



Semantic Web

Vorlesung

Dr. Harald Sack

Hasso-Plattner-Institut für Softwaresystemtechnik

Universität Potsdam

Wintersemester 2008/09



<http://sw0809.blogspot.com/>

Blog zur Vorlesung: <http://sw0809.blogspot.com/>

Semantic Web - Vorlesungsinhalt

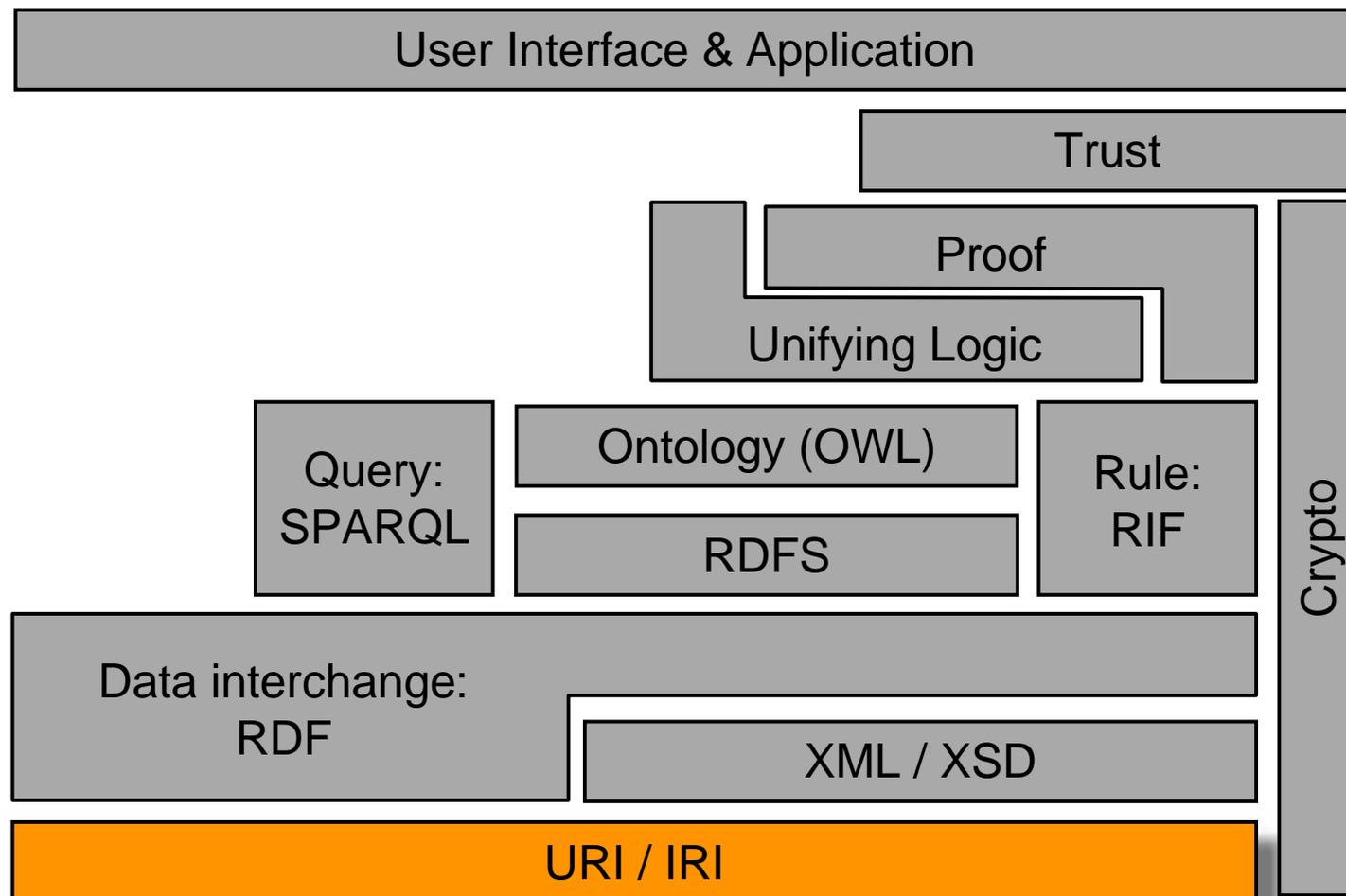
2

1. Einführung
2. **Die Sprachen des Semantic Web**
3. Wissensrepräsentation
4. Web of Trust
5. Ontology Engineering
6. Semantic Web Anwendungen

2. Die Sprachen des Semantic Web

3

Semantic Web Architecture



- 1 **06.11.2008 – Vorlesung Nr. 2**
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13

2. Die Sprachen des Semantic Web

2.1 Uniform Resource Identifier - URI

2.2 Universelle Vokabularien mit XML

2.3 Einfache semantische Strukturen mit RDF und RDFS

2.4 Abfragesprache SPARQL

2.5 Semantik im WWW - μ formats, RDFa, GRDDL

2.6 Warum RDF/RDFS noch nicht ausreicht

2. Die Sprachen des Semantic Web

2.1 Uniform Resource Identifier - URI

5

Uniform Resource Identifier - URI

- Ein **Uniform Resource Identifier** (URI) definiert ein einfaches und erweiterbares Schema zur weltweit eindeutigen Identifikation von Ressourcen (RFC 3986).
- Ressource kann jedes Objekt sein, das (im Kontext der jeweiligen Anwendung) eine klare Identität besitzt ,
 - also z.B. (z.B. Webseiten, Bücher, Orte, Personen, Beziehungen zwischen diesen Dingen, abstrakte Konzepte usw.)
- Das URI Konzept ist in verschiedenen Anwendungsbereichen bereits etabliert,
 - wie z.B. Web (URL, PRN, pURL), Bücher (ISBN), Digital Object Identifier (DOI)

Uniform Resource Identifier - URI

- URI umfasst sowohl
 - **Adressangaben (Locator)**
 - Uniform Resource Locator (URL, RFC 1738)
 - gibt an, wo eine Web-Ressource zu finden ist
 - kann sich im Lebenszyklus der Web-Ressource ändern
 - **Identitätsbezeichner (Namen)**
 - Uniform Resource Name (URN, RFC 2141)
 - persistenter Bezeichner für eine Web-Ressource
 - bleibt für die gesamte Lebensdauer konstant

2. Die Sprachen des Semantic Web

2.1 Uniform Resource Identifier - URI

7

URI - Generische Syntax

- US-ASCII Encoding
- Percent Encoding für reservierte Zeichen, oder Zeichen, die nicht im US-ASCII Encoding existieren
 - `pct-encoded = "%" HEXDIG HEXDIG`
- Reservierte Zeichen mit funktionalen Aufgaben
 - `reserved = gen-delims / sub-delims`
 - `gen-delims = ":" / "/" / "?" / "#" / "[" / "]" / "@"`
 - `sub-delims = "!" / "$" / "&" / "'" / "(" / ")" / "*" / "+" / "," / ";" / "="`
- Erlaubte Zeichen
 - `unreserved = ALPHA / DIGIT / "-" / "." / "_" / "~"`
- Erweiterung auf Universal Character Code (Unicode/ISO 10646)
 - **Internationalized Resource Identifier (IRI, RFC 3987)**
 - Bsp.: `http://كق#فغ=عظ;طض=صش?سز/وذخ/حج.ثت.با`

2. Die Sprachen des Semantic Web

2.1 Uniform Resource Identifier - URI

URI - Generische Syntax

URI = schema"://" [userinfo"@"] host[:port] [path] ["?" query] ["#" fragment]

- schema: z.B. http, ftp, mailto,...
- userinfo: z.B. user:password
- host: z.B. Domain-Name, IPv4/IPv6-Adressen
- port: z.B. 80 für Standard http-Port
- path: z.B. Pfadangabe im WWW-Server Filesystem
- query: z.B. Parameter, die an Anwendung weitergegeben werden
- fragment: z.B. Angabe eines bestimmten Dokumententeilbereichs

2. Die Sprachen des Semantic Web

2.1 Uniform Resource Identifier - URI

9

URIs im Semantic Web

- Was tun, wenn eine URI für eine Ressource noch nicht existiert?
- Selbstdefinierte URIs
 - Zur Vermeidung von Überschneidungen, Nutzung der eigenen Website
 - Ermöglicht auch Ablage von zugehöriger Dokumentation an gleicher Stelle
- Trennung von URI für Ressource (*Beschriebenes*) und deren Dokumentation (*Beschreibendes*) durch URI-Referenzen (durch "#" angehängte Fragmente) oder content negotiation

z.B.: http://hpi.ui-potsdam.de/harald_foaf.rdf#URI

- 1
- 06.11.2008 – Vorlesung Nr. 2**
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13

2. Die Sprachen des Semantic Web

2.1 Uniform Resource Identifier - URI

2.2 Universelle Vokabularien mit XML

2.2.0 Warum (X)HTML nicht ausreicht...

2.2.1 XML Sprachstandard

2.2.2 XML DTD

2.2.3 XML Schema (XSD)

2.2.4 XML Sprachfamilie

2.2.5 Warum XML noch nicht ausreicht...

2.3 Einfache semantische Strukturen mit RDF und RDFS

2.4 Abfragesprache SPARQL

2.5 Semantik im WWW - μ formats, RDFa, GRDDL

2.6 Warum RDF/RDFS noch nicht ausreicht

2. Die Sprachen des Semantic Web

2.2 XML / 2.2.0 Warum HTML nicht ausreicht...

11

Semantik und HTML

- HTML diene ursprünglich zur Beschreibung der Struktur eines Hypermedia-Dokuments
- **Browser-Wars** – Browser-Hersteller (Netscape/MS) definierten eigenständig proprietäre Erweiterungen der HTML-Syntax zur Formatierung von Strukturelementen
 - z.B. Tabellen mit absoluten Maßangaben
 - z.B. Schrifttyp, Schriftauszeichnungen, Schriftgröße, Farbe, etc...
- Problem: **Formatierung vs. Semantik**

```
<td width="261">  
<font color="#00ff00" face="Arial, Helvetica, sans-serif,"  
size="+1">
```

fetter Text

2. Die Sprachen des Semantic Web

2.2 XML / 2.2.0 Warum HTML nicht ausreicht...

Semantik und HTML

- **Meta-Tags** können zusätzliche Semantik in das HTML-Dokument einbringen

```
<meta name="description"  
      content="Homepage of Dr. Harald Sack, ... ">  
<meta name="keywords" content="homepage teaching ... ">  
<meta name="Autor"    content="Harald Sack">
```

- **Problem:**
 - Erweiterungen **nicht standardisiert**
 - leicht manipulierbar → Missbrauch

2. Die Sprachen des Semantic Web

2.2 XML / 2.2.0 Warum HTML nicht ausreicht...

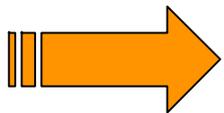
13

Semantik und HTML

- Semantik der Meta-Tags beruht ausschließlich auf der privaten Vereinbarung zwischen Informationsanbieter und Informationskonsumenten

```
<meta name="PLZ"           content="07743">  
<meta name="Ort"          content="Jena">  
<meta name="Strasse"      content="Ernst-Abbe-Platz">  
<meta name="„Hausnummer" content="2">
```

- **Problem:**
 - Syntax und Semantik gilt **nur für spezielle Vereinbarung**



Fazit:

HTML war zwar für das immense Wachstumstempo des WWW mitverantwortlich, behindert aber dessen Weiterentwicklung

2. Die Sprachen des Semantic Web

2.2 XML / 2.2.0 Warum HTML nicht ausreicht...

µformats -- Microformate

- XHTML Markup, mit dem es möglich ist, (in einem begrenzten Umfang) Semantik in einem HTML-Dokument auszudrücken
- Anwendungen können leichter Daten aus HTML-Dokumenten extrahieren
- Verwendung in folgenden XHTML-Tag **Attributen**:
 - **class**
 - **rel**
 - **rev**
- Vordefinierte Standard-Microformate:
 - hCard - Personendaten (→ vCard, RFC2426)
 - hCalendar – Events
 - rel-Tag – social tagging
 - XFN – XHTML Friends Network
 - ...

2. Die Sprachen des Semantic Web

2.2 XML / 2.2.0 Warum HTML nicht ausreicht...

15

µFormats -- Microformate

- Beispiel:

```
<span class="vcard" >  
  <span class="fn">Joe Blow</span>  
  <span class="org">The Example Company</span>  
  <span class="tel">604-555-1234</span>  
  <a class="url" href="http://example.com/" >  
    http://example.com/</a>  
</span>
```

- 1 **06.11.2008 – Vorlesung Nr. 2**
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13

2. Die Sprachen des Semantic Web

2.1 Uniform Resource Identifier - URI

2.2 Universelle Vokabularien mit XML

2.2.0 Warum (X)HTML nicht ausreicht...

2.2.1 XML Sprachstandard

2.2.2 XML DTD

2.2.3 XML Schema (XSD)

2.2.4 XML Sprachfamilie

2.2.5 Warum XML noch nicht ausreicht...

2.3 Einfache semantische Strukturen mit RDF und RDFS

2.4 Abfragesprache SPARQL

2.5 Semantik im WWW - μ formats, RDFa, GRDDL

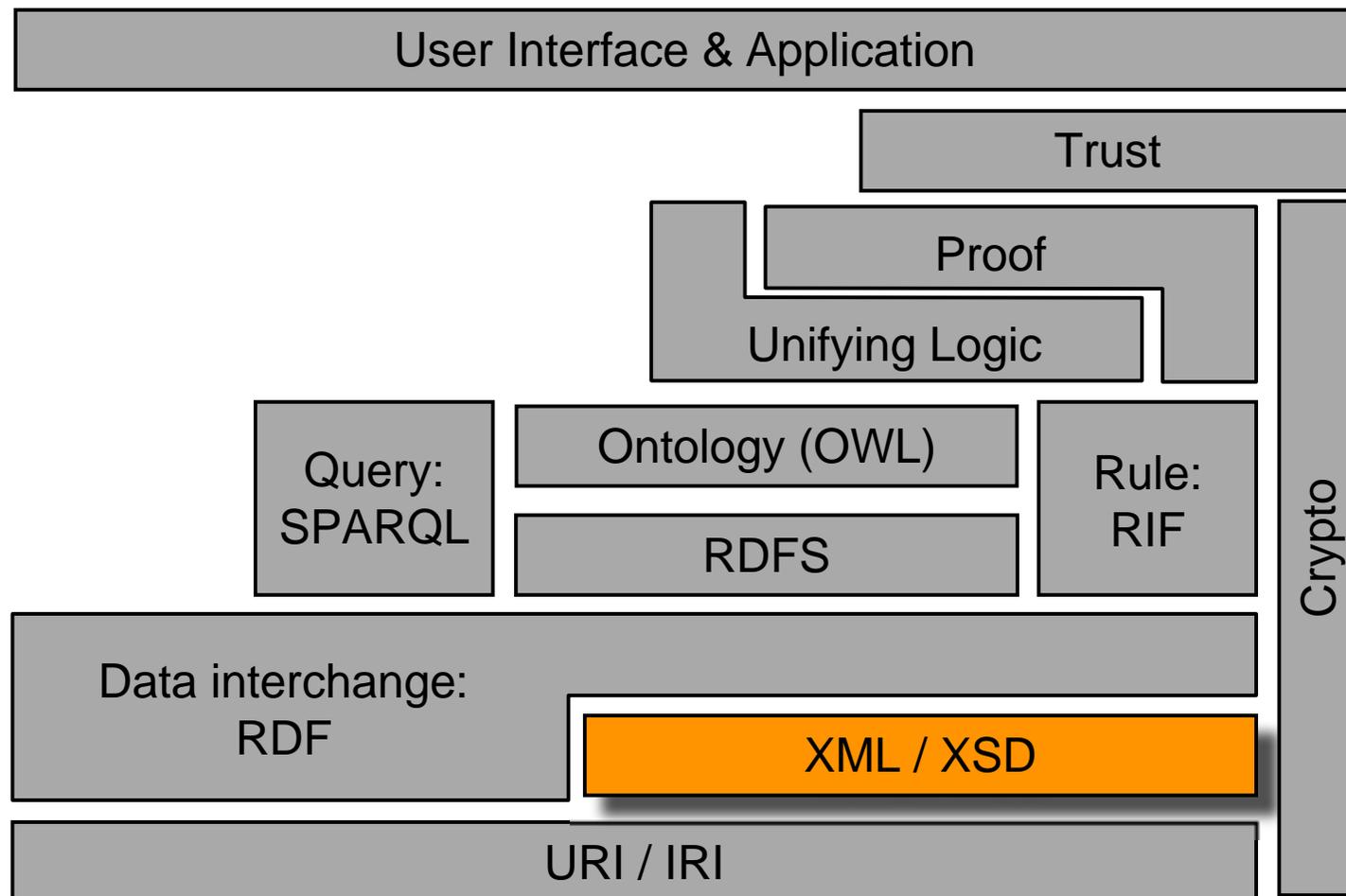
2.6 Warum RDF/RDFS noch nicht ausreicht

2. Die Sprachen des Semantic Web

2.2 XML / 2.2.1 XML Sprachstandard

17

Semantic Web Architecture



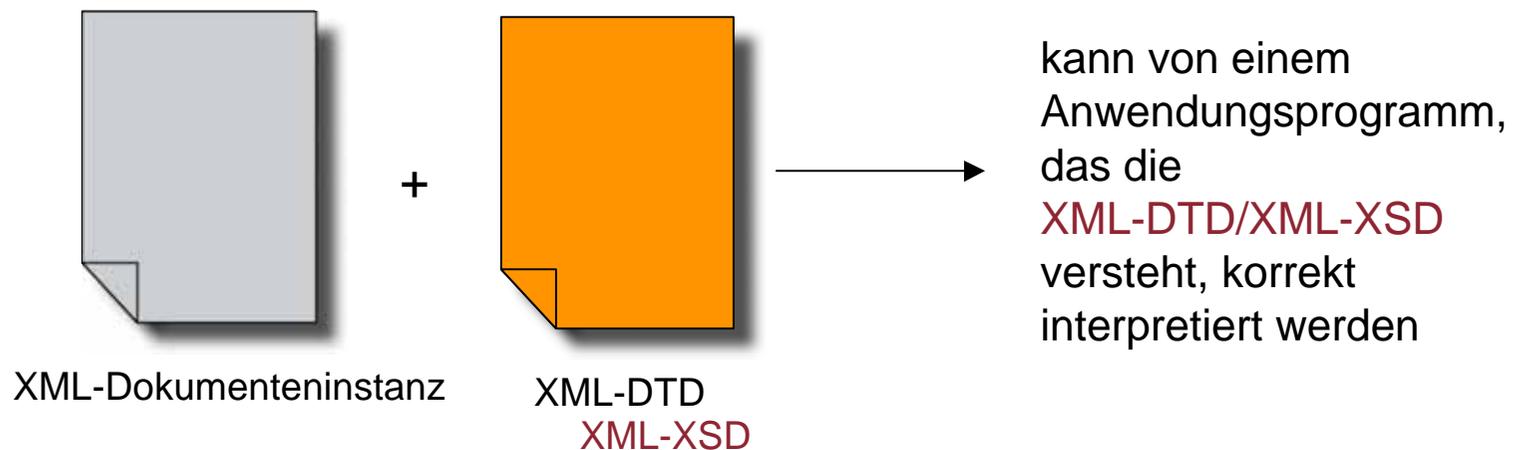
2. Die Sprachen des Semantic Web

2.2 XML / 2.2.1 XML Sprachstandard

18

XML – Extensible Markup Language

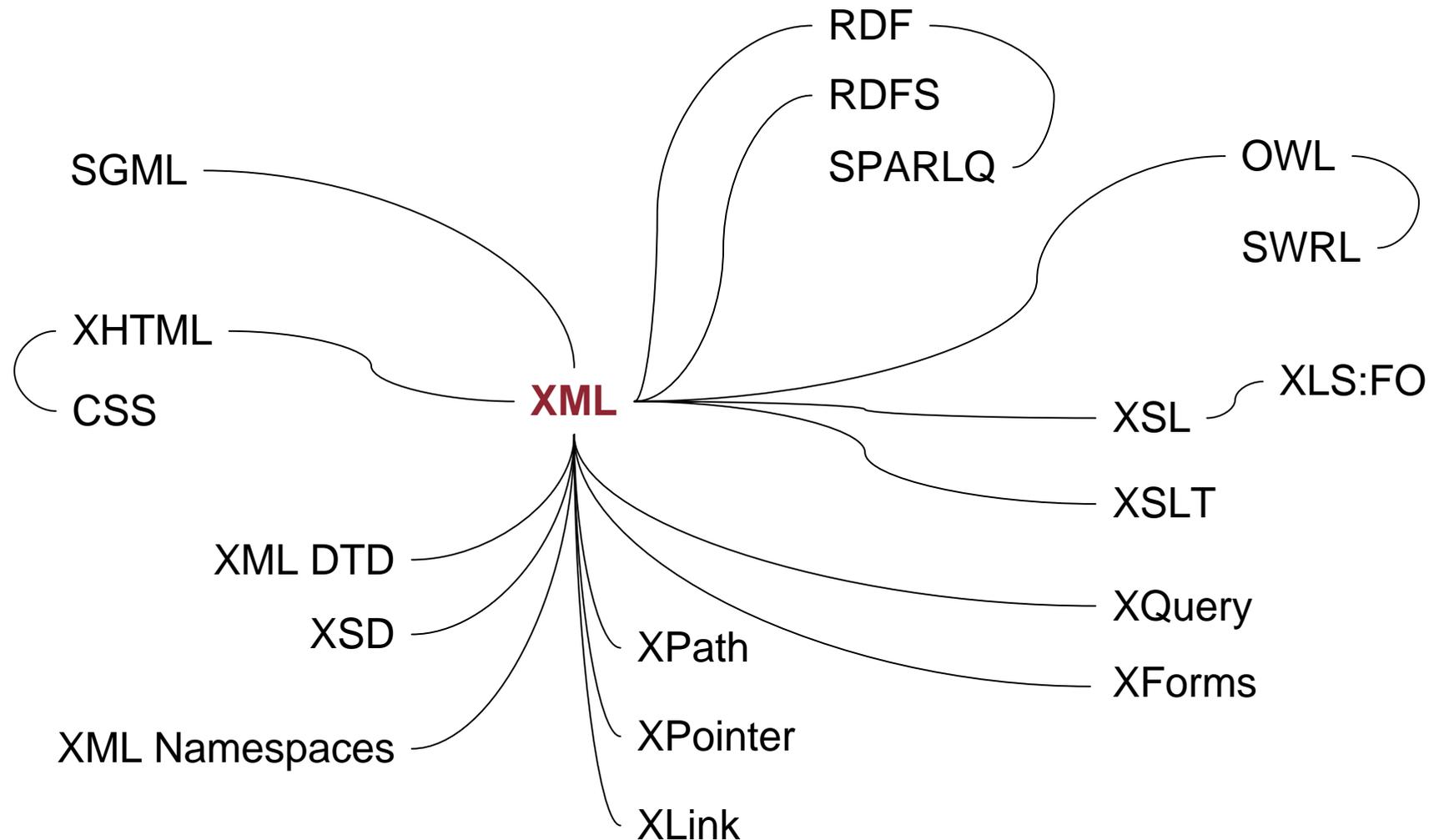
- Was XML alles kann
 - XML gestattet die Definition beliebiger neuer Tags (Metasprache zur Definition neuer Markupssprachen)
 - die Definition neuer Tags erfolgt in einem speziellen Dokument, der Document Type Definition (DTD) / XML Schema Definition



2. Die Sprachen des Semantic Web

2.2 XML / 2.2.1 XML Sprachstandard

19

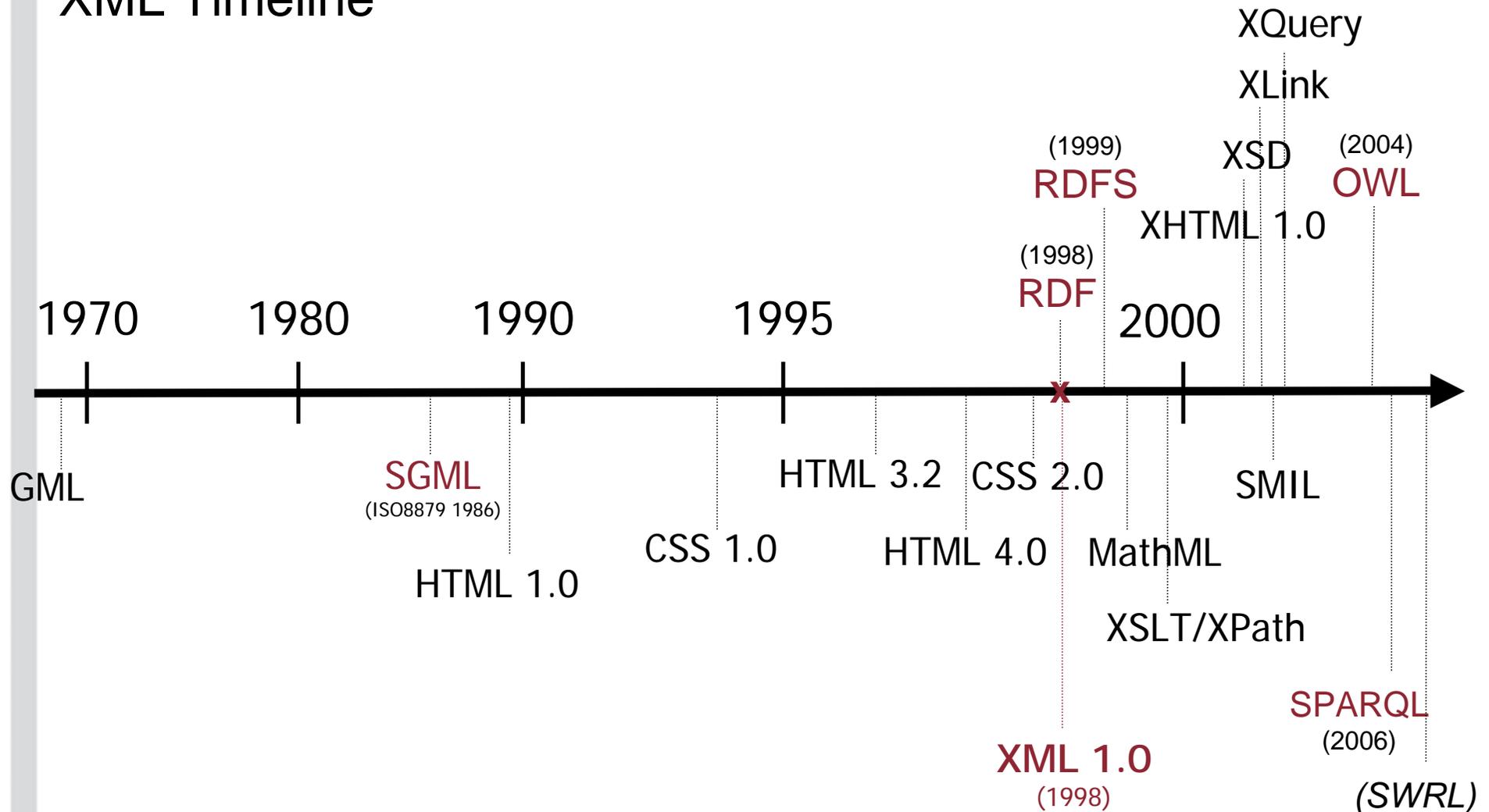


2. Die Sprachen des Semantic Web

2.2 XML / 2.2.1 XML Sprachstandard

20

XML Timeline



2. Die Sprachen des Semantic Web

2.2 XML / 2.2.1 XML Sprachstandard

21

XML – Extensible Markup Language

- Was XML alles kann

```
<h2>
Max Mustermann
</h2>

<p>
Sesamstr. 49a <br>
<b>93123 Bad Sulzdetfurth </b>
</p>
```

HTML

```
<Adresse>
  <Vorname> Max </Vorname>
  <Nachname> Mustermann </Nachname>
  <Straße> Sesamstr. </Straße>
  <Hausnummer> 49a </Hausnummer>
  <PLZ> 93123 </PLZ>
  <Ort> Bad Sulzdetfurth </Ort>
</Adresse>
```

XML



XML als semistrukturiertes Austauschdatenformat (Ein-/Ausgabe)
für beliebige Anwendungen

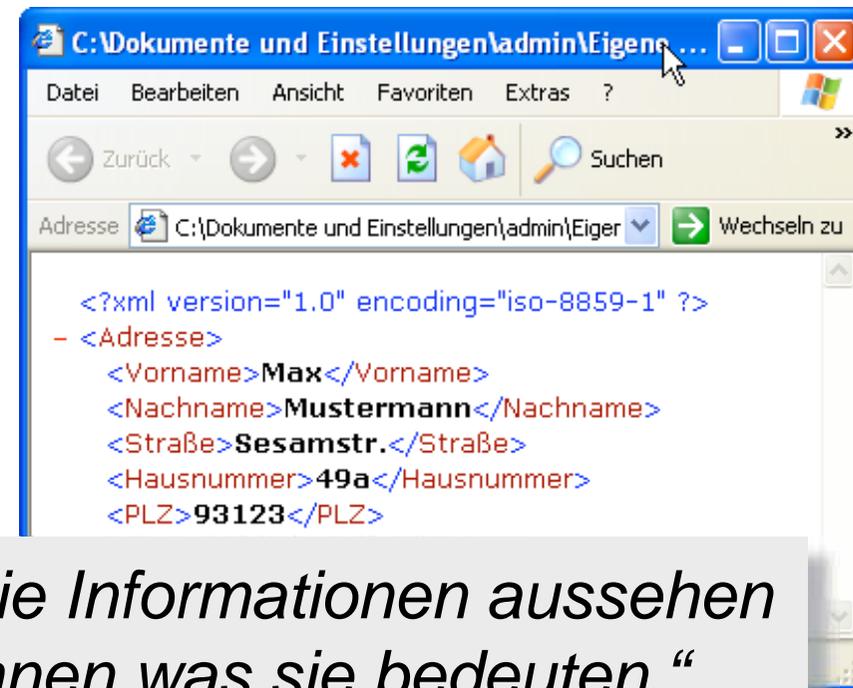
2. Die Sprachen des Semantic Web

2.2 XML / 2.2.1 XML Sprachstandard

22

XML – Extensible Markup Language

- Was XML alles kann



„HTML sagt Ihnen wie die Informationen aussehen sollen, aber XML sagt Ihnen was sie bedeuten.“

Quelle: Charles F. Goldfarb, „The XML Handbook“

2. Die Sprachen des Semantic Web

2.2 XML / 2.2.1 XML Sprachstandard

23

XML – Extensible Markup Language

SGML (Standard Generalized Markup Language)



HTML



XML



XHTML

- Instanz von SGML
- Vermischung von Inhalt und Layout
- Darstellung von Information
- Teilmenge von SGML
- Strikte Trennung von Inhalt und Layout
- Beschreibung von Information

XHTML ist Instanz von XML

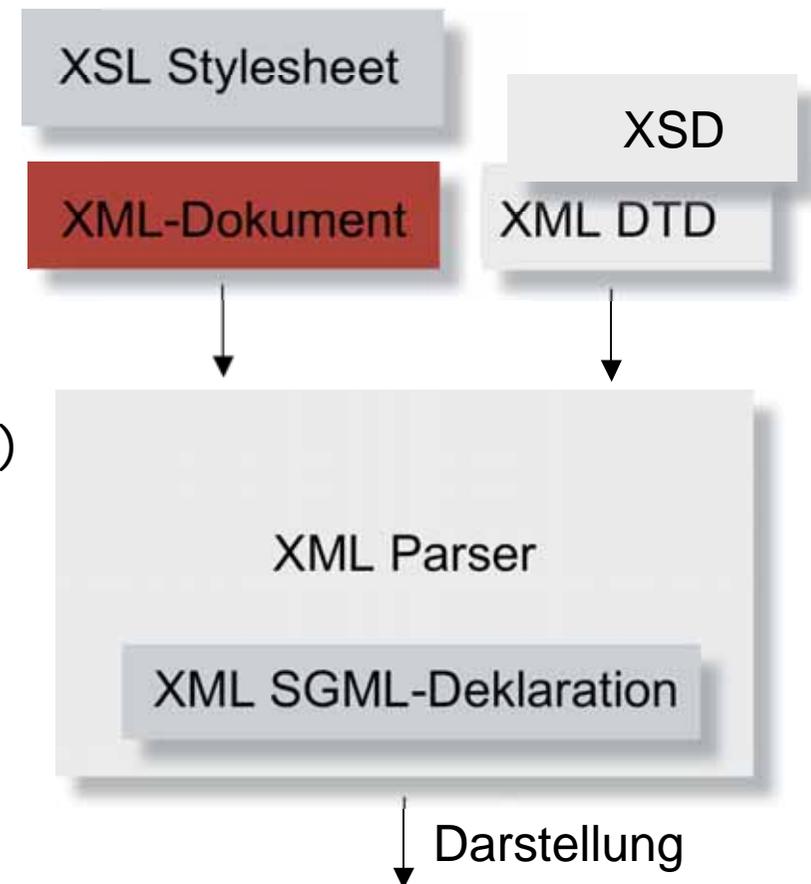
2. Die Sprachen des Semantic Web

2.2 XML / 2.2.1 XML Sprachstandard

24

XML – Extensible Markup Language

- HTML unterstützt stets nur einen vordefinierten Dokumententyp
- XML erlaubt die Definition eigener Dokumenten-Typen über die Definition einer zugehörigen **XML DTD** (Document Type Definition) bzw. XML Schema Definition (**XSD**)



2. Die Sprachen des Semantic Web

2.2 XML / 2.2.1 XML Sprachstandard

25

XML – Extensible Markup Language

```
<?xml version="1.0" encoding="iso-8859-1"?>
```

```
<!DOCTYPE address SYSTEM "Beispiel.dtd">
```

```
<adresse>
```

```
  <vorname> Max </vorname>
```

```
  <nachname> Mustermann </nachname>
```

```
  <straße> Sesamstr. </straße>
```

```
  <hausnummer> 49a </hausnummer>
```

```
  <plz prefix="D"> 93123 </plz>
```

```
  <ort> Bad Sulzdetfurth </ort>
```

```
</adresse>
```

XML Deklaration

XML DTD

XML Tags

XML Attribute

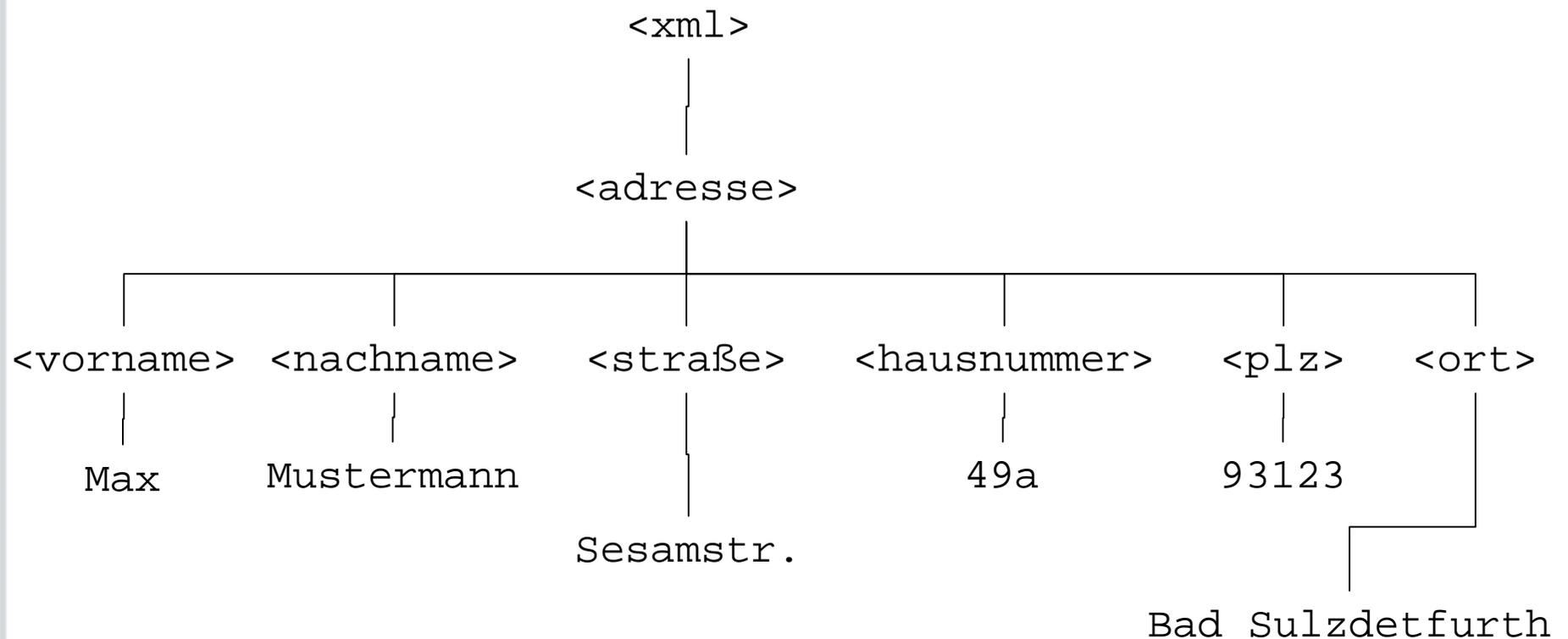
2. Die Sprachen des Semantic Web

2.2 XML / 2.2.1 XML Sprachstandard

26

XML – Extensible Markup Language

- Veranschaulichung von XML-Daten als gerichteter Graph



2. Die Sprachen des Semantic Web

2.2 XML / 2.2.1 XML Sprachstandard

27

XML – Deklaration

- XML-Prolog gibt Direktiven zur XML-Kodierung
 - version: obligatorische Versionsnummer
 - encoding: **UTF-8/16/32** ...
 - standalone: **yes/no** (wird DTD benötigt?)

```
<?xml version="1.0" encoding="UTF-8" standalone="yes">
```

2. Die Sprachen des Semantic Web

2.2 XML / 2.2.1 XML Sprachstandard

28

XML – Deklaration

- XML-Referenzen:
 - Angabe des verwendeten XML-DTDs
 - Referenz auf externe Teilmenge (globale URI)

```
<!DOCTYPE name SYSTEM/PUBLIC "URI">
```

- Referenz auf interne Teilmenge

```
<!DOCTYPE name [DTD-Definitionen]>
```

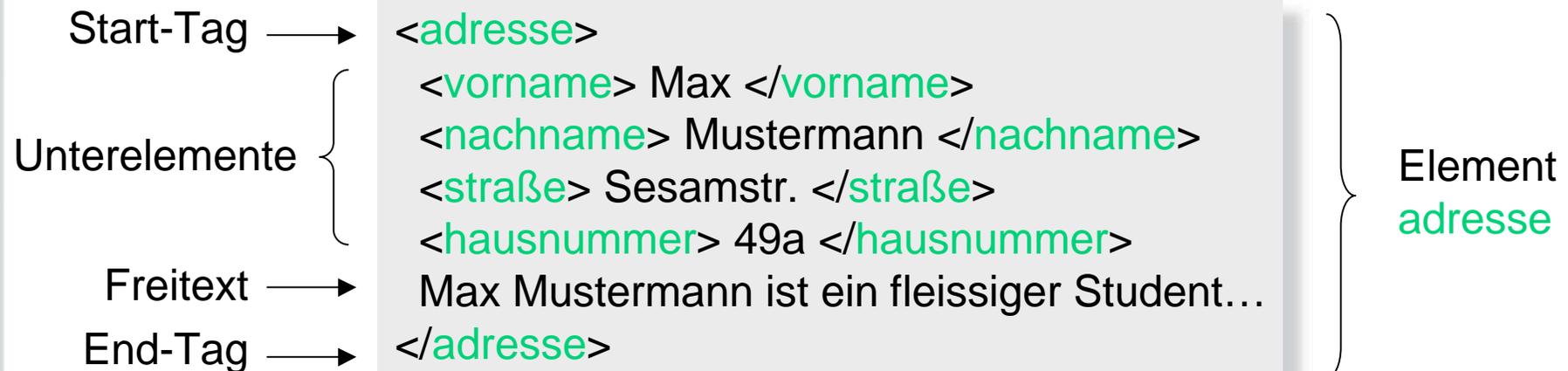
2. Die Sprachen des Semantic Web

2.2 XML / 2.2.1 XML Sprachstandard

29

XML – Element

- Beschreibung eines Objekts, das durch passende Markierungen (tags) wie `<vorname>` und `</vorname>` geklammert ist
- Inhalt eines Elements: Text und/oder weitere Elemente (Unterelemente)
- Elemente können beliebig geschachtelt sein
- Leere Elemente: `<plz></plz>` kurz: `<plz/>`



2. Die Sprachen des Semantic Web

2.2 XML / 2.2.1 XML Sprachstandard

30

XML – Attribute

- Name-Zeichenkettenwert-Paar
- Assoziiert mit einem Element
- Alternative Möglichkeit, Daten zu beschreiben

```
<adresse>  
  <vorname> Max </vorname>  
  <nachname> Mustermann </nachname>  
  <ort plz="01234"> Bad Sulzdetfurth </ort>  
  ...
```

┌──────────┐
Attribut **plz**

2. Die Sprachen des Semantic Web

2.2 XML / 2.2.1 XML Sprachstandard

31

Wohlgeformtes und gültiges XML

Max Mustermann
Sesamstr. 49a
93123 Bad Sulzdetfurth

```
<adresse>  
  <vorname> Max </vorname>  
  <nachname> Mustermann </nachname>  
  <straße> Sesamstr. </straße>  
  <hausnummer> 49a </hausnummer>  
  <ort plz="93123">Bad Sulzdetfurth</ort>  
</adresse>
```

Wohlgeformte XML-Dokumente:

Syntaktisch korrekt, d.h. alle angefangenen Tags werden wieder geschlossen und sind stets korrekt geschachtelt,
keine Mehrfachattribute, nur ein Wurzelement

2. Die Sprachen des Semantic Web

2.2 XML / 2.2.1 XML Sprachstandard

32

Wohlgeformtes und gültiges XML

```
<adresse>  
  <vorname> Max </vorname>  
  <nachname> Mustermann </nachname>  
  <straße> Sesamstr. </straße>  
  <hausnummer> 49a </hausnummer>  
  <ort plz="93123">Bad Sulzdetfurth</ort>  
</adresse>
```

```
<!ELEMENT adresse (vorname+,  
  nachname, straße, hausnummer,  
  ort)>
```

```
<!ELEMENT vorname CDATA>  
<!ELEMENT nachname CDATA>  
<!ELEMENT straße CDATA>
```

....

Document Type Definition (DTD)

gültige (valide) XML-Dokumente:

wohlgeformt und zu einem assoziierten Schema uneingeschränkt konform

2. Die Sprachen des Semantic Web

2.2 XML / 2.2.1 XML Sprachstandard

33

Schemata in XML

- **DTD – Document Type Definitions:**
 - Einfache Grammatik für ein XML-Dokument
 - Deklaration von Elementen, Attributen, u.a.
 - beschränkt die beliebige Verschachtelung von Elementen und Attributen
 - Teil des XML-Standards (“Erbe“ von SGML)
- **XML-Schema:**
 - Komplexere Datendefinitionssprache
 - standardisierte Basistypen, z.B. float, double, decimal, boolean
 - Typen und typisierte Objektreferenzen
 - Klassenhierarchien / Vererbung
 - Konsistenzbedingungen
 - Standard (W3C Recommendation) in Ergänzung zu XML

- 1
- 06.11.2008 – Vorlesung Nr. 2**
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13

2. Die Sprachen des Semantic Web

2.1 Uniform Resource Identifier - URI

2.2 Universelle Vokabularien mit XML

2.2.0 Warum (X)HTML nicht ausreicht...

2.2.1 XML Sprachstandard

2.2.2 XML DTD

2.2.3 XML Schema (XSD)

2.2.4 XML Sprachfamilie

2.2.5 Warum XML noch nicht ausreicht...

2.3 Einfache semantische Strukturen mit RDF und RDFS

2.4 Abfragesprache SPARQL

2.5 Semantik im WWW - μ formats, RDFa, GRDDL

2.6 Warum RDF/RDFS noch nicht ausreicht

2. Die Sprachen des Semantic Web

2.2 XML / 2.2.2 Document Type Definition (DTD)

35

- XML-DTD definiert eine kontextfreie Grammatik für ein XML-Dokument
- Extended Backus-Naur Form (**EBNF**)
 - --> Spezieller Interpreter notwendig
- XML-DTDs unterstützen lediglich den Datentyp "text"
- XML-DTDs unterstützen nicht das Vererbungskonzept

```
<adresse>
  <vorname> Max </vorname>
  <nachname> Mustermann </nachname>
  <straße> Sesamstr. </straße>
  <hausnummer> 49a </hausnummer>
  <ort plz="93123">Bad Sulzdetfurth</ort>
</adresse>
```

XML

```
<!DOCTYPE adresse[
  <!ELEMENT adresse (vorname+,
    nachname, straÙe, hausnummer,
    ort)>
  <!ELEMENT vorname (#PCDATA)>
  <!ELEMENT nachname (#PCDATA)>
  <!ELEMENT straÙe (#PCDATA)>
  <!ELEMENT ort (#PCDATA)>
  <!ATTLIST ort plz CDATA #REQUIRED>
```

```
....
]>
```

XML-DTD

2. Die Sprachen des Semantic Web

2.2 XML / 2.2.2 Document Type Definition (DTD)

36

DTD Elementdeklaration

- beschreibt die Einschränkungen des Inhalts eines Elements
- Syntax:
 <!ELEMENT Name (Definition)>
- Einziger atomarer Typ: #PCDATA
 (Parsed Character DATA)
- (a,b,c) Liste von Unterelementen
- (a|b|c) Alternativen
- Kardinalitäten:
 - * keinmal oder beliebig oft
 - + einmal oder beliebig oft
 - ? kein- oder einmal
 - (ohne Angabe): genau einmal
- EMPTY: Erzwingen von leerem Element

```
<!DOCTYPE adresse[
  <!ELEMENT adresse (vorname+,
    nachname, straße, hausnummer,
    ort)>
  <!ELEMENT vorname (#PCDATA)>
  <!ELEMENT nachname (#PCDATA)>
  <!ELEMENT straße (#PCDATA)>
  <!ELEMENT ort (#PCDATA)>
  <!ATTLIST ort plz CDATA #REQUIRED>
  ....
]>
```

XML-DTD

2. Die Sprachen des Semantic Web

2.2 XML / 2.2.2 Document Type Definition (DTD)

37

DTD Attributdeklaration

- Name-Zeichenkettenwert-Paar
- Assoziiert mit einem Element
- Syntax: `<!ATTLIST Element Attributname1 Typ1 Zusatz1 Attributname2 ...>`
- Typ:
 - **CDATA** Zeichenkette
 - **ID** OID
 - **IDREF** Referenzen
 - **IDREFS** Menge von Referenzen
- Zusatz:
 - **REQUIRED** zwingend
 - **IMPLIED** optional
 - (Initialwert)

```
<!DOCTYPE adresse[
  <!ELEMENT adresse (vorname+,
    nachname, straÙe, hausnummer,
    ort)>
  <!ELEMENT vorname (#PCDATA)>
  <!ELEMENT nachname (#PCDATA)>
  <!ELEMENT straÙe (#PCDATA)>
  <!ELEMENT ort (#PCDATA)>
  <!ATTLIST ort plz CDATA #REQUIRED>
  ....
]>
```

XML-DTD

2. Die Sprachen des Semantic Web

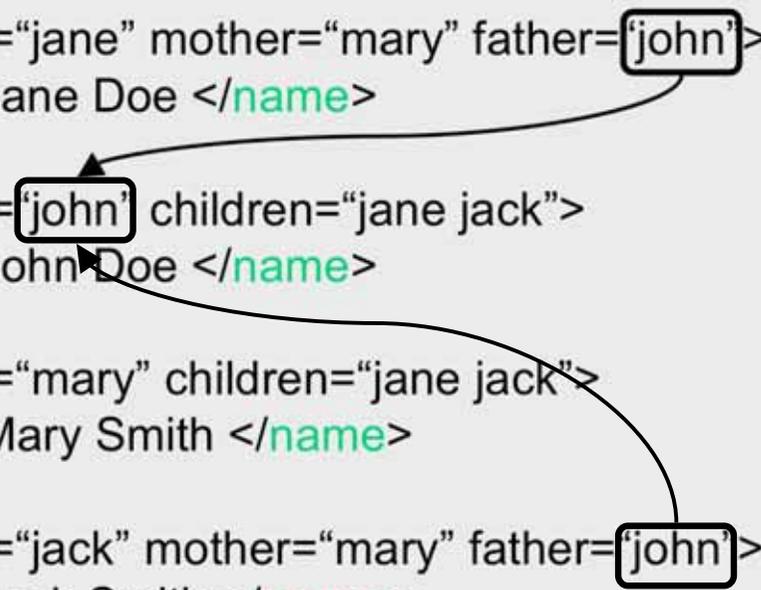
2.2 XML / 2.2.2 Document Type Definition (DTD)

38

DTD - OID und Referenzen

- DTDs erlauben die Deklaration von OIDs, Referenzen und Referenzmengen als Attribute

```
<family>
  <person id="jane" mother="mary" father="john">
    <name> Jane Doe </name>
  </person>
  <person id="john" children="jane jack">
    <name> John Doe </name>
  </person>
  <person id="mary" children="jane jack">
    <name> Mary Smith </name>
  </person>
  <person id="jack" mother="mary" father="john">
    <name> Jack Smith </name>
  </person>
</family>
```



XML

```
<!DOCTYPE family [
  <!ELEMENT family (person*)>
  <!ELEMENT person (name)>
  <!ELEMENT name (#PCDATA)>
  <!ATTLIST person
    id ID #REQUIRED
    mother IDREF #IMPLIED
    father IDREF #IMPLIED
    children IDREFS #IMPLIED>
]>
```

XML-DTD

2. Die Sprachen des Semantic Web

2.2 XML / 2.2.2 Document Type Definition (DTD)

39

DTDs - Bewertung

- DTDs sind kontextfreie Grammatiken
→ rekursive Definitionen möglich

- **Probleme:**

- Ungewollte Festlegung der Reihenfolge:

```
<!ELEMENT person ( name, phone ) >
```

- **Workaround:**

```
<!ELEMENT person ( ( name, phone ) | ( phone, name ) ) >
```

- Kann teilweise zu vage werden:

```
<!ELEMENT person ( ( name | phone | email ) * ) >
```

- Referenzen können nicht eingeschränkt (typisiert) werden

- Alle Elementnamen sind global in einem Namensraum

```
<!DOCTYPE [  
<!ELEMENT buch (titel,  
    subtitel?, autor+, preis)>  
<!ATTLIST buch ISBN ID  
    #REQUIRED>  
<!ELEMENT titel (#PCDATA)>  
<!ELEMENT subtitel (#PCDATA)>  
<!ELEMENT autor (#PCDATA)>  
<!ELEMENT preis (#PCDATA)>  
...]>
```

- 1 **06.11.2008 – Vorlesung Nr. 2**
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13

2. Die Sprachen des Semantic Web

2.1 Uniform Resource Identifier - URI

2.2 Universelle Vokabularien mit XML

2.2.0 Warum (X)HTML nicht ausreicht...

2.2.1 XML Sprachstandard

2.2.2 XML DTD

2.2.3 XML Schema (XSD)

2.2.4 XML Sprachfamilie

2.2.5 Warum XML noch nicht ausreicht...

2.3 Einfache semantische Strukturen mit RDF und RDFS

2.4 Abfragesprache SPARQL

2.5 Semantik im WWW - μ formats, RDFa, GRDDL

2.6 Warum RDF/RDFS noch nicht ausreicht