

Vorlesung Semantic Web



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Vorlesung im Wintersemester 2011/2012

Dr. Heiko Paulheim

Fachgebiet Knowledge Engineering

Einführung

- Was ist das Semantic Web?
- Bausteine des Semantic Web
- Grundlagen: URIs, Unicode, XML

Was ist das Semantic Web?

- Artikel von Tim Berners-Lee, Jim Hendler, und Ora Lassila:

„The Web is the killer app of the Internet.
The Semantic Web is another killer app
of that magnitude.“

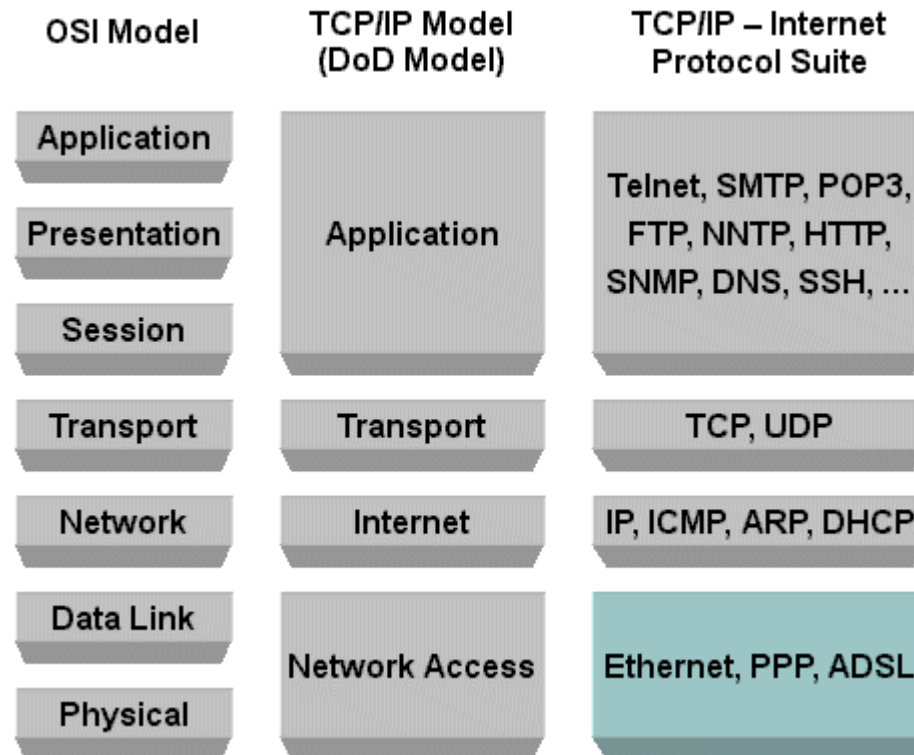


Berners-Lee et al. (2001): *The Semantic Web*. In: Scientific American, Mai 2001.

Web vs. Internet?



TECHNISCHE
UNIVERSITÄT
DARMSTADT



Chin-Shiuh Shieh (2000): *TCP/IP - Internet Protocol Suite and Ethernet*.
<http://bit.kuas.edu.tw/~csshie/teach/np/tcpip/index.html>

Das „klassische“ Web

- HTTP-Protokoll
- URLs
- HTML als Auszeichnungssprache
 - plus CSS, JavaScript, ...
 - plus weitere mehr oder weniger standardisierte Formate (GIF, JPEG, Flash, ...)
- Browser als universeller Client

Das „klassische“ Web

▪ Hypertext: verlinkte Dokumente

Das World Wide Web

Der Grundstein für das World Wide Web wurde in den 90er-Jahren durch [Tim Berners-Lee](#) am [CERN](#) gelegt.

...

Tim Berners-Lee

Tim Berners-Lee (geboren 1955) gilt als einer der Erfinder des [World Wide Web](#).

...

CERN

Das CERN ist ein europäisches Forschungszentrum in der Nähe von Genf.

...

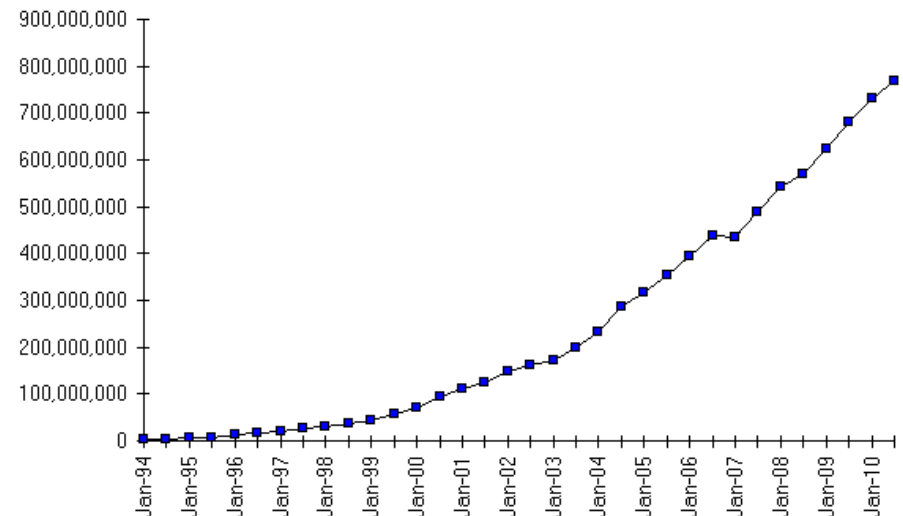
Das „klassische“ Web



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- 1989: Hypertext-Konzept von Tim Berners-Lee am CERN
- 1991: Erste HTML-Version (20 Elemente)
- 1993: Mosaic-Browser, ca. 500 Webserver (weltweit)
- 1994: Volltext-Suchmaschinen (WebCrawler, Lycos),
Gründung des W3C
- 1995: Amazon, eBay,
Internet Explorer
- 1998: Google
- 2001: Wikipedia
- 2003: Skype
- 2004: Facebook, Firefox
- 2006: Twitter, WikiLeaks
- ...

Internet Domain Survey Host Count



Source: Internet Systems Consortium (www.isc.org)

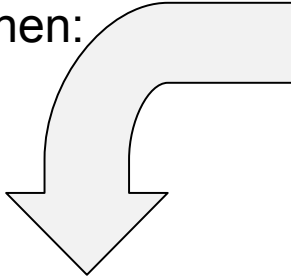


Das „klassische“ Web



TECHNISCHE
UNIVERSITÄT
DARMSTADT

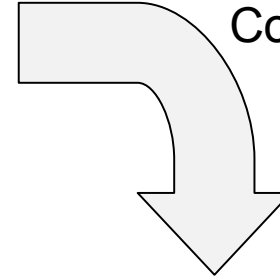
aus Sicht des
Menschen:



Dr. Mark Smith
Physician
Main St. 14
Smalltown
Mon-Fri 9-11 am
Wed 3-6 pm

```
<html>
...
<b>Dr. Mark Smith</b>
<i>Physician</i>
Main St. 14
Smalltown
Mon-Fri 9-11 am
Wed 3-6 pm
...
</html>
```

aus Sicht des
Computers:



Print in bold: „hmf298hmmhudsa“
Print in italics: „mj2i9ji0“
Print normal: „fdsah
02hfadsh0um2m0adsmf0ihm
asdfjkköfdsa298ndsfmij32mio
lk2mjpoimjiofdpmsajiomjm“

Informationssuche im Web



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Volltextsuche nach Stichwörtern (z.B. Google):

- „Mark Smith“
- „Physician in Smalltown“
- „Doctor in Smalltown“
- „Physician in Smalltown with opening hours on Wednesday afternoon“
- „Somebody in Smalltown who can fix a broken leg“

```
<html>
...
<b>Dr. Mark Smith</b>
<i>Physician</i>
Main St. 14
Smalltown
Mon-Fri 9-11 am
Wed 3-6 pm
...
</html>
```

=> „klassisches“ Web für Suche zu unflexibel

=> intelligente Agenten können es nicht nutzen

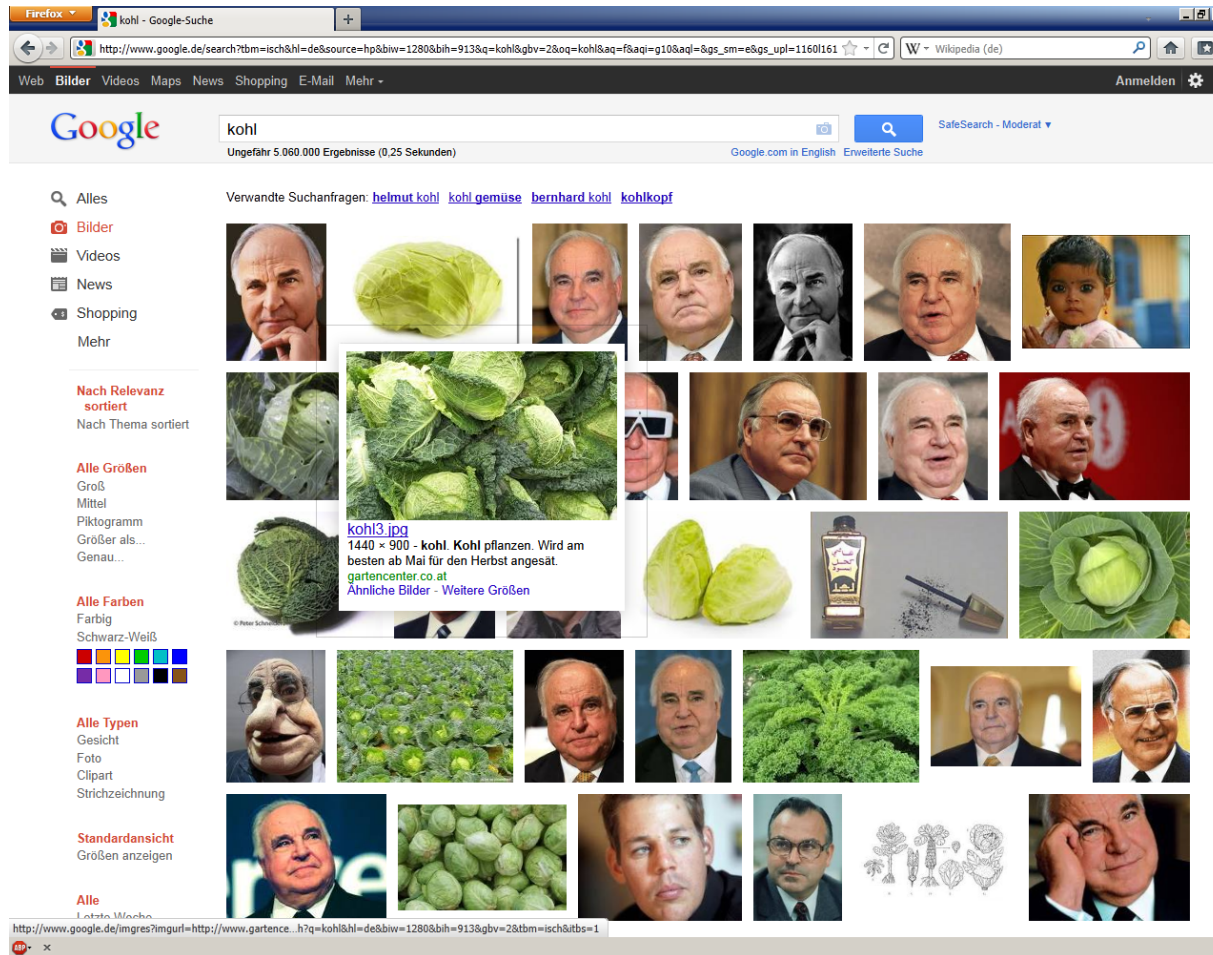
Probleme des Web

- Informationen finden
 - Stichwortbasierte Textsuche statt echter Fragen
 - verschiedene natürliche Sprachen
 - Homonyme/Polyseme
 - Synonyme
- Information verarbeiten
 - Formate (Encodings, Bilder, Videos, PDFs, ...)
- Information verwerten
 - Verteilt auf verschiedene Seiten
 - Bsp.: Information zu Buchautor auf Verlagsseite, Adresse auf Uni-Seite



<http://geekandpoke.typepad.com/geekandpoke/2011/08/coders-love-unicode.html>

Probleme des Web



Probleme des Web

Pfahls-Prozess: Richter rollen
die Skandale der Ära Kohl
neu auf

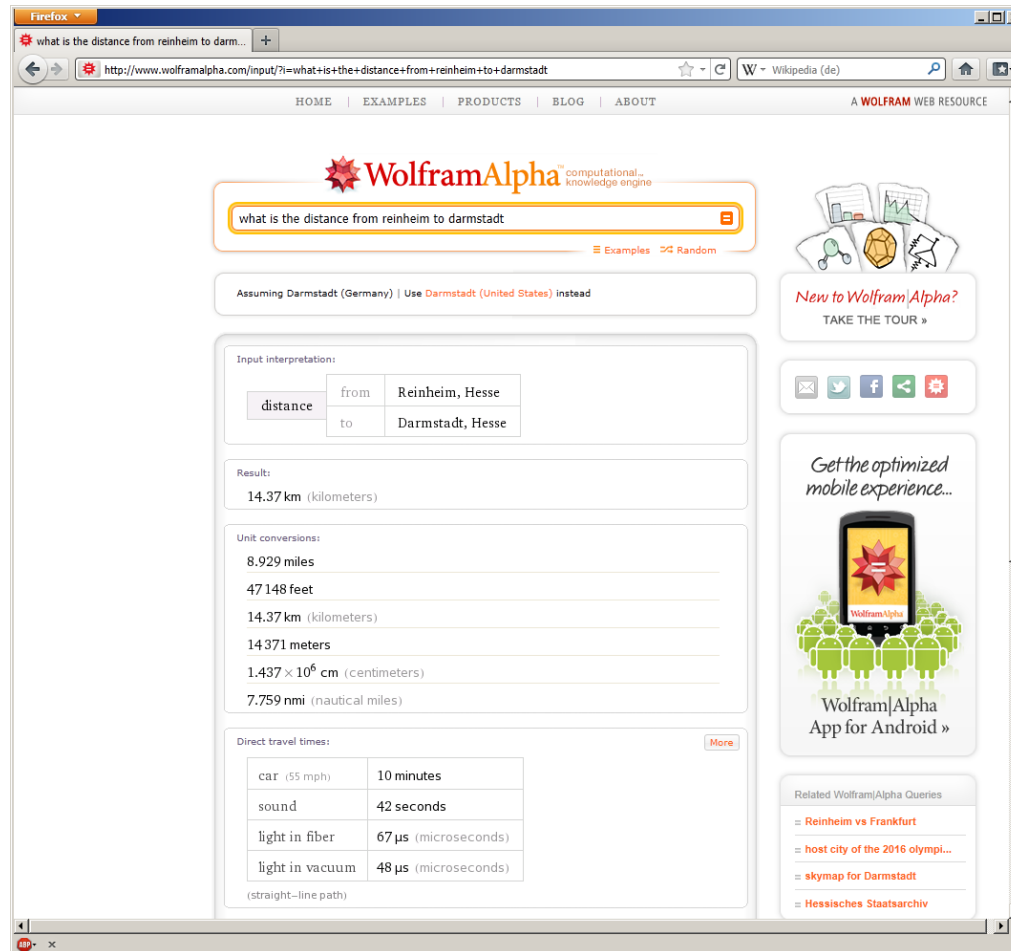
?

?

?



Beispiel: Wolfram Alpha



Firefox

what is the distance from reinheim to darmstadt

http://www.wolframalpha.com/input/?i=what+is+the+distance+from+reinheim+to+darmstadt

HOME | EXAMPLES | PRODUCTS | BLOG | ABOUT

A WOLFRAM WEB RESOURCE

WolframAlpha™ computational knowledge engine

what is the distance from reinheim to darmstadt

Examples Random

Assuming Darmstadt (Germany) | Use Darmstadt (United States) instead

Input interpretation:

distance	from	Reinheim, Hesse
	to	Darmstadt, Hesse

Result:

14.37 km (kilometers)

Unit conversions:

8.929 miles

47 148 feet

14.37 km (kilometers)

14 371 meters

1.437×10^6 cm (centimeters)

7.759 nmi (nautical miles)

Direct travel times:

car (55 mph)	10 minutes
sound	42 seconds
light in fiber	67 μ s (microseconds)
light in vacuum	48 μ s (microseconds)

(straight-line path)

New to Wolfram|Alpha?
TAKE THE TOUR »

Get the optimized mobile experience...

Wolfram|Alpha App for Android »

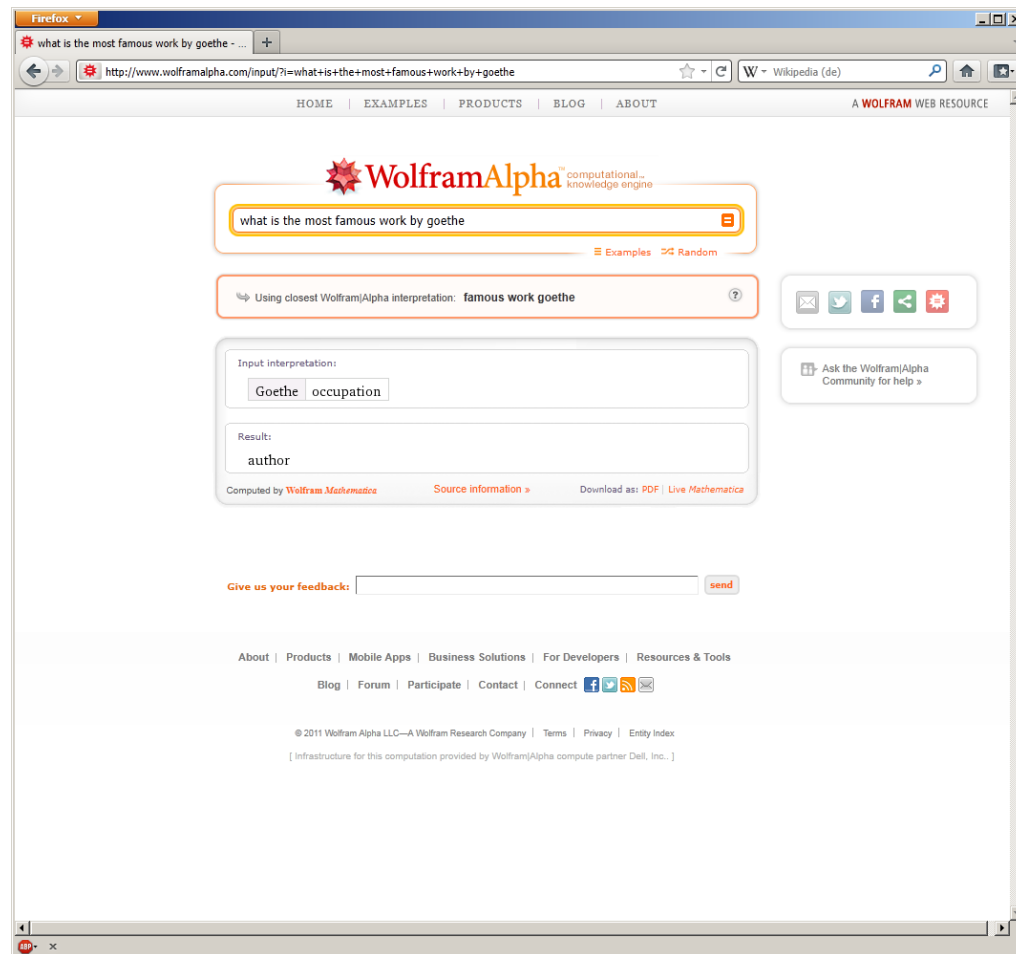
Related Wolfram|Alpha Queries

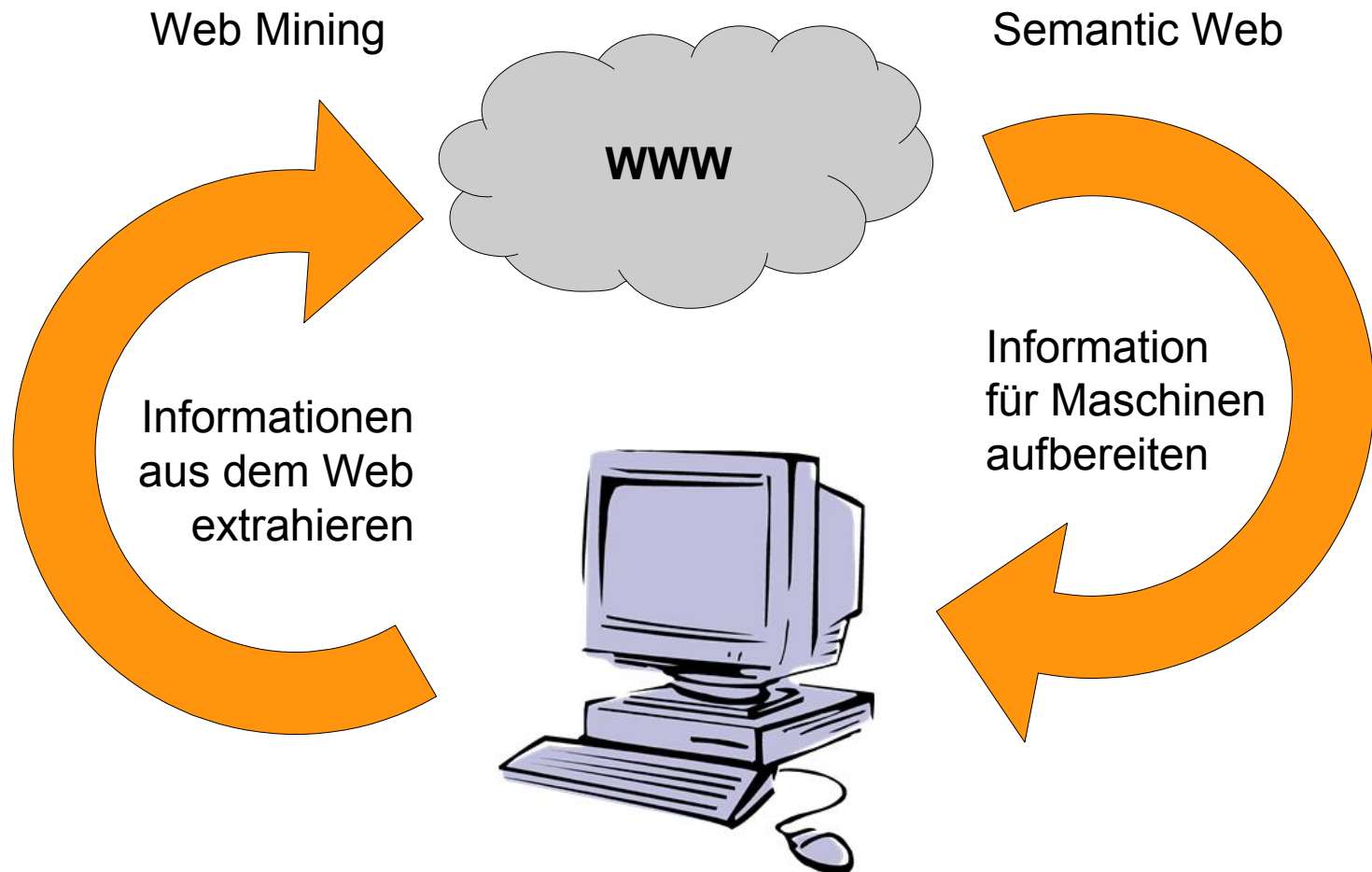
- Reinheim vs Frankfurt
- host city of the 2016 olympi...
- skymap for Darmstadt
- Hessisches Staatsarchiv

Beispiel: Wolfram Alpha



TECHNISCHE
UNIVERSITÄT
DARMSTADT





Idee des Semantic Web

- Information in maschinenlesbarer Form bereitstellen
- (Semantische) Verweise zwischen Seiten nutzbar machen
- Schlussfolgern ermöglichen
- Komplexe (nützliche!) Abfragen ermöglichen

Semantic Web – Aufbau



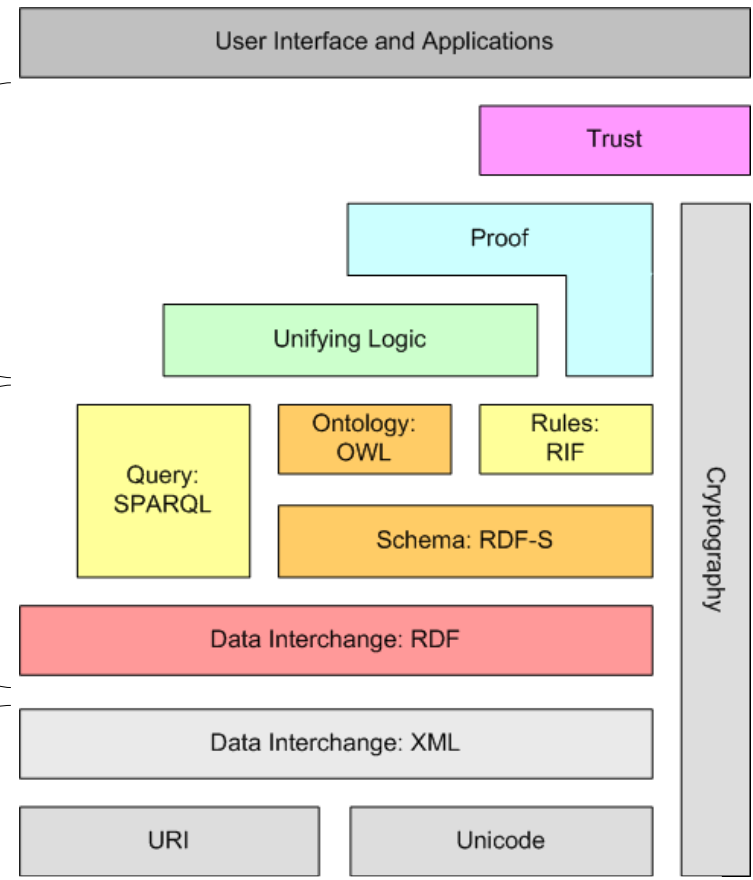
TECHNISCHE
UNIVERSITÄT
DARMSTADT



here be dragons...

Semantic-Web-
Technologie
(Fokus der Vorlesung)

Technische
Grundlagen

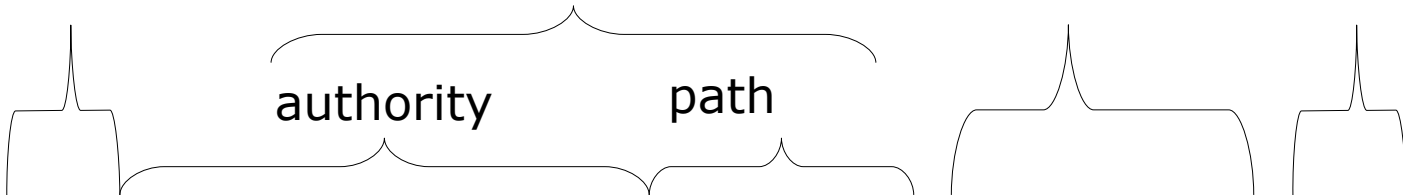


Berners-Lee (2009): *Semantic Web and Linked Data*
<http://www.w3.org/2009/Talks/0120-campus-party-tbl/>

Uniform Resource Identifiers (URIs)

- Als „Universal Resource Identifier“ erstmals vorgeschlagen von Tim Berners-Lee (1994): IETF RFC 1630
- Standard: IETF RFC 3986 (2005)
- Dienen zur Benennung und zum Auffinden von Ressourcen im Internet

URI = scheme ":" hier-part ["?" query] ["#" fragment]



http://example.com:8042/over/there?name=ferret#nose

- Uniform Resource Locators (IETF RFC 1738, 1994) sind eine Untermenge von URIs
- URIs können *beliebige* Dinge mit *beliebigen* Namen versehen
- ein URL ist der Name einer Resource [im Internet]
- URL-typische Präfixe:
 - http
 - ftp
 - mailto
 - telnet
 - file
 - ...

URLs im Web



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- Häufigste Verwendung:
(HTTP-)Links
- Links haben in der Regel
keine Metainformation!

Tim Berners-Lee

Tim Berners-Lee (geboren
1955) gilt als einer der Erfinder
des [World Wide Web](#).

...

Das World Wide Web

Der Grundstein für das World
Wide Web wurde in den 90er-
Jahren durch [Tim Berners-Lee](#)
am [CERN](#) gelegt.

...

<http://www.w3.org/WWW/>

<http://www.w3.org/People/Berners-Lee/>

Zeichensätze im Web

- ASCII („American Standard Code for Information Interchange“) ISO 646 (1963), 127 Zeichen, davon 95 druckbar:

!"#\$%&'()*+,-./0123456789:;<=>?
@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_
`abcdefghijklmnopqrstuvwxyz{|}~

- Erweiterung auf 8 Bit: ISO 8859-1 bis -16 (1998)
 - deckt Zeichen der europäischen Sprachen ab
 - bekannt ist insbesondere 8859-1 („Latin-1“)
- Das Web spricht aber viel mehr Sprachen...

وللحب علامات يقفوها الف
فأولها إدمان النظر، والعج
سراؤها، والمعبرة لضمائرها
بر لا يطرف، يتنقل بتنقل
ن مال، كالحرباء مع الشمس

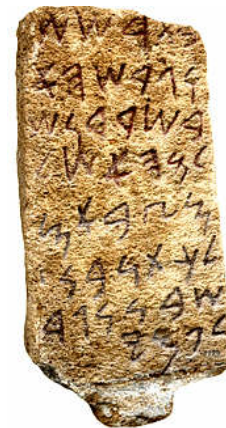
我爱中国
国中爱我

Unicode



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- ISO 10646
 - erste Version 1991 (Europa, Nahost, Indien)
 - Version 6.0 im Oktober 2010
 - 17 Codebereiche à 16 Bit
 - deckt selbst exotischste Sprachen ab



Unicode

00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	
10	11	12	13	14	15	16	17	18	19	1A	1B	1C	1D	1E	1F	
20	21	22	23	24	25	26	27	28	29	2A	2B	2C	2D	2E	2F	
30	31	32	33	34	35	36	37	38	39	3A	3B	3C	3D	3E	3F	
40	41	42	43	44	45	46	47	48	49	4A	4B	4C	4D	4E	4F	
50	51	52	53	54	55	56	57	58	59	5A	5B	5C	5D	5E	5F	
60	61	62	63	64	65	66	67	68	69	6A	6B	6C	6D	6E	6F	
70	71	72	73	74	75	76	77	78	79	7A	7B	7C	7D	7E	7F	
80	81	82	83	84	85	86	87	88	89	8A	8B	8C	8D	8E	8F	
90	91	92	93	94	95	96	97	98	99	9A	9B	9C	9D	9E	9F	
A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	AA	AB	AC	AD	AE	AF	
B0	B1	B2	B3	B4	B5	B6	B7	B8	B9	BA	BB	BC	BD	BE	BF	
C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	CA	CB	CC	CD	CE	CF	
D0	D1	D2	D3	D4	D5	D6	D7	D8	D9	DA	DB	DC	DD	DE	DF	
E0	E1	E2	E3	E4	E5	E6	E7	E8	E9	EA	EB	EC	ED	EE	EF	
F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	FA	FB	FC	FD	FE	FF	

■ Lateinische Schriften und Symbole

■ Lautschriften

■ Andere europäische Schriften

■ Nahost- und Südwestasiatische Schriften

■ Afrikanische Schriften

■ Südasiatische Schriften

■ Südostasiatische Schriften

■ Ostasiatische Schriften

■ CJK-Ideogramme

■ Kanadische Silben

■ Symbole

■ Diakritika

■ UTF-16-Surrogates und privater Nutzungsbereich

■ Verschiedene Zeichen

■ Nicht belegte Codebereiche

Quelle: Wikimedia Commons

Informationsrepräsentation in XML

XML (eXtensible Markup Language)

- Standardisiert vom W3C (1998)
- Universelles Datenaustauschformat



```
<physician>
  <name>Dr. Mark Smith</name>
  <address>
    <street>Main St.</street>
    <number>14</number>
    <city>Smalltown</city>
  </address>
  <telephone>
    <number>+44 123 456789</number>
  </telephone>
  <hours>
    <monday>9-11 am</monday>
    <tuesday>9-11 am</tuesday>
    ...
  </hours>
</physician>
```


- Tags (beliebig definierbar):
 - Paare:
`<physician> ... </physician>`
 - Empty-Element-Tags:
`<young />`
- Attribute:
`<physician location="Smalltown">`
- Schachtelung (genau ein Root-Element!):
`<physician>`
 `<address> ... </address>`
`</physician>`

XML: wohlgeformte Dokumente



TECHNISCHE
UNIVERSITÄT
DARMSTADT

```
<physician>
  <name>Dr. Mark Smith</name>
  <address>
    <street>Main St.</street>
    <number>14</number>
    <city>Smalltown</city>
  </address>
  <telephone>
    <number>+44 123 456789</number>
  </telephone>
  <hours>
    <monday>9-11 am</monday>
    <tuesday>9-11 am</tuesday>
    ...
  </hours>
</physician>
```

```
<physician>
  <name>Dr. Mark Smith</name>
  <address>
    <street>Main St.</street>
    <number>14</number>
    <city>Smalltown</city>
  <telephone>
    <number>+44 123 456789</number>
  </address>
  </telephone>
  <hours>
    <monday>9-11 am</monday>
    <tuesday>9-11 am</tuesday>
    ...
  </hours>
</physician>
```

- HTML-Dokumente sind i.d.R. keine wohlgeformten XML-Dokumente!

```
<p> Look at this! <img src=smiley.gif> <br>
```

- XHTML: HTML als wohlgeformte XHTML-Dateien
- Vom W3C standardisiert (seit 2000)

```
<p> Look at this!  <br/> </p>
```



XPath: Zugriff auf XML-Daten



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- Abfragesprache für XML
- Standardisiert vom W3C (1999, Version 2.0: 2010)

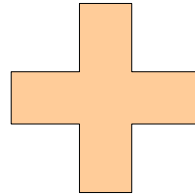
`/physician[name='Dr. Mark Smith']/telephone/number`

```
<physician>
  <name>Dr. Mark Smith</name>
  <address>
    <street>Main St.</street>
    <number>14</number>
    <city>Smalltown</city>
  </address>
  <telephone>
    <number>+44 123 456789</number>
  </telephone>
  <hours>
    <monday>9-11 am</monday>
    <tuesday>9-11 am</tuesday>
    ...
  </hours>
</physician>
```

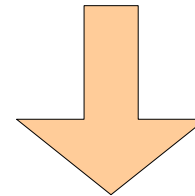
XSLT: Transformation von XML-Dokumenten

- Verarbeitung von XML-Dokumenten basierend auf Stylesheets
- Standardisiert vom W3C (1999)
- Verwendet XPath

```
<physician>
  <name>Dr. Mark Smith</name>
  <address>
    <street>Main St.</street>
    <number>14</number>
    <city>Smalltown</city>
  </address>
  <telephone>
    <number>+44 123 456789</number>
  </telephone>
  <hours>
    <monday>9-11 am</monday>
    <tuesday>9-11 am</tuesday>
    ...
  </hours>
</physician>
```



```
<xsl:template match="/pyhsician">
  <b>
    <xsl:value-of select="name"/>
  </b>
</xsl:template>
```



```
<b>Dr. Mark Smith</b>
```

Namensräume in XML



- Elemente gleichen Namens können mehrfach vorkommen
- ...aber mit anderem Inhalt (und anderer Semantik!)
- Wie können wir solche Elemente unterscheiden?

```
<physician>
  <name>Dr. Mark Smith</name>
  <address>
    <street>Main St.</street>
    <number>14</number>
    <city>Smalltown</city>
  </address>
  <telephone>
    <number>+44 123 456789</number>
  </telephone>
  <hours>
    <monday>9-11 am</monday>
    <tuesday>9-11 am</tuesday>
    ...
  </hours>
</physician>
```

Namensräume in XML



- Mit Präfix unterscheidbar (Schreibweise: `prefix:name`)
- Ein Namensraum ist ein URI
- Default-Namensraum möglich

```
<physician xmlns      ="http://www.med.com/physician"
           xmlns:addr="http://www.med.com/addr">
  <name>Dr. Mark Smith</name>
  <addr:address>
    <addr:street>Main St.</addr:street>
    <addr:number>14</addr:number>
    <addr:city>Smalltown</addr:city>
  </addr:address>
  <telephone>
    <number>+44 123 456789</number>
  </telephone>
  <hours>
    <monday>9-11 am</monday>
    <tuesday>9-11 am</tuesday>
    ...
  </hours>
</physician>
```

XML: Document Type Definition (DTD)



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- Definiert gültige Elemente für einen XML-Dokumenttyp
 - Name
 - zulässige Attribute
 - zulässige Kind-Elemente
- DTD ist Teil der W3C-XML-Spezifikation
- XML-Dokumente, die auf eine DTD passen, heißen „gültig“

XML: Document Type Definition (DTD)



TECHNISCHE
UNIVERSITÄT
DARMSTADT

```
<!DOCTYPE physician [  
  
  <!ELEMENT physician (  
    name,  
    address*,  
    telephone?,  
    fax?,  
    hours)>  
  
  <!ELEMENT address (  
    street,  
    number,  
    city)>  
  
  <!ELEMENT street (#PCDATA)>  
  
    ...  
  ]>
```



```
<!DOCTYPE physician SYSTEM  
  "physician.dtd">  
  
<physician>  
  <name>Dr. Mark Smith</name>  
  <address>  
    <street>Main St.</street>  
    <number>14</number>  
    <city>Smalltown</city>  
  </address>  
  <telephone>  
    <number>+44 123 456789</number>  
  </telephone>  
  <hours>  
    <monday>9-11 am</monday>  
    <tuesday>9-11 am</tuesday>  
    ...  
  </hours>  
</physician>
```

XML: Document Type Definition (DTD)

- Definition von Kind-Elementen und deren Reihenfolge:

```
<!ELEMENT address(street,nr,addtl*,zip,city,state?) >
```

- ? und * markieren optionale und wiederholbare Elemente

- Definition von Attributlisten:

```
<!ATTLIST person title CDATA>
```

- mögliche Zusätze: #REQUIRED, #FIXED, #IMPLIED, "..."
- Aufzählung zulässiger Werte: (dr|prof)

- Definition von Entitäten:

```
<!ENTITY sw "Semantic Web">
```

- Können als Abkürzung verwendet werden: &sw;

- W3C-Standard (seit 2004)
- XML-Schema-Dateien sind selbst XML-Dateien
- Flexibler als DTDs:
 - Minimale und maximale Anzahl von Elementen
 - Kombinationen von Elementen (entweder oder, Auswahl ohne Reihenfolge, ...)
 - Datentypen (Zahlen, Daten, ...), eigene Definitionen möglich
 - Unterstützung von Namespaces
 - Modulare Schemata möglich

XML Schema



TECHNISCHE
UNIVERSITÄT
DARMSTADT

```
<xs:schema elementFormDefault="qualified"
xmlns:xs="http://www.w3.org/2001/XMLSchema">
```

```
  <xs:element name="physician">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="name"
          type="xs:string">
        <xs:element name="address">
          <xs:complexType>
            <xs:sequence>
              <xs:element name="street"
                type="xs:string">
              ...
            </xs:sequence>
          </xs:complexType>
        </xs:element>
        ...
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

```
<physician xmlns:xsi=
"http://www.w3.org/2001/XMLSchema-instance"
xsi:noNamespaceSchemaLocation=
"physician.xsd">
  <name>Dr. Mark Smith</name>
  <address>
    <street>Main St.</street>
    <number>14</number>
    <city>Smalltown</city>
  </address>
  <telephone>
    <number>+44 123 456789</number>
  </telephone>
  <hours>
    <monday>9-11 am</monday>
    <tuesday>9-11 am</tuesday>
    ...
  </hours>
</physician>
```

XML Schema – modulare Schemata



TECHNISCHE
UNIVERSITÄT
DARMSTADT

```
<xs:schema elementFormDefault="qualified"
xmlns:xs="http://www.w3.org/2001/XMLSchema"
xmlns:addr="http://www.address.com/">
```

```
  <xs:import
    namespace="http://www.address.com/"
    schemaLocation="address.xsd"/>
  <xs:element name="physician">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="name"
          type="xs:string">
          <xs:element ref="addr:address" />
          ...
        </xs:sequence>
      </xs:complexType>
    </xs:element>
  </xs:schema>
```

```
<xs:schema elementFormDefault="qualified"
xmlns:xs="http://www.w3.org/2001/XMLSchema">
```

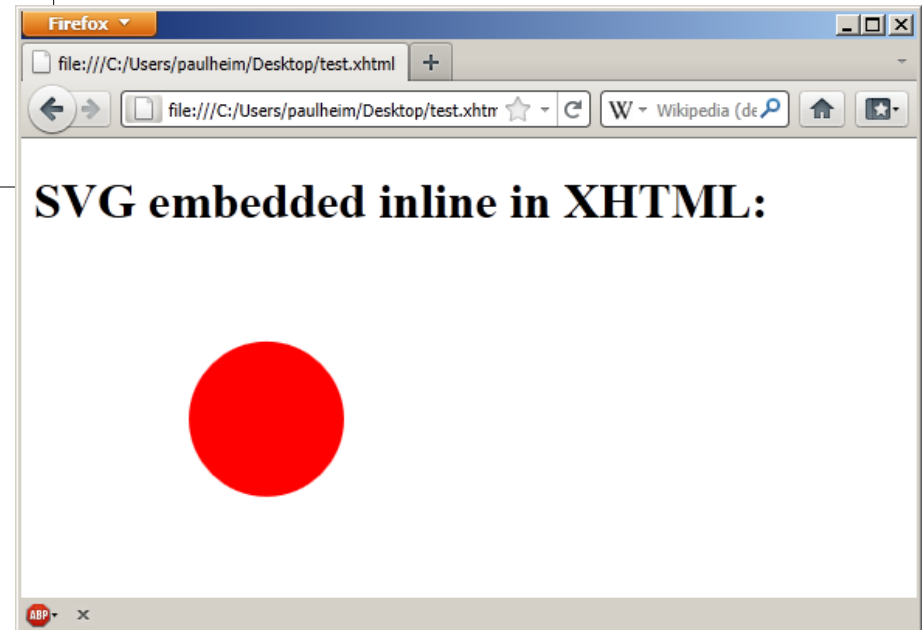
```
  <xs:element name="address">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="street"
          type="xs:string">
          ...
        </xs:sequence>
      </xs:complexType>
    </xs:element>
  </xs:schema>
```

Beispiel: Modulare Schemata in XHTML



TECHNISCHE
UNIVERSITÄT
DARMSTADT

```
<html xmlns="http://www.w3.org/1999/xhtml"
      xmlns:svg="http://www.w3.org/2000/svg">
  <body>
    <h1>SVG embedded inline in XHTML:</h1>
    <svg:svg width="300px" height="200px">
      <svg:circle cx="150" cy="100" r="50"
        fill="#ff0000"/>
    </svg:svg>
  </body>
</html>
```



https://developer.mozilla.org/En/SVG:Namespaces_Crash_Course

- Alternative zu XML Schema
- Standardisiert (ISO/IEC 19757-2, 2003)



International
Organization for
Standardization

- Unterschiede zu XML Schema:
 - weniger flexibel in Kardinalitäten
 - kein expliziter Link zwischen XML-Dokument und Schema
 - kein eigenes Typsystem (kann XML Schema Datatypes verwenden)
 - bessere Unterstützung von schwach strukturierten Inhalten
 - XML-basierte und kompakte Darstellung möglich

```
<element name="physician"
xmlns="http://relaxng.org/ns/structure/1.0">
  <element name="address">
    <group>
      <element name="street">
        <text/>
      </element>
      <element name="number">
        <text/>
      </element>
      <element name="city">
        <text/>
      </element>
    </group>
  </element>
  ...
</element>
```

```
<physician>
  <name>Dr. Mark Smith</name>
  <address>
    <street>Main St.</street>
    <number>14</number>
    <city>Smalltown</city>
  </address>
  <telephone>
    <number>+44 123 456789</number>
  </telephone>
  <hours>
    <monday>9-11 am</monday>
    <tuesday>9-11 am</tuesday>
    ...
  </hours>
</physician>
```


- Fokus: *Validierung* von XML-Dokumenten
- Standardisiert (ISO/IEC 19757-3, 2006)
- Verwendet Regeln zur Validierung
- Regeln basieren auf XPath-Ausdrücken
- Ausführung der Validierung mit XSLT möglich
- Fehlermeldungen werden direkt im Schema definiert



International
Organization for
Standardization

```
<?xml version="1.0" encoding="utf-8"?>
<schema
  xmlns="http://purl.oclc.org/dsdl/schematron"
  <title>Physician validation schema</iso:title>

  <pattern>
    <rule context="physician">
      <assert test="address">A physician must have an
        address</assert>
      ...
      <assert test="hours/*">Hours must not be empty
        </assert>
    </rule>
  </pattern>
</schema>
```

```
<physician>
  <name>Dr. Mark Smith</name>
  <address>
    <street>Main St.</street>
    <number>14</number>
    <city>Smalltown</city>
  </address>
  <telephone>
    <number>+44 123 456789</number>
  </telephone>
  <hours>
    <monday>9-11 am</monday>
    <tuesday>9-11 am</tuesday>
    ...
  </hours>
</physician>
```

Schematron – Ausführung

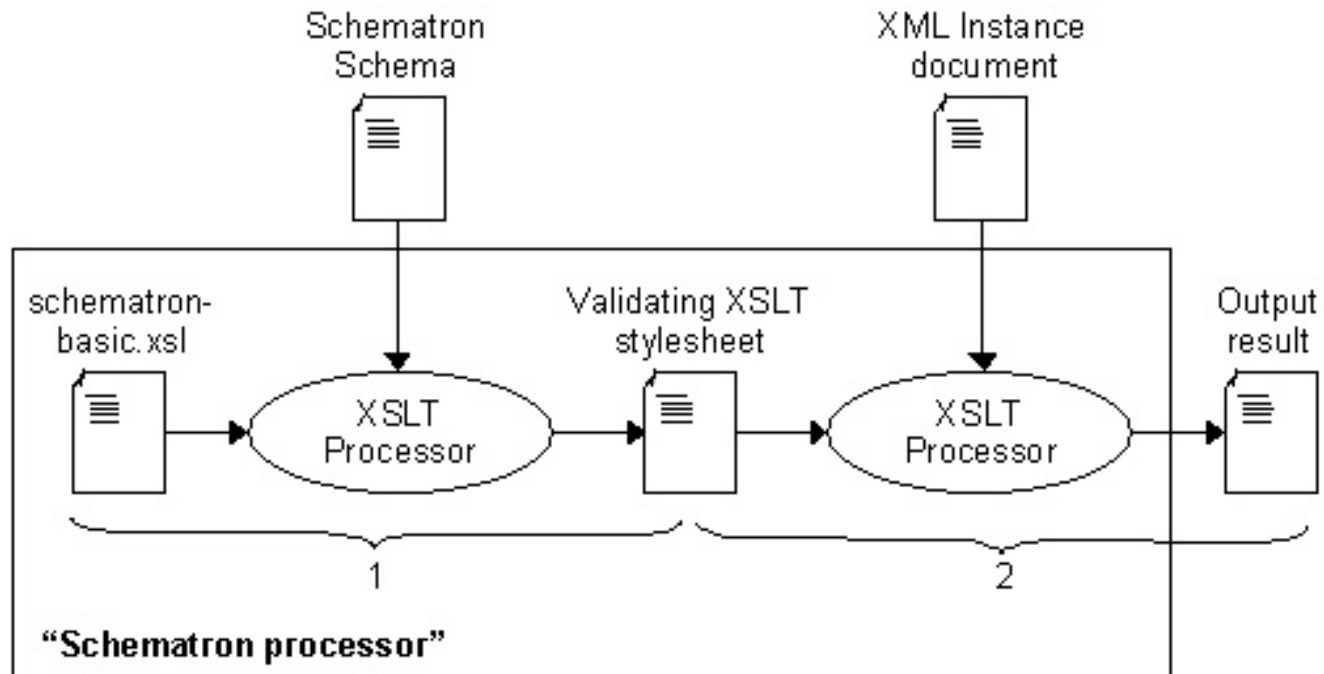


Figure 1: Schematron processing

XML Schema, DTD & Co – Was wird hier definiert?



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- Syntax – σύνταξις („zusammen“ + „Ordnung“)
 - Welche Elemente gibt es?
 - Wie werden sie angeordnet?
 - Welche Kombinationen sind zulässig?
- Gegensatz: Semantik - σημαίνει („bezeichnen“)
 - Wie interpretiert man den Inhalt von Elementen?
 - In welchem *inhaltlichen* Zusammenhang stehen Elemente?

Syntax und Semantik: Exkurs in die Sprachwissenschaft



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- Syntax: wie bildet man korrekte Wörter und Sätze?

„Dieser Satz kein Verb.“

„Die träumende Lampe ~~schenkst~~ schenkt dem müden
Wasserhahn unaufmerksam ~~eine~~ einen abgesägten Saft.“

- Semantik: was bedeutet ein Wort/Satz/Text?

Syntax und Semantik: Exkurs in die Sprachwissenschaft



TECHNISCHE
UNIVERSITÄT
DARMSTADT

The screenshot shows the Duden online website in a Firefox browser. The page title is 'Lampe - Artikel - duden.de'. The URL is 'http://www.duden.de/rechtschreibung/Lampe_Leuchtkoerper_Leuchte'. The page features a search bar with the text 'Suchen' and a 'Text überprüfen' button. Below the search bar, the word 'Lampe' is displayed with its grammatical information: 'Wortart: Substantiv, feminin' and 'Häufigkeit: ■■■■'. The 'Rechtschreibung' section shows the word 'Lampe' with a 'Nach oben' button. The 'Bedeutungsübersicht' section is circled in red and contains two numbered entries: 1. 'als Träger einer künstlichen Lichtquelle (besonders von Glühbirnen) dienendes, je nach Zweck sehr unterschiedlich gestaltetes, hängendes, stehendes oder auch frei bewegliches Gerät' and 2. '(besonders Fachsprache) künstliche Lichtquelle (z. B. Glühlampe)'. The 'Wussten Sie schon?' section contains a 'Nach oben' button. The 'Inhalte:' section lists various links: Rechtschreibung, Bedeutungsübersicht, Wussten Sie schon?, Synonyme zu Lampe, Aussprache, Herkunft, Grammatik, Typische Verbindungen (computergeneriert), Bedeutungen, Beispiele und Wendungen, and Blättern. The 'Kennen Sie schon?' section is also visible.



XML Schema, DTD & Co – Was wird hier definiert?



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Personalverzeichnis
des Krankenhauses:

```
<physician>
  <name>Dr. Mark Smith</name>
  <address>
    <street>Main St.</street>
    <number>14</number>
    <city>Smalltown</city>
  </address>
  <telephone>
    <number>+44 123 456789</number>
  </telephone>
  <hours>
    <monday>9-11 am</monday>
    <tuesday>9-11 am</tuesday>
    ...
  </hours>
</physician>
```

(wahrscheinlich)
die Privatadresse

Gelbe Seiten:

```
<physician>
  <name>Dr. Mark Smith</name>
  <address>
    <street>Main St.</street>
    <number>14</number>
    <city>Smalltown</city>
  </address>
  <telephone>
    <number>+44 123 456789</number>
  </telephone>
  <hours>
    <monday>9-11 am</monday>
    <tuesday>9-11 am</tuesday>
    ...
  </hours>
</physician>
```

(wahrscheinlich)
die Adresse der Praxis

?
=

XML Schema, DTD & Co – Was wird hier definiert?



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- XML Schema / DTD definiert die *Syntax* eines XML-Dokuments, nicht die *Semantik*
- Tag-Namen sind für Maschinen nicht a priori interpretierbar
 - das macht die Informationssuche nicht leichter...
 - Semantik der Daten (ver-)steckt hart verdrahtet in der Anwendung
- Das Semantic Web soll hier Abhilfe schaffen
 - *Semantic Web ist/kann mehr als XML!*

```
<2nf3oiü*>
  <34f0>Dr. Mark Smith</34f0>
  <rm4935r>
    <e2m4>Main St.</e2m4>
    <dur3>14</dur3>
    <jfa34>Smalltown</jfa34>
  </rm4935r>
  <d24r3fmö>
    <deß5>+44 123 456789</deß5>
  </d24r3fmö>
  <vsfif>
    <f02>9-11 am</f02>
    <fj9>9-11 am</fj9>
    ...
  </vsfif>
</2nf3oiü*>
```


Zusammenfassung



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- Probleme des klassischen Web
 - Nicht für Maschinen nutzbar
- URIs
 - eindeutige Identifier für Ressourcen
 - URL = dereferenzierbarer URI
- Unicode
 - ein einheitlicher Zeichensatz für alle
- XML
 - XPath
 - XSLT
 - Schemasprachen

Semantic Web – Aufbau



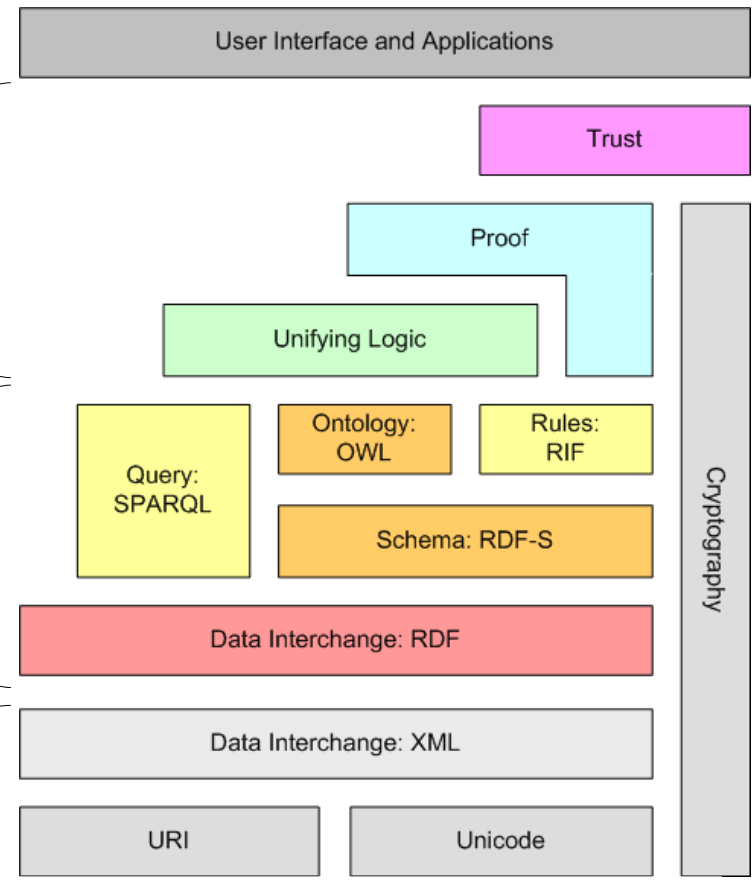
TECHNISCHE
UNIVERSITÄT
DARMSTADT



here be dragons...

Semantic-Web-
Technologie
(Fokus der Vorlesung)

Technische
Grundlagen



Berners-Lee (2009): *Semantic Web and Linked Data*
<http://www.w3.org/2009/Talks/0120-campus-party-tbl/>



Vorlesung Semantic Web



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Vorlesung im Wintersemester 2011/2012

Dr. Heiko Paulheim

Fachgebiet Knowledge Engineering