Netzwerkeffekte XML



- XML hat eindeutig noch *nicht* die kritische Masse von Nutzern erreicht.
- Aber: Zum ersten mal in der Geschichte haben sich *alle* großen IT-Unternehmen auf einen Standard geeinigt: XML, SOAP und WSDL