

9 SGML

9.1 Einführung

9.2 Geschichtlicher Überblick

9.3 Mißverständnisse über SGML

9.4 Nutzung von SGML

9.5 SGML-Deklaration

9.6 Beispiele

Ein Textstück in verschiedenen Schriftarten und -graden.

Dies ist ein vollständiger Satz.

Dies ist ein vollständiger Satz.

Dies ist ein vollständiger Satz.

Dies ist ein vollständiger Satz.

Dies ist ein vollständiger Satz.

Dies ist ein vollständiger Satz.

Dies ist ein vollständiger Satz.

Dies ist ein vollständiger Satz.

DIES IST EIN VOLLSTÄNDIGER SATZ.

Tex-Beispiel:

```
% sample.tex
\documentclass[pdf,azure,slideBW,nocolorBG]{prosper}
\begin{document}
```

```
\begin{slide}{Folien mit Prosper}
```

Formeln:

```
\[
\int_{-\infty}^{\infty} e^{-x^2} \, dx = \sqrt{\pi}
\]
```

```
\[
e^{i\pi} = -1
\]
```

Beobachtungen:

```
\begin{itemize}
\item  $\sqrt{2}$  ist irrational.
\item  $e$  und  $\pi$  sind transzendent.
\item  $42$  und  $42$  sind verschieden in  $\text{\LaTeX}$ .
\end{itemize}
```

```
\end{slide}
```

```
\end{document}
```

Folien mit Prosper

Formeln:

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$$

$$e^{i\pi} = -1$$

Beobachtungen:

- $\sqrt{2}$ ist irrational.
- e und π sind transzendent.
- 42 und 42 sind verschieden in \LaTeX .

Einführendes Beispiel:

Ein geschmücktes Dokument:

```
<!DOCTYPE bericht SYSTEM "Beschreibung.dtd">
<bericht>
<titel> Besuch bei Fa. ABC </titel>
<datum> 4. April 2003 </datum>
<person> Claire McDonald </person>
<abschnitt M="1"> <kopf> Besuchsziele </kopf>
<p> Der Besuch bei Fa. ABC dient der Pflege ... </p>
</abschnitt>
</bericht>
```

Deklaration der Schmuckelemente:

```
<!ELEMENT bericht -- (titel, datum, person,
                        abschnitt+)>
<!ELEMENT (titel, datum, person) -- (#PCDATA)>
<!ELEMENT abschnitt -- (kopf, (liste | p)+)>
<!ATTLIST abschnitt M NUMBER #IMPLIED>
<!ELEMENT kopf -- (#PCDATA)>
<!ELEMENT liste -- (kopf?, posten+)>
<!ELEMENT (p, posten) -- ((#PCDATA) |
                        hervorhebung)* >
<!ELEMENT hervorhebung -- (#PCDATA) >
```

Überblick über SGML:

- (i) SGML ist eine Metasprache zur Erstellung von Auszeichnungssprachen. Vordefinitionen sorgen dafür, daß SGML direkt als Auszeichnungssprache für Dokumente genutzt werden kann.
- (ii) Jedes SGML-Dokument ist typisiert. Ein Typ wird in einer Datentypdefinition festgelegt.
- (iii) Jede einzelne Datentypdefinition ist hierarchisch strukturiert.
- (iv) Einem Dokument können gleichzeitig mehrere Datentypdefinitionen zugeordnet sein.
- (v) SGML enthält eine Makrofazilität.

Beispiel für die Notwendigkeit mehrerer unabhängiger Auszeichnungen:

Gedichtband

Gedicht

Verfasser

Titel

Strophenfolge

Strophe

Zeile

Für die Paginierung eines Gedichtbandes benötigt man eine andere Auszeichnung, denn ein Gedicht kann mehrere Seiten umspannen oder eine Seite kann mehrere Gedichte aufnehmen.

Auszug aus dem Umfeld von XML:

Extensible Markup Language (XML) 1.0 (Third Edition)
W3C Recommendation 4th February 2004,
François Yergeau, Tim Bray, Jean Paoli,
C. M. Sperberg-McQueen, Eve Maler

XML 1.1, W3C Recommendation, 4th February 2004,
François Yergeau, John Cowan, Tim Bray, Jean Paoli,
C. M. Sperberg-McQueen, Eve Maler. (edited in place
15 April 2004)

Namespaces in XML W3C Recommendation 14 January
1999, Tim Bray, Dave Hollander, Andrew Layman

Namespaces in XML 1.1 W3C Recommendation, 4th
February 2004, Andrew Layman, Richard Tobin, Tim
Bray, Dave Hollander

xml:id W3C Recommendation, 9 September 2005. Jonathan
Marsh, Daniel Veillard, Norman Walsh.

XML Base W3C Recommendation, 27 June 2001. Jonathan
Marsh.

...

XSLT, XPath, Xpointer, XLink, XInclude,
XQuery, XML-Schema,
XML Signature, XML-Encryption.

Geschichtlicher Überblick:

- 1969: GML (Generalized Markup Language) wird von Charles Goldfarb, Edward Mosher und Raymond Lorie definiert und installiert.
- 1978: Ein Unterkomitee von ANSI übernimmt die Aufgabe, eine Textbeschreibungssprache zu definieren. Charles Goldfarb ist ein Mitglied dieses Komitees.
- 1980: Ein erster Vorschlag für SGML wird publiziert.
- 1985: In Großbritannien wird eine internationale SGML-Benutzergruppe gegründet.
- 1986: Ein ISO-Standard für SGML wird verabschiedet: "ISO 8879:1986 Information processing --- Text and office systems --- Standard Generalized Markup Language (SGML)."
- 1990: Tim Berners-Lee entwickelt parallel zu einem Browser für den NeXT-Rechner die Grundzüge der Hypertext Markup Language.
- 1993: HTML 1.0 wird im Browser Mosaic implementiert.
- 1996: Eine erste Fassung von XML (Extensible Markup Language) wird auf der Konferenz SGML'96 vorgestellt.
- 1997: Die Spezifikation von HTML 4.0 wird abgeschlossen.
- 1998: XML 1.0 wird publiziert.
Version 1 der Synchronized Multimedia Integration Language (SMIL 1) wird publiziert.

Mißverständnisse über SGML:

**Quelle: P. Flynn: Understanding SGML and XML
Tools, Kluwer Academic Publishers,
Boston (1998)**

SGML is a word processor or DTP system from ...

XML is a Microsoft product.

SGML can be downloaded from the Internet.

SGML is too difficult.

SGML is a programming language.

SGML is inflexible.

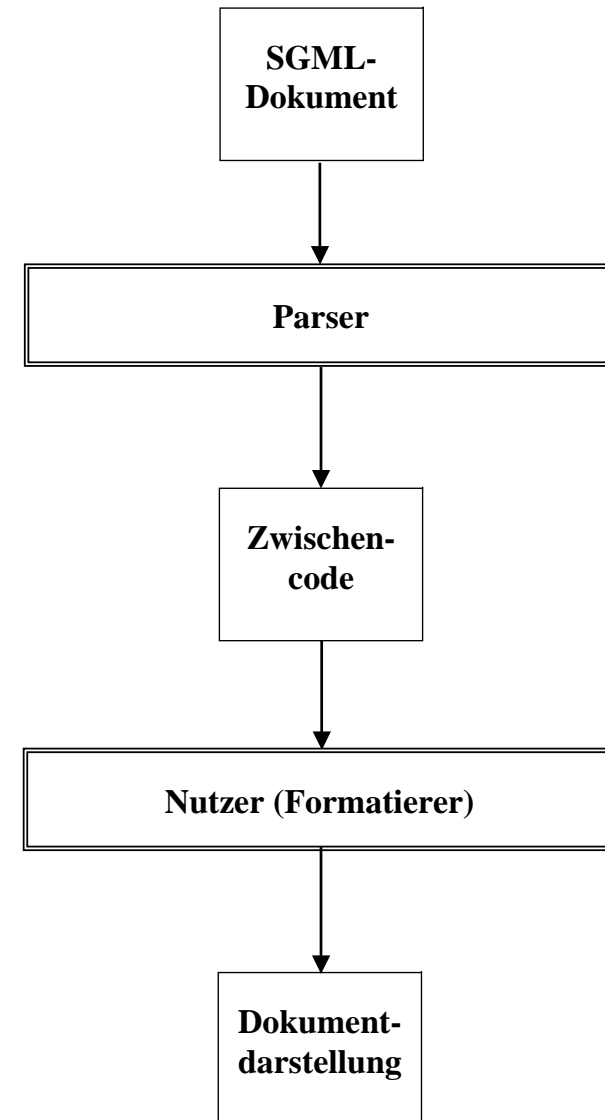
SGML is too expensive.

There aren't no WYSIWYG Interfaces to SGML.

SGML (or XML) predefines the tag names.

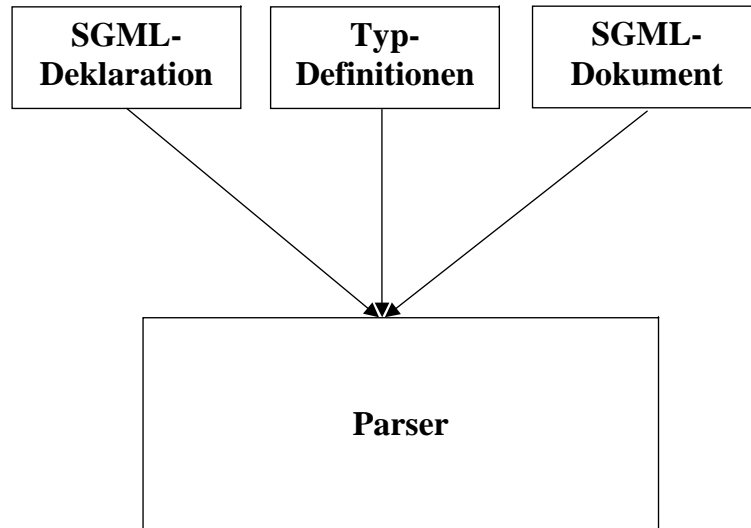
SGML is a US military word processor.

Arbeitsweise eines SGML-Programms:

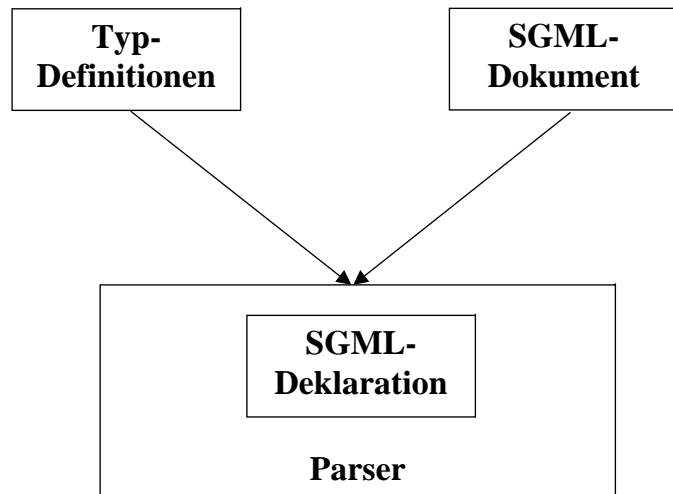


SGML-Parser:

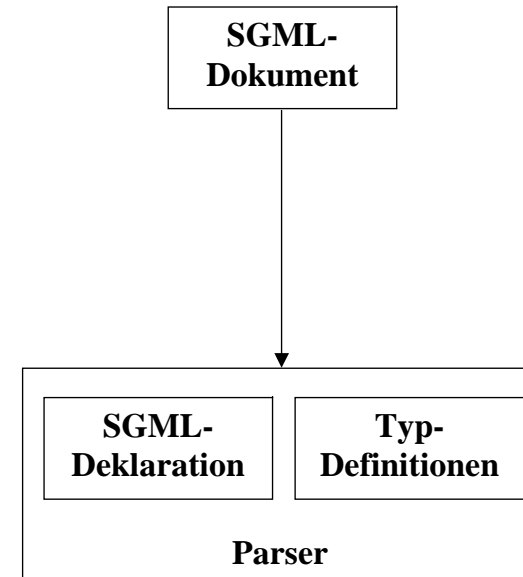
a) Allgemeiner Parser:



b) Spezialisierter Parser:



c) Spezialisierter Parser:



SGML-Deklaration:

Eine SGML-Deklaration legt die konkrete Syntax für eine SGML-Anwendung fest.

Beispiel für eine SGML-Deklaration:

```
<!SGML "ISO 8879:1986"  
CHARSET  
  BASESET "ISO 646:1983//CHARSET  
          International Reference Version  
          (IRV)//ESC 2/5 4/0"  
  
  DESCSET 0    9    UNUSED  
          9    2    9  
          11   2    UNUSED  
          13   1    13  
          14  18    UNUSED  
          32  95    32  
          127  1    UNUSED  
  
CAPACITY    PUBLIC  
  "ISO 8879:1986//CAPACITY Reference//EN"  
  
SCOPE       DOCUMENT  
  
SYNTAX      PUBLIC  
  "ISO 8879:1986//SYNTAX Reference//EN"
```

FEATURES

| | |
|----------|--------------|
| MINIMIZE | DATATAG NO |
| | OMITTAG YES |
| | RANK NO |
| | SHORTTAG YES |

| | |
|------|-------------|
| LINK | SIMPLE NO |
| | IMPLICIT NO |
| | EXPLICIT NO |

| | |
|-------|-----------|
| OTHER | CONCUR NO |
| | SUBDOC NO |
| | FORMAL NO |

APPINFO NONE

>

Abkürzungen:

Baset = Base Character Set

Descset = Described Character Set

Vereinbarung:

`<!Typangabe Parameter und Kommentare>`

Die Typangabe erfolgt durch reservierte Bezeichner oder den Kommentaranfang --.

Kommentare werden durch ein doppeltes Minuszeichen eingeführt und beendet.

Dokumentvereinbarung:

`<!DOCTYPE Bezeichner Suchhinweis - und/oder - Deklarationen >`

Eine Struktur für den Suchhinweis ist
PUBLIC Beschreibung,
wobei Beschreibung sich untergliedert in:

`pm//Besitzer//Klasse Spezifikation//Sprache.`

Bedeutung:

`pm = "+":` Besitzer ist bei der ISO registriert.

`pm = "-":` Besitzer ist nicht registriert.

`pm = leer:` Sonderformen, wie ISO oder ISBN.

Klasse: CAPACITY, CHARSET, DOCUMENT, DTD, ELEMENTS, ENTITIES, LPD, NONSGML, NOTATION, SHORTREF, SUBDOC, SYNTAX, TEXT.

Sprache: Die Angabe der Sprache erfolgt nach ISO 639.

`<!-- Beispiel: Elementvereinbarungen -->`

`<!DOCTYPE dokument`

`[<!element dokument - - any>`

`<!element test - - any>]>`

`<dokument>`

`<test>`

Text A

`</test>`

`<test>Text B`

`</test>`

`<test>Text C</test>`

`<test>Text D`

`</test>`

`</dokument>`

Ausgabe des Parsers nsgmls:

(DOKUMENT

(TEST

-Text A

)TEST

-\n

(TEST

-Text B

)TEST

-\n

(TEST

-Text C

)TEST

-\n

(TEST

-Text D\n

)TEST

)DOKUMENT

```

<!-- Beispiel: Auslassen von Klammern -->
<!DOCTYPE liste
  [<!ELEMENT liste - - (eintrag)*>
   <!ELEMENT eintrag - O CDATA >]>
<!-- Überliest CDATA öffnende Klammern? -->
<liste>
<eintrag>
Ein erster Eintrag.
</eintrag>
<eintrag>
Ein zweiter Eintrag.
<eintrag>
Ein dritter Eintrag.
</liste>

```

Ausgabe des Parsers nsgmls:

```

(LISTE
(EINTRAG
-Ein erster Eintrag.
)EINTRAG
(EINTRAG
-Ein zweiter Eintrag.\n<eintrag>\nEin dritter Eintrag.
)EINTRAG
)LISTE

```

```

<!-- Beispiel: Modellgruppen -->
<!DOCTYPE bericht [
  <!ELEMENT (vorwort|abschnitt|epilog) - O CDATA>
  <!ELEMENT bericht - - (vorwort+, kapitel*,
                        epilog?)>
  <!ELEMENT kapitel O O (abschnitt*)> ]>
<bericht>
<vorwort> Sehr geehrte Leser!</vorwort>
<kapitel>
<abschnitt> Hard- und Software </abschnitt>
<kapitel>
<abschnitt> Hardware </abschnitt>
<abschnitt> Software </abschnitt>
<epilog> Nun das Fazit.
</bericht>

```

Ausgabe des Parsers nsgmls:

```

(BERICHT
(VORWORT
- Sehr geehrte Leser!
)VORWORT
(KAPITEL
(ABSCHNITT
- Hard- und Software
)ABSCHNITT
)KAPITEL
(KAPITEL
(ABSCHNITT
- Hardware
)ABSCHNITT
(ABSCHNITT
- Software
)ABSCHNITT
)KAPITEL
(EPILOG
- Nun das Fazit.
)EPILOG
)BERICHT

```

```

<!-- Beispiel: Ein- und Ausschluß -->
<!DOCTYPE abhandlung
  [<!ELEMENT abhandlung -- (absatz)+>
   <!ELEMENT absatz -- (#PCDATA|fussnote)*>
   <!ELEMENT fussnote -- (absatz)
    -(fussnote) +(absatz)>]>
<abhandlung>
<absatz>
Dies ist ein Absatz mit einer Fußnote.
<fussnote><absatz>
Hier enthält eine Fußnote wieder eine Fußnote.
<fussnote><absatz>???
```

Ausgabe des Parsers nsgmls:

```

(ABHANDLUNG
(ABSATZ
-Dies ist ein Absatz mit einer Fußnote.\n
(FUSSNOTE
(ABSATZ
-Hier enthält eine Fußnote wieder eine
Fußnote.nsgmls:b005.sgm:12:9:E: document type does not allow
element "FUSSNOTE" here
\n
(FUSSNOTE
(ABSATZ
-???)
)ABSATZ
)FUSSNOTE
-\nzzz
)ABSATZ
)FUSSNOTE
-\nxxx
)ABSATZ
)ABHANDLUNG
```

```

<!-- Beispiel: ENTITY -->
<!-- Beispiel für eine Ersetzungskaskade -->
<!DOCTYPE a
  [<!ELEMENT a -- ANY>
   <!ENTITY % b "test &d; test2 &e; ende">
   <!ENTITY c "%b;" >
   <!ENTITY % e " Parameter e ">
   <!ENTITY d "- aus d und e: %e; -">
   <!ENTITY e "## e ##">
  ]>
<a>
&c;
</a>
```

Ausgabe des Parsers nsgmls:

```

(A
-test - aus d und e: Parameter e - test2 ## e ## ende
)A
```