

<XML/>

Eine kleine Einführung

Historisches

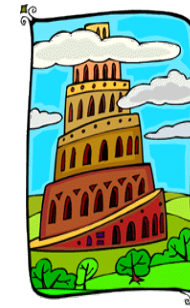
- Am Anfang war...



- Es kam das Wort...



- Das Chaos nahm seinen Lauf...



- Und wurde auf die Spitze getrieben...



Historisches

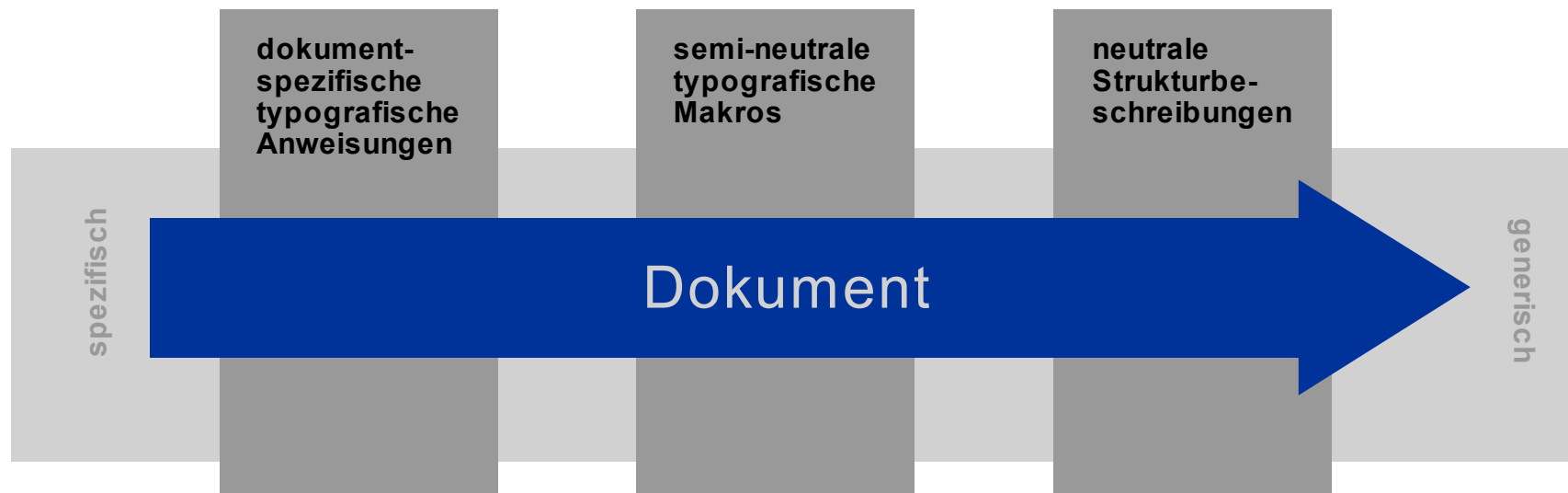
- Digitaler Datenaustausch wurde notwendig
- Formatanweisungen sollten helfen
 - Makros für Strukturen (Header, Footnote, Margin usw.)
 - Leider applikationsabhängig
 - Leider eingeschränkte Menge
- Probleme, Probleme, Probleme
- Etwas neues musste her... und Goldfarb sprach:

GML
das war 1969

```
:h1.Chapter 1: Introduction
:p.GML supported hierarchical containers, such as
:ol
:li.Ordered lists (like this one),
:li.Unordered lists, and
:li.Definition lists
:eol.
```

SGML — Standard General Markup Language

- Seit 1985 ISO Standard (ISO 8879)
- Trennung von Form (Layout) und Inhalt
- Deskriptiv – keine speziellen Verarbeitungsanweisungen
- Unspezifisch – Erstellung eigener Anwendungen mittels DTD
- Sehr mächtig – aber sehr komplex (*Design by Committee*)



HTML — Hypertext Markup Language

- SGML Dialekt mit Hypertextelementen
 - Einfache Hyperlinks, unidirektional
- Ab 1989 von Tim Berners-Lee am CERN entwickelt
- HyTime Implementierung (ISO 10744)
- Basis des World Wide Web (WWW)

Der Sündenfall durch Netscape

- vermischen von Form (Layout) und Inhalt
<center> background color usw...

XML — Extensible Markup Language

- Grundlegende Idee 1996 von Jon Bosak (Sun)
- Wurde durch W3C entwickelt
- SGML Subset - strenger definiert
- Seit Februar 1998 W3C Empfehlung (Version 1.0)
 - Seit 2004 Version 1.1
- Basis des XML Universums
- Keine Erweiterung von HTML
 - XHTML ist XML Anwendung

XML – Begriffe I

- **DTD – Document Type Descriptor**
Kann eine XML Anwendung formal beschreiben (Dokumentstruktur)
- **XML Anwendung**
Ein, durch eine DTD beschriebenes Dokumentenformat – z.B. XHTML
- **Dokument**
Instanz einer XML Anwendung – Beispielsweise eine XHTML Seite
- **Element**
 - Inhaltseinheit eines Dokument. Beispielsweise `<name>sascha</name>`
 - Elemente können (und müssen) hierarchisch geschachtelt werden.
 - Elemente werden auch als Knoten bezeichnet.
- **Tag**
Bezeichnung für den jeweils öffnenden und schliessenden Teil eines Elementes.
- **Attribut**
Informationen innerhalb von öffnenden Tags. Im Gegensatz zu SGML bestehen Attribute immer aus einem Namen und einem Wert.

- XML Anwendungen können durch DTDs beschrieben werden
 - Ein XML Dokument ist ohne DTD nicht validierbar
 - Nicht validierbare Dokumente müssen mindestens *wohlgeformt* sein
- XML ist hierarchisch
 - Öffnende Tags müssen geschlossen werden
<text></text> oder <text/>

```
<!-- FACT definition for simple triple. -->
<!ELEMENT fact (#PCDATA | subject | predicate | object)>
<!ATTLIST fact source CDATA #REQUIRED>

<!ELEMENT subject (#PCDATA) >
<!ATTLIST subject type CDATA #REQUIRED>

<!ELEMENT predicate (#PCDATA) >
<!ATTLIST predicate type CDATA #REQUIRED>

<!ELEMENT object (#PCDATA) >
<!ATTLIST object type CDATA #IMPLIED>
```


XML – Begriffe II

- XML Prolog
Beschreibt Version, Deklaration und Encoding eines Dokuments
- Stylesheet
Beschreibt die Formatierung eines XML Dokumentes
- Stylesheetanweisungen
Beschreibt die Quelle einer Formatierung ausserhalb des Dokumenteninhaltes
- <!DOCTYPE
Bindet in einem Dokument einen Verweis auf die DTD oder diese komplett ein
- Namespace
Erlaubt das Vermischen mehrerer Dokumentenformate. Wird z.B. im Semantik Web und Webservices Bereich stark eingesetzt
- XML Schema
Ermöglicht erweiterte Definitionen von Struktureigenschaften (z.B. Mengen)
 - Nur noch DTD für Schemadefinition nötig
 - Es existieren konkurrierende Schema Standards (W3C, Relax)

XML - Beispiel

```
<?xml version='1.0' encoding='UTF-8'?>    <!-- Prolog mit XML Version und Encoding -->

<!-- Stylesheetformatierungsanweisung -->
<?xml-stylesheet type='text/xsl' href='http://www.speexx.de/rdf-style.xsl' ?>

<!-- DTD Deklaration fehlt, da es für RDF keine gibt. -->

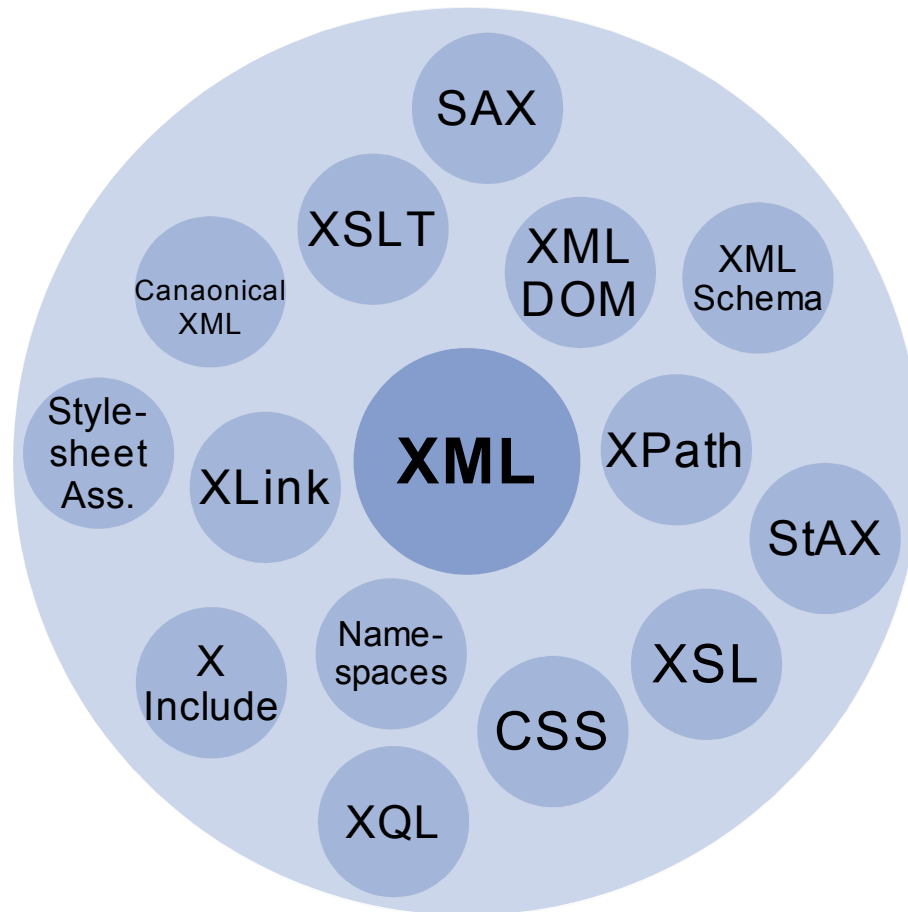
<!-- Beginn des Dokumentes -->
<rdf:RDF xmlns:rdf='http://www.w3.org/1999/02/22-rdf-syntax-ns#'
  xmlns='http://purl.org/rss/1.0/'
  xmlns:dc='http://purl.org/dc/elements/1.1/'
  xml:base='http://www.speexx.de'>

  <channel rdf:about='http://www.speexx.de/'>
    <title>SpeexX Home Links</title>
    <link>http://www.speexx.de/</link>
    <description>SpeexX Home Links</description>
    <items>
      <rdf:Seq>
        <rdf:li resource='http://www.speexx.de/ocean/annotator/index.html' />
      </rdf:Seq>
    </items>
  </channel>

  <!-- Beschreibung eines Eintrages -->
  <item rdf:about='http://www.speexx.de/ocean/annotator/index.html'>
    <dc:date>2005-03-18</dc:date>
    <link>http://www.speexx.de/ocean/annotator/index.html</link>
    <title>Erster Release der SpeexX Ocean Annotator API für semantische Textannotationen</title>
    <dc:description/>
  </item>

</rdf:RDF>
```

Das XML Universum



Das XML Universum

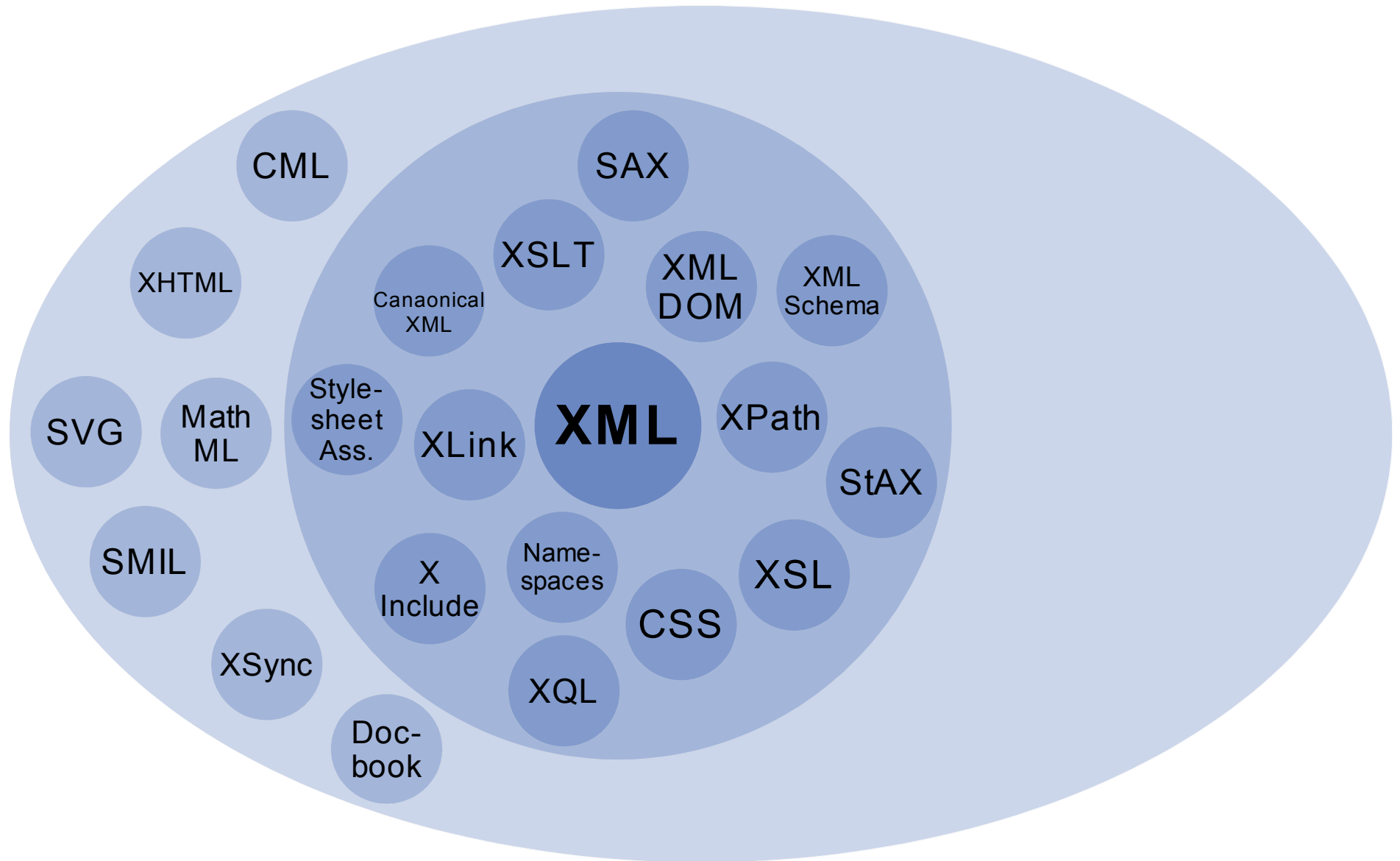
- **XPath**
Abfragesprache zur Ermittlung von Elementknoten
- **XQL – XML Query Language**
Abfragesprache zur Ermittlung von Elementknotenmengen
- **XSL, XSLT, CSS, Stylesheet Assoziation**
Sprachen zum formatieren und transformieren von XML Dokumenten
 - XSL – XML Stylesheet Language
 - XSLT - XML Stylesheet Language Transformation
 - CSS – Cascading Stylesheets – Wird für XHTML Layouts verwendet
- **Canonical XML**
Regelt die Formatfreiheit von XML.

In XML können Elemente und Attribute relativ frei angeordnet werden
Dabei gibt es Probleme, wenn XML Elemente kryptographisch signiert werden müssen
- **XInclude**
Spezifikation zum zusammenführen von verteilten XML Dokumenten

Das XML Universum

- **DOM**
Document Object Model
API zum traversieren durch XML Bäume – Sehr ressourcenhungrig
Ein gutes Beispiel für das Antipattern *Design by Committee*
 - **SAX**
Simple API for XML
Sogenanntes *push*-System zur Verarbeitung von XML Dokumenten
 - **StAX**
Streaming API for XML
Java basierte API zum durchlaufen von XML Dokumenten
- Vereint die Vorteile von DOM und SAX und eliminiert deren Nachteile teilweise

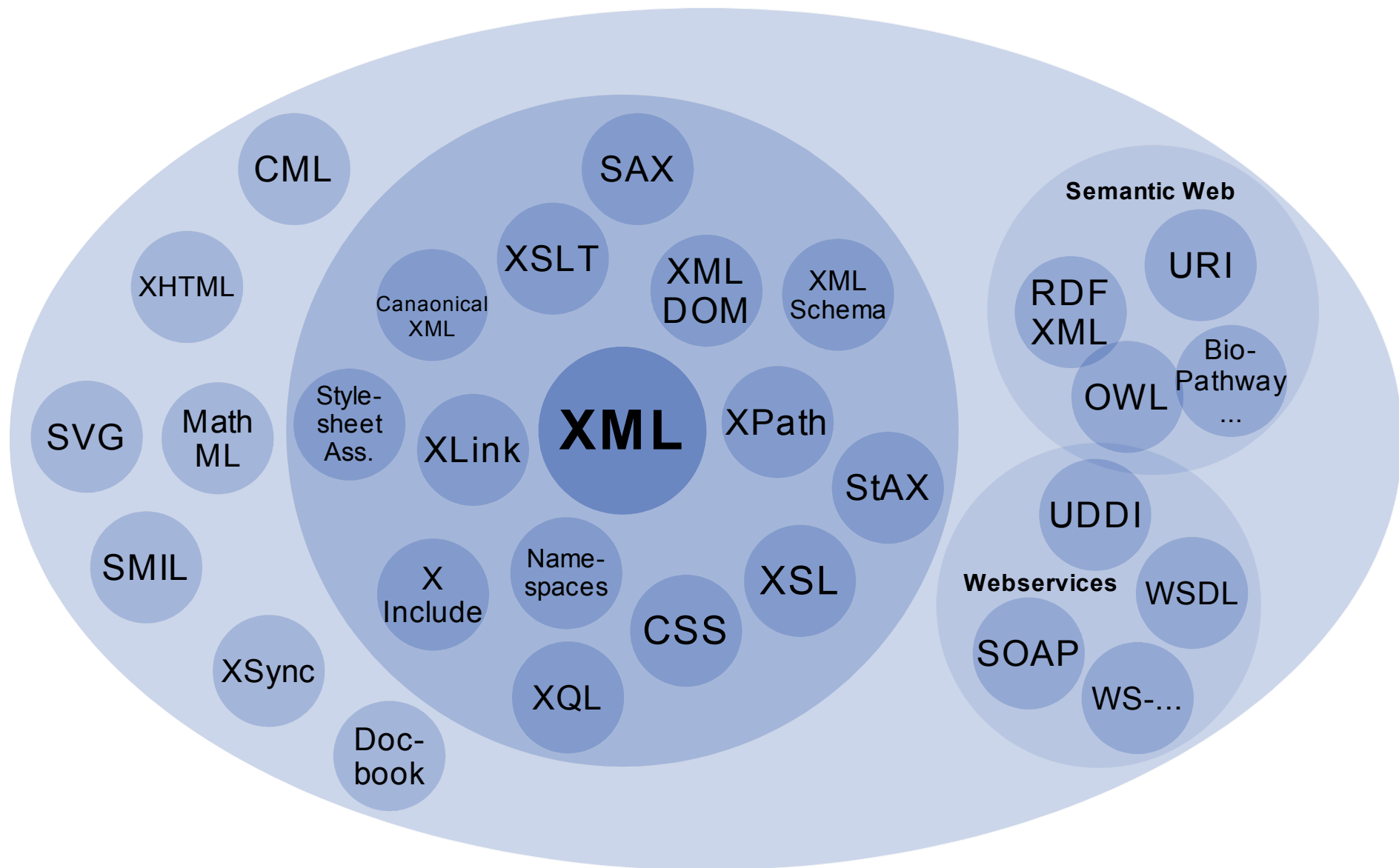
Das XML Universum



Das XML Universum

- **XHTML**
HTML in XML definiert
- **CML**
Chemical Markup Language
- **SVG**
Scalable Vector Graphics
- **Docbook**
Format für von technische und andere Dokumentationen
- **SMIL**
Synchronized Multimedia Integration Language

Das XML Universum



Das XML Universum – Semantic Web

- **URI**
Uniform Resource Identifier – dient z.B. dem Auffinden von Dokumenten
Subtyp ist z.B. die URL (<http://www.speexx.de/>)
- **RDF/XML**
Resource Description Framework in der XML Notation
- **OWL**
Web Ontology Language – Sprache zur Definition von Ontologien
- **BioPathway und andere
OWL Anwendungen**

Das XML Universum – Webservice

- SOAP
SOAP ist SOAP ist SOAP – SOAP war mal *Simple Object Access Protocol*
Definiert einen RPC Mechanismus
- WSDL
Webservice Description Language
Beschreibt, wie eine SOAP Nachricht auszusehen hat
- UDDI
Universal Description, Discovery and Integration
 - Eine Art Gelbe Seiten für Webservices
 - In Kombination mit RDF und OWL für das semantische auffinden von Webservices möglich

Fragen