

# *Auszeichnungssprachen (Markup Languages)*

## *SGML*

Besim Kabashi, M.A. / Matthias Bethke, M.A.

Computerlinguistik / Fach "Linguistische Informatik"  
Universität Erlangen-Nürnberg

Proseminar  
**Textkodierung und Textauszeichnung**  
13.6.2006

# Überblick

1 Textauszeichnung: Grundlagen

2 SGML

# Überblick

1 Textauszeichnung: Grundlagen

2 SGML

# Textauszeichnung (Markup)

## Definition

Strukturbeschreibung von Texten; Hervorhebung und formatierung.

## Beispiel

HTML (HyperText Markup Language)

```
<title>&Uuml;bung</title>
```

Vgl. auch die Struktur eines HTML-Dokumentes.

# Textauszeichnung (Markup)

## Definition

Strukturbeschreibung von Texten; Hervorhebung und formatierung.

## Beispiel

HTML (HyperText Markup Language)

```
<title>&Uuml;bung</title>
```

Vgl. auch die Struktur eines HTML-Dokumentes.

# Auszeichnungssprachen (Markup Languages)

## Einige Typen

- Roff (runoff, '60 AT&T/UNIX),
- RTF
- T<sub>E</sub>X, L<sup>A</sup>T<sub>E</sub>X
- SGML, TEI, HTML, XML, XHTML, ...

## Frage

Für welche Zwecke sind sie geeignet?

# Auszeichnungssprachen (Markup Languages)

## Einige Typen

- Roff (runoff, '60 AT&T/UNIX),
- RTF
- T<sub>E</sub>X, L<sup>A</sup>T<sub>E</sub>X
- SGML, TEI, HTML, XML, XHTML, ...

## Frage

Für welche Zwecke sind sie geeignet?

# Auszeichnungssprachen (Markup Languages)

## Einige Typen

- Roff (runoff, '60 AT&T/UNIX),
- RTF
- T<sub>E</sub>X, L<sup>A</sup>T<sub>E</sub>X
- SGML, TEI, HTML, XML, XHTML, ...

## Frage

Für welche Zwecke sind sie geeignet?

# Überblick

1 Textauszeichnung: Grundlagen

2 **SGML**

# SGML

## Standard Generalized Markup Language (1986) ISO-8879

- Vorläufer war GML (Generalized Markup Language), **Charles Goldfarb** (IBM).
- Eine DTD (Document Type Definition) definiert die Struktur eines SGML-Dokuments.

### Beispiel

Vgl. (korpora/BNC-world/Texts/ ...)

# SGML

## Standard Generalized Markup Language (1986) ISO-8879

- Vorläufer war GML (Generalized Markup Language), **Charles Goldfarb** (IBM).
- Eine DTD (Document Type Definition) definiert die Struktur eines SGML-Dokuments.

## Beispiel

Vgl. (korpora/BNC-world/Texts/ ...)

# SGML

## Eigenschaften

- Eine XML/SGML-Anwendung gibt an, was für Typen von abstrakten und konkreten Informationseinheiten es gibt, gibt ihnen Namen zur eindeutigen Identifizierung und spezifiziert ggf. weitere Beschreibungsmerkmale.
- Diese Informationstypen werden durch Regeln miteinander in Beziehung gesetzt.
- Diese Regeln werden zu einer Grammatik zusammengefasst.
- Reale Informationseinheiten werden mit diesen Typen in Beziehung gesetzt und in Baumform angeordnet.

Aus [Lobin 2000: 4].

# SGML-Anwendungen

- TEI (Text Encoding Initiative)
- HTML (HyperText Markup Language)
- XML (Extensible Markup Language)
- XHTML (Extensible HyperText Markup Language)
  
- Vgl. auch <http://www.w3.org/>

# DTD (1)

## Document Type Definitions

“A DTD is essentially a context-free grammar, with its own notation for describing the variables and productions.”

(HOPCROFT ET AL. 2003 : 199)

# DTD (2)

## Document Type Definitions

- DTD-Form

```
<!DOCTYPE name-of-DTD  
  [ element definition(s) ]>
```

- Element

```
<!ELEMENT element-name  
  ( description of the element )>
```

- Special term (Variable input data)

```
#PCDATA, #CDATA
```

# Einheiten der Informationsmodellierung

## Ebenen

1. Die Ebene der konkreten Daten
2. Die Ebene der abstrakten Einheiten, die Daten Funktionen zuordnen oder gruppieren.

Aus [Lobin 2000: 9ff.].

## Abstrakte Daten (Elemente)

1. Daten-Elemente (konkrete Daten / Typ von Daten bzw. Zeichenfolgen)
2. Container-Elemente (Elemente dieses Typs enthalten selbst wiederum Elemente, wobei die enthaltenen Elemente Daten- oder Container-Elemente sein können)

Nach [Lobin 2000: 9ff.].

# Einheiten der Informationsmodellierung: Beispiele

## Konkrete Daten

```
<!ELEMENT titel (#PCDATA)>
```

## Abstrakte Daten

```
<!ELEMENT buch (autor, titel, ort, verlag, jahr)>
```

```
<!ELEMENT buch  
  ((autor+ & titel), ort?, verlag?, jahr?)>
```

```
<!ELEMENT katalog (buch)*>
```

# Leere und gemischte Elementtypen

## Leeres Element: EMPTY

```
<!ELEMENT unbekannt EMPTY>
```

```
<!ELEMENT autor ((vorname+, familienname)  
| unbekannt)>
```

## Gemischte Elemente

```
<!ELEMENT autor ((vorname+, familienname)  
| #PCDATA>
```

# Attribute

Obligatorische, vorbelegte, fakultative und un spezifizierte Attribute

## Eigenschaften von Elementen

```
<!ATTLIST buch  
  cover (soft|hard) #REQUIRED>
```

```
<!ATTLIST buch  
  cover (soft|hard) ``soft``>
```

Weitere Schlüsselwörter von Attribute:

```
#FIXED, #IMPLIED, #CURRENT, ``unspecified``
```

# DSSSL und HyTime

- Document Style and Semantic Specification Language;  
ISO 10179: 1996
- Hypermedia and Time-based Structuring Language;  
ISO 10744: 1997

# Literatur

- Charles F. GOLDFARB: The SGML Handbook. Oxford : Clarendon Press, 1990.
- Eliotte Rusty HAROLD und W. Scott MEANS: The XML CD Bookshelf, 2. Auflage. Cambridge : O'Reilly, 2002.
- John E. HOPCROFT, Rajeev MOTWANI & Jeffrey D. ULLMAN: Introduction to Automata Theory, Languages, and Computation. 2nd. Edition. Upper Saddle River N.J. : Pearson IE 2003. (= HOPCROFT ET AL. 2003)
- Henning LOBIN: Informationsmodellierung in XML und SGML. Berlin : Springer, 2001.