

# Vorlesung Datenbanksysteme II

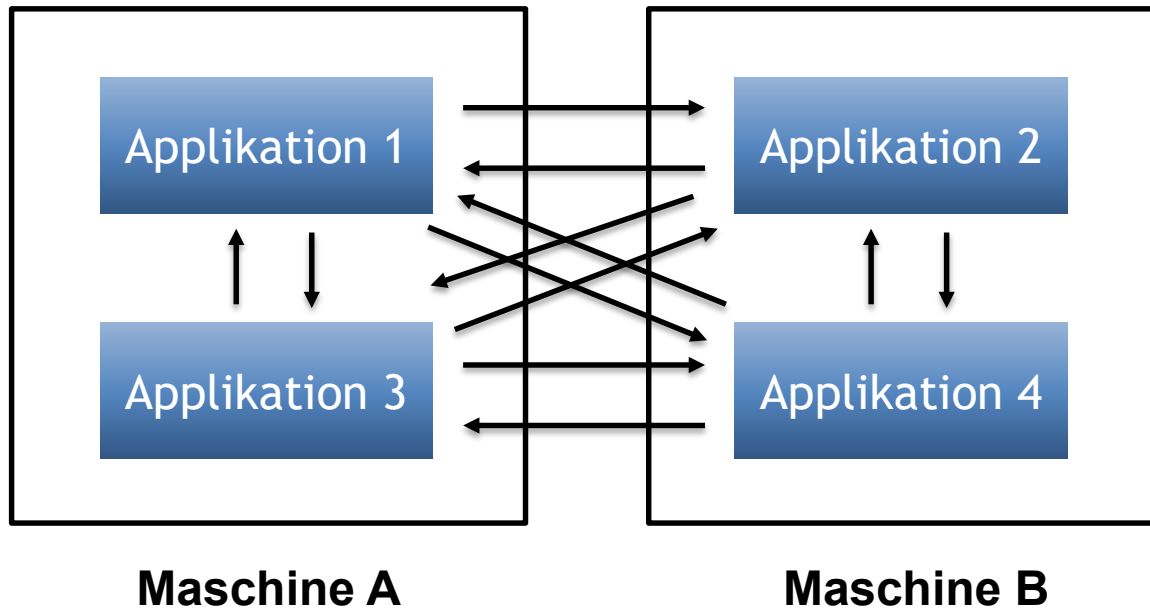
## Transportdateien und die SGML

# Inhalt

- Grundlagen von Transportdateien
- Einführung in die SGML
- Visualisierung strukturierter Daten

# Grundlagen von Transportdateien

# Motivation: Datenaustausch zwischen Applikationen



## Herausforderungen beim Datenaustausch

- Applikationen laufen auf verschiedenen (physikalischen/virtuellen) **Maschinen** mit unterschiedlichen Prozessoren, Speichern...
- Applikationen stammen von verschiedenen Herstellern, die häufig eigene (proprietäre) **Daten- und Dateiformate** verwenden.
- Applikationen sind in verschiedenen Sprachen implementiert, die auf unterschiedlichen **Konzepten zur Datenmodellierung** basieren.

## Technische Aspekte des Datenaustauschs

- Daten-Darstellung auf Ebene der Rechnerarchitektur (32 vs. 64 bit, ...)
- Daten-Codierung auf Datei-Ebene (lesbare vs. binäre Formate, ...)
- Daten-Transport zwischen Maschinen (Speichermedien, Übertragungsprotokolle, ...)
- ...

 hier nicht weiter Thema

## Internationale Standards für Datenformate

- EDIF (Electronic Design Interchange Format)
- STEP/EXPRESS (Data Modeling Language, standardized as ISO 10303-11)
- UN/EDIFACT (United Nations Electronic Data Interchange for Administration, Commerce and Transport)
- ...  
➔ Bisher kein allgemein akzeptierter Standard

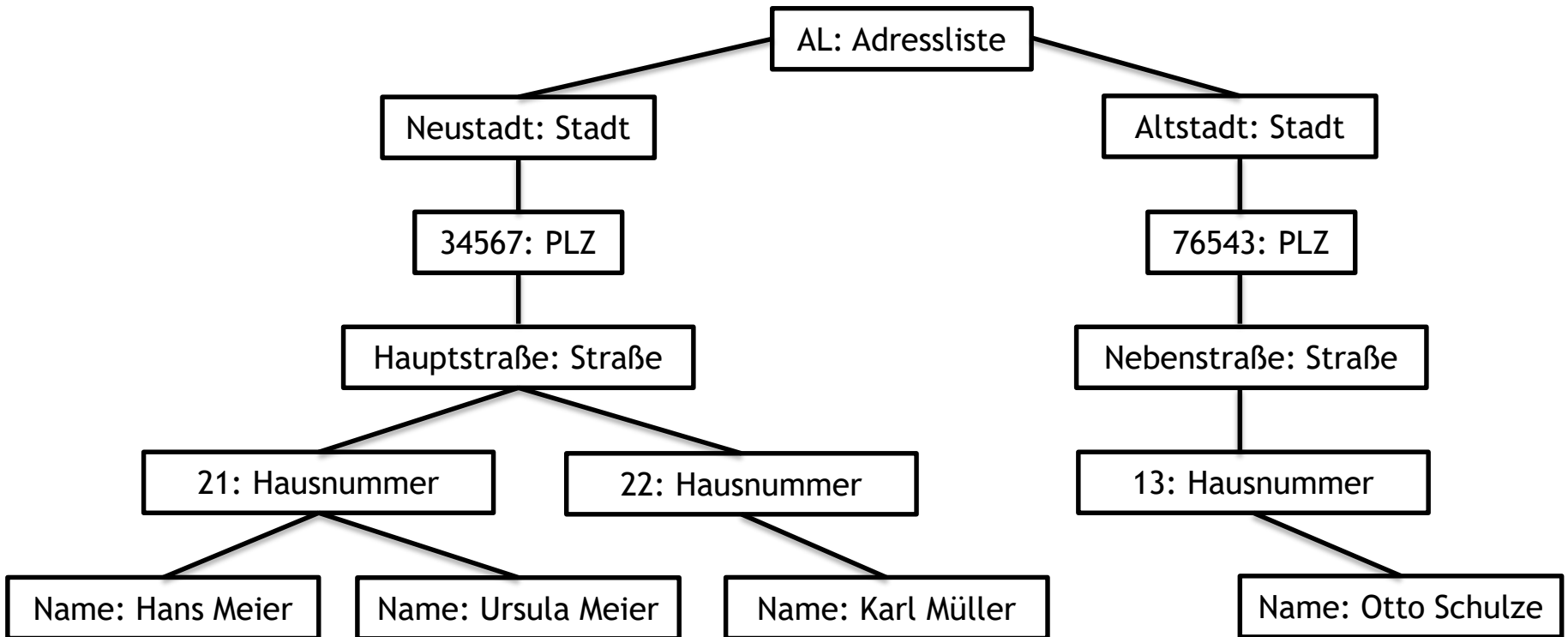
# Konzepte zur Datenmodellierung

Name	Straße	Hausnr.	PLZ	Stadt
Hans Meier	Hauptstr.	21	34567	Neustadt
Ursula Meier	Hauptstr.	21	34567	Neustadt
Karl Müller	Nebenstr.	13	76543	Altstadt
Otto Schulze	Hauptstr.	22	34567	Neustadt
...	...	...	...	...

Beispiel: Adressliste im Datenmodell „Tabelle“

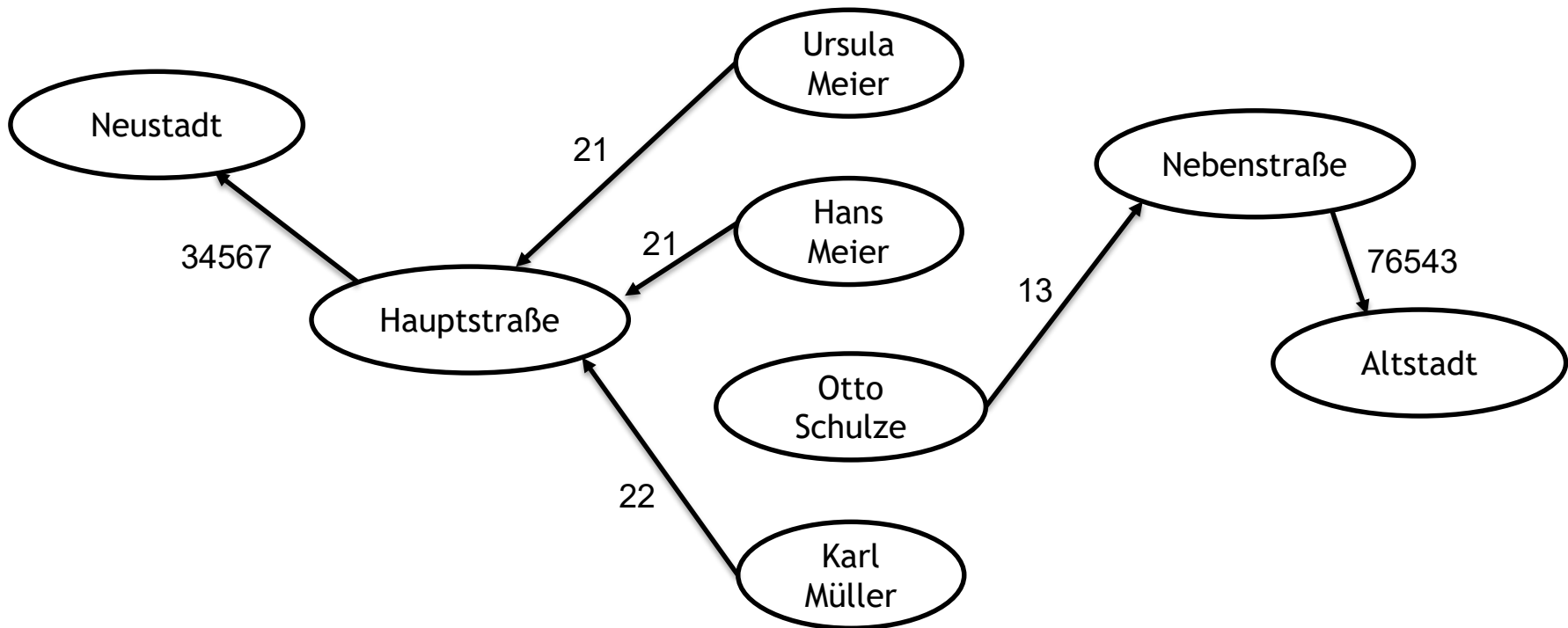


# Konzepte zur Datenmodellierung



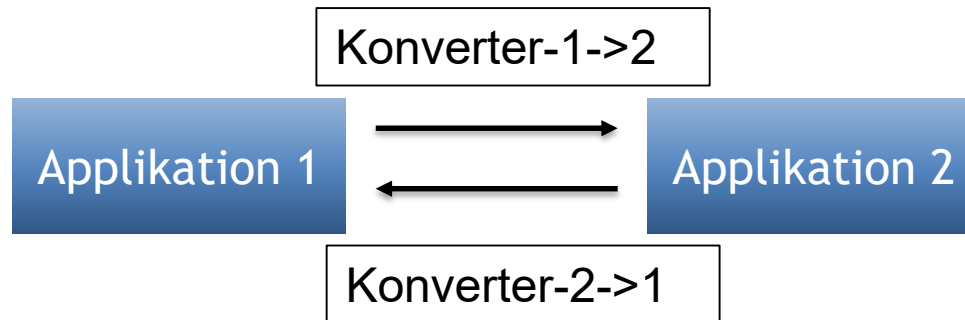
Beispiel: Adressliste im Datenmodell „Baum“

# Konzepte zur Datenmodellierung



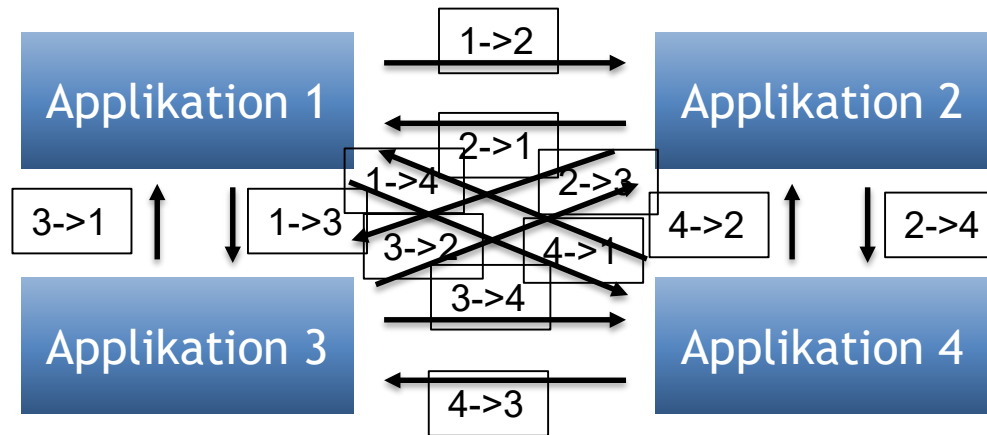
Beispiel: Adressliste im Datenmodell „Graph“

## Ansatz: Formatkonversion



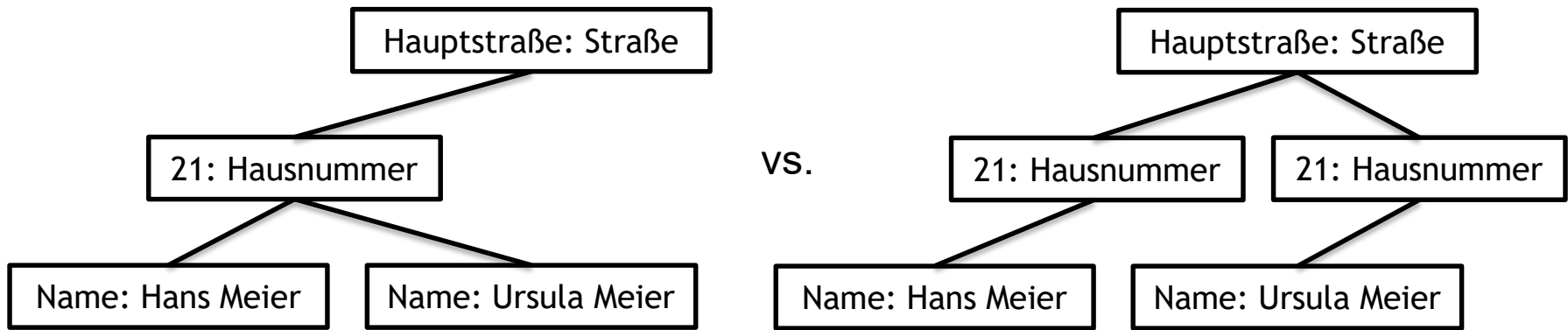
- Implementierung kompatibler Datenformat-Konversionsroutinen pro Applikations-Paar und verwendeten Dokumenttypen

## Diskussion: Formatkonversion



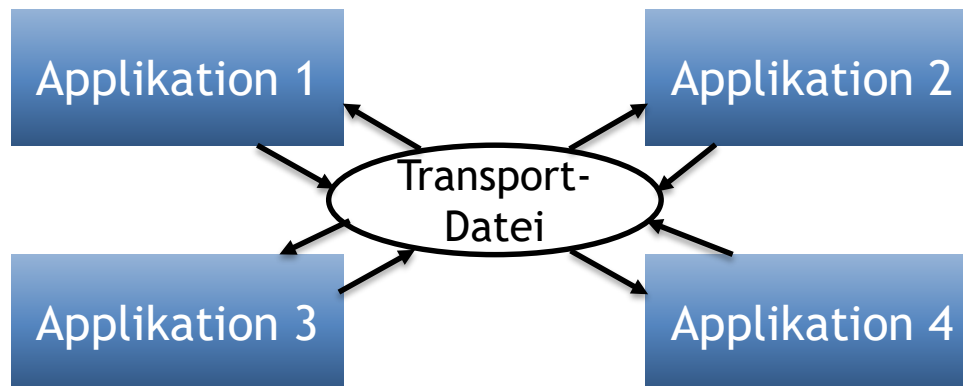
- Gegebenenfalls sehr viele und sehr spezifische Konverter notwendig
- Gegebenenfalls Informationsverlust/Redundanz bei der Konversion

## Diskussion: Formatkonversion



*Übungsaufgabe: Ist dieser semantische Unterschied in den anderen Datenmodellen (direkt) darstellbar?*

## Lösungsansatz: Transportdateien

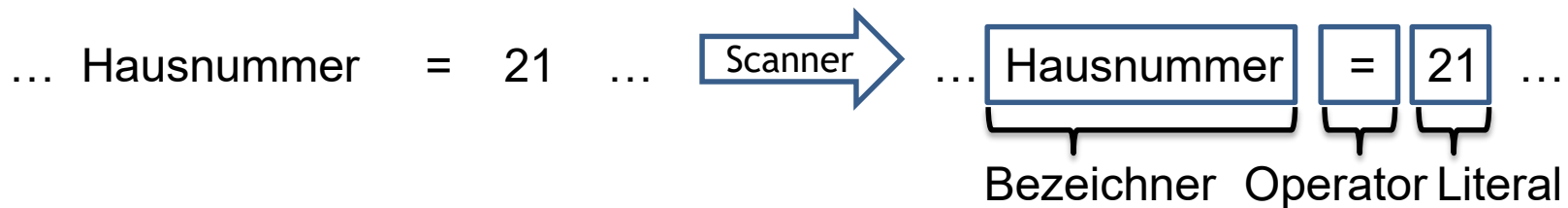


- Ziel: Sternförmige Konversion zwischen **allen** Applikationen mittels **einem einheitlichen** Format für Transportdateien

## Diskussion: Transportdateien

- Werden sehr viele verschiedene Dokumenttypen und Datenmodellen verwendet, ist der Aufwand nach wie vor sehr hoch.
- In der Praxis besteht der Konverter zumeist vor allem aus Scannern und Parsern.

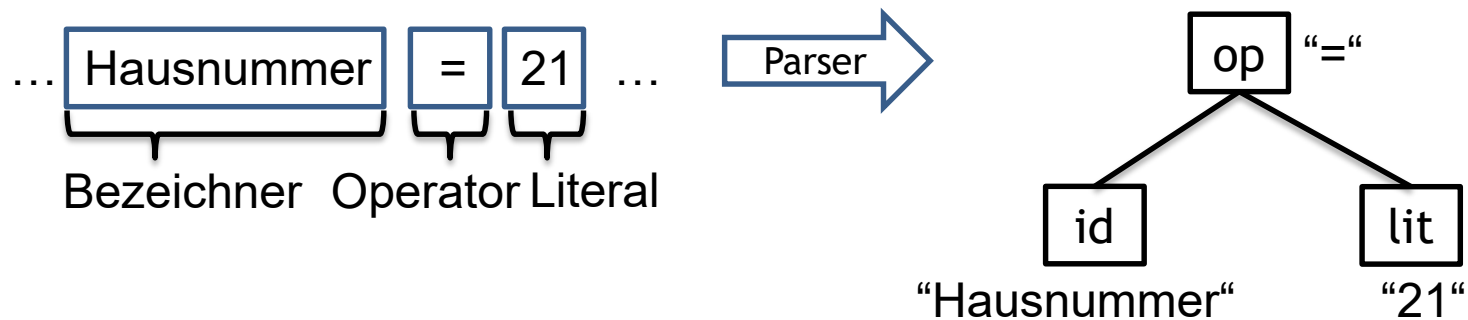
## Exkurs: Scanner



- Der Scanner führt eine **lexikalische Analyse** des Eingabetextes gemäß der Zeichentypen der Eingabesprache durch.
- Der Scanner überführt dazu den Eingabetext in **Folgen von Token** (Zeichen)



## Exkurs: Parser

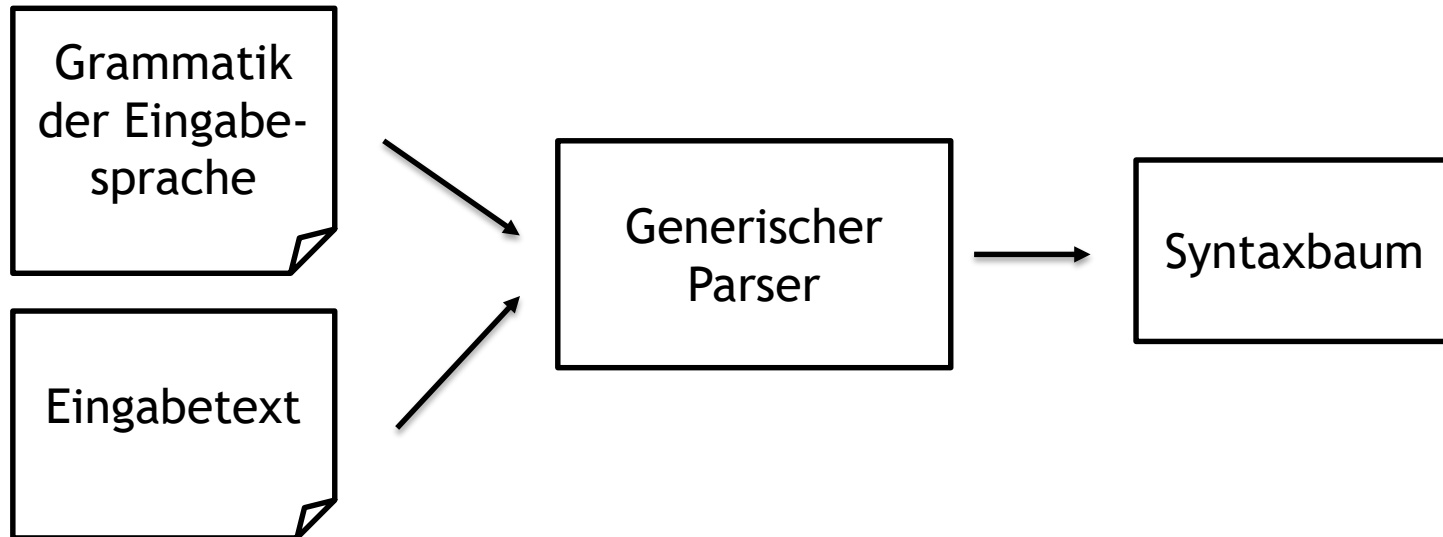


- Der Parser führte eine syntaktische Analyse des Eingabetextes gemäß der **Grammatik** der Eingabesprache durch.
- Der Parser überführt dazu die Folge von Token in einen (attribuierten) **Syntaxbaum**.

## Diskussion: Scanner und Parser

- Scanner wird zumeist als Bestandteil des Parsers angesehen und nicht extra erwähnt.
- Grammatik häufig im Quelltext des Parsers „fest verdrahtet“.
- Für Formatkonversionen aufwändige Konstruktion vieler (oft ähnlicher) Scanner/Parser notwendig.

# Generische Parser

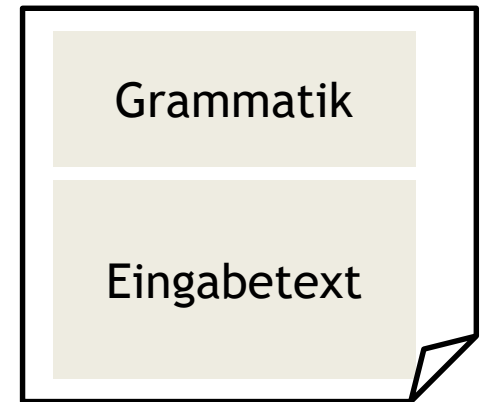


- Generischer Parser erhält neben dem Eingabetext als Parameter eine Beschreibung der Grammatik des Eingabetextes.

## Generische Parser

Übergabe der Grammatik an den generischen Parser:

- Über „Konstante“ im Programm
- Separat über lokale Ressourcen-Datei oder globale Referenz (z.B. URL)
- Als integraler Bestandteil der Transportdatei



Transportdatei

Notation von Grammatiken in **Meta-Sprachen**.

# Sprachebenen

<b>Sprachebene</b>	<b>Der Eingabetext repräsentiert...</b>	<b>Die Eingabesprache ist bestimmt durch...</b>
2	Grammatik der Meta-Sprache	Meta-Meta-Sprache
1	Grammatik der Nutzdatensprache	Meta-Sprache
0	Nutzdaten / Fakten der Problemdomäne	Nutzdatensprache

## Ebene 0: Nutzdaten



Meier, Hans, Hauptstr., 21, 34567, Neustadt  
Meier, Ursula, Hauptstr., 21, 34567, Neustadt  
Müller, Karl, Nebenstr., 13, 76543, Altstadt  
Schulze, Otto, Hauptstr., 22, 34567, Neustadt  
...

- Textuelle Darstellung von Realwelt-Entitäten aus einer Problemdomäne.
- Beispiel: eine konkrete Adressliste.

## Ebene 1: Grammatiken

Adressliste	::= Adresse*
Adresse	::= Name ', ' Vorname ', ' Straße ' ' Hausnummer ', ' PLZ ', ' Ort
Name	::= Zeichen*
Straße	::= Zeichen*
Hausnummer	::= Ziffer*
PLZ	::= Ziffer*
Ort	::= Zeichen*
Ziffer	::= '0' '1' '2' '3' '4' '5' '6' '7' '8' '9'

- (Textuelle) Darstellung (eines Teils) der Syntax der Eingabesprache von Ebene 0.
- Beispiel: Grammatik einer Adressliste (in BNF).

## Ebene 2: Meta-Grammatiken

Grammatik	::= Produktion*
Produktion	::= Nonterminal '::=' Ausdruck
Ausdruck	::= '(' Ausdruck ')'   Ausdruck '**'   Ausdruck '+'   Ausdruck '?'   Sequenz   Alternative   Terminal
Sequenz	::= Ausdruck ' ' Ausdruck
Alternative	::= Ausdruck ' ' Ausdruck
....	

- (Textuelle) Darstellung (eines Teils) der Syntax der Sprache, die auf Ebene 1 verwendet wird, um Grammatiken zu beschreiben.
- Beispiel: Grammatik der BNF (in BNF).



## Diskussion: Meta-Sprachen

- Beobachtung: die grundlegenden syntaktischen Strukturen von Meta-Sprachen und der durch diese beschriebenen konkreten Sprachen können gleich sein!
- Trotzdem ist die Bedeutung der Sprachkonstrukte hinsichtlich der damit verfassten Dokumente semantisch verschieden.

## Beispiele für Meta-Sprachen

- Backus-Naur-Form (BNF)
- Teile der Standard Generalized Markup Language (SGML)
- Teile der eXtensible Markup Language (XML)

# Transportdateien vs. Datenbanken

## Gemeinsamkeiten:

- Speicherung von strukturierten Daten auf Basis eines konzeptuellen Modells (z.B. ER oder OOAM).
- Unterstützung von Sichten (durch Einlese-Filter bei Transportdateien).
- Lesbarkeit ist zweitrangig.

# Transportdateien vs. Datenbanken

## Unterschiede:

- Transportdateien dienen dem Austausch kleiner Datenmengen.
- Inhalte von Transportdateien werden zumeist vollständig und rein sequentiell durch eine Applikation verarbeitet.

# Einführung in die SGML

## Historische Notizen

### Standard Generalized Markup Language (SGML)

- Standard für den Austausch textueller Dokumente in Büroinformations- und ähnlichen Systemen.
- Prinzip der „schwach“ strukturierte Daten
- Seit Anfang der 80er Jahre durch die ISO standardisiert.
- Sehr komplexe Meta-Sprache (500 Seiten Spezifikation), nicht weit verbreitet.

## Historische Notizen

### eXtensible Markup Language (XML)

- vereinfachte Variante der SGML (nur noch 50 Seiten Spezifikation).
- Erste Version 1998 vom W3C-Konsortium vorgestellt.
- Aktuell ist die fünfte Ausgabe von 2008 in Gebrauch.
- Integration von Nutzdaten und Strukturmerkmalen in einem Dokument.

# Strukturmerkmale in textuellen Dokumenten

## Submission Template for ACM Papers\*

Insert Your Subtitle Here

First A. Author†  
First Department Name  
First Institution/University Name,  
City, State, Country, email@email.com

Second B. Author  
Second Department Name  
Second Institution/University Name,  
City, State, Country, email@email.com

### ABSTRACT

This submission template allows authors to submit their papers for review to an ACM Conference or Journal without any output design specifications incorporated at this point in the process. The ACM "Submission Template" is a single column MS-Word document that allows authors to type their content into the pre-existing set of paragraph formatting styles applied to the sample placeholder text here, or copy-and-paste their text and then apply the respective paragraph styles (Windows: you can open the Styles task pane from the Home tab [it can also be opened with the keyboard shortcut Alt+Ctrl+Shift+S], MAC OS: you can access the Styles pane at the right of the Home toolbar.) Highlight a section that you want to designate with a certain style, and then select the appropriate style from the list. To view which style is being used in any part of this document, place your cursor on your text and look at the "Current style" field in the Styles pane.

It is beneficial to create your document in draft mode with the style panel open in the left-side panel. If the panel is not immediately visible when the Submission Template is opened, you will need to open the panel manually—for Windows: click on the following from the main ribbon above: File > Options > Advanced > Display > Style area pane width in Draft and Outline views. Set the style area width (1-1.5" is a good starting value), for MAC: go to the "View" menu and select "Draft"; then go to the "Word" menu and select "Preferences" and then "View," under the "Window" section insert "1.5" inches under the style area width.

For your reference, below is a list of the paragraph styles to be applied to the respective elements of your text:

STYLE TAG	DEFINITION	STYLE TAG	DEFINITION
This document	main title of article	ListParagraph	list item
SubTitle	subtitle of article	Section	math sections
Author	author name	Equation	block equations
Abstract	author affiliation information	AlgorithmCaption	caption for algorithm
ListItem	list item in outline	Equation	equation
Section	heading for section	Text	heading for subequation
Text	abstract text	Text	acknowledgments text
CCSItem	heading for CCS Concepts	Equation	equation of page
CCSDescription	CCS term	Equation	number for the part
KeyWordList	heading for keywords	Section	heading for references
Keywords	keywords text	Text	references
Footnote	footnote to title	EquationList	equation heading level 1
FootnoteList	author's (IEEE) D.F.	EquationList	equation heading level 1
FootnoteList	heading level 1	EquationList	equation heading level 1
FootnoteList	heading level 2	TableCaption	title of table
FootnoteList	heading level 3	Table	column head of table
FootnoteList	heading level 4	TableCaption	footnote to table
FootnoteList	footnote paragraph after heading	FigureCaption	figure caption text
ParaContinue	subsequent paragraphs of figure text	Image	figure
DisplayEquation	numbered math equation	Equation	digital object identifier
EquationFormalEquation	unnumbered equation	Text	label

If a paper is accepted for publication, authors will be instructed on the next steps: to submit a revised document file after applying the Word version of the ACM Master Article Template to their paper. Building on the work authors have already completed using this "Submission Template," the ACM Master Article Template will contain macros, functionalities, and validations that need to be applied before a paper can be submitted for publication. Authors must then follow the submission instructions found on their respective publication's web page. Once your submission is received, your paper will be processed to produce the formatted Word, PDF, and HTML5 output formats, which will be provided to you for review, revision/resubmission (if applicable), and approval.

CCS CONCEPT insertion: The new template enables you to import required indexing concepts for your article from the ACM Computing Classification System (CCS) using an indexing support tool found in the ACM Digital Library (DL). The tool generates formatted text after you have selected your terms. To insert your CCS terms into your document, copy and paste the formatted text from the CCS tool using the "https://dl.acm.org/ccs/ccs.cfm" link into the "CCS CONCEPTS" section.

WWW '19, 2019-04-23, 2019, Lyon

First A. Author and Second B. Author

An additional step is necessary to ensure that the proper CCS terms are added to the Digital Library citation page: from the "view CCS TeX Code" listing, click on "Show the XML Only." Highlight and copy the XML code from the window. You must insert the XML code into your Word document's properties: from your Word document, click on "File", then click on the "Info" tab on the left-hand side panel, then click "Properties" and select "Show All Properties." Click within the "Comments" metadata field and paste the XML data.

### CCS CONCEPTS

Insert your first CCS term here - Insert your second CCS term here - Insert your third CCS term here

### KEYWORDS

Insert comma delimited author-supplied keyword list, Keyword number 2, Keyword number 3, Keyword number 4

### ACM Reference format:

First A. Author and Second B. Author. 2019. Submission Template for ACM Papers: Insert Your Subtitle Here. In *Proceedings of WWW '19: The Web Conference 2019* (WWW '19). ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3179876.3186135>

### 1 Introduction: About this Template

All style elements are specified in this template to facilitate the production of your paper and to have the styles consistent throughout. The paragraph styles are built-in and examples of the styles are provided throughout this document. Save as you go and backup your work regularly!

### 1.1 More about the submission template

This submission version of your paper should not have headers or footers. Balasubramanian (2000). It should remain in a one-column format—please do not alter any of the styles or margins.<sup>1</sup>

1.1.1 How to apply Display Math. There are two types of math equations: the **numbered display math equation** and the **un-numbered display math equation**. Below are examples of both.<sup>2</sup>

$$\sqrt{b^2 - 4ac} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad (1)$$

The **DisplayFormula** style (above) is applied in the numbered

math

equa

The

unnum

1

2

3

4

5

6

7

8

contains an equation number Berron and Grimes (2012) on the right—this element distinguishes it from the numbered equation. Please note: the subsequent text after the **DisplayFormula** (numbered equation) or **DisplayFormulaUnnum** (unnumbered equation) must have the paragraph style **ParaContinue** applied.

1.1.2 Figures and Tables. Figure(s) and table(s) are "float elements" which should be inserted after their first text reference, and have specific styles for identification. Insert a figure and apply the "Image" paragraph style to it. For the figure caption, apply the style "FigureCaption." Below is an example of a figure and caption with the styles applied.



Figure 1: Figure caption and image above the caption [in draft mode the image will not appear on the screen.]

Authors can also insert a (Abdelghaffar, Kamel, and Duquenois 2010) multi-part figure above a single caption. Every inserted figure must have the "Image" style applied. Below are instructions regarding how to insert a multi-part figure in your paper.

(i) If the author wants to insert two multi-part images, they must draw a one row and one

## Beispiel: pdf-Darstellung eines Beispieldokuments für das ACM Latex-Template

\* Place the footnote text for the title (if applicable) here.

† Place the footnote text for the author (if applicable) here.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

WWW '19, 2019-04-23, 2019, Lyon  
© 2019 Association for Computing Machinery.  
ACM ISBN 978-1-4503-5030-0...\$15.00  
<https://doi.org/10.1145/3179876.3186135>



# Strukturmerkmale in textuellen Dokumenten

Submission Template for ACM Papers\*

Subtitle Here

City, State, Country, email@email.com

**Titelabschnitt**

**Zweispaltiger Seitenaufbau**

**Fettschrift**

**Absätze**

**Eingebettete Tabellen**

STYLE TAG	DEFINITION	STYLE TAG	DEFINITION
The document	main title of article	ListParagraph	list item
SubTitle	sub-title of article	SectionName	math elements
Author	author name	Equation-Block	block equation
Algorithm	author affiliation information	AlgorithmCaption	caption for algorithm
Footnote	footnote or sidebar	Algorithm	algorithm
SectionName	heading for section	FigureCaption	heading for sub-figure
Text	abstract text	Figure	acknowledgments text
SectionName	heading for CCS concepts	Equation-Block	equation of page
CCSDescription	CCS terms	Equation-Block	number for the page
		Equation-Block	heading for subsection
		Equation-Block	appendix heading level 1
		Equation-Block	appendix heading level 1
		Equation-Block	appendix heading level 1
		Equation-Block	title of table
		Equation-Block	column level of table
		Equation-Block	document title
		Equation-Block	figure caption text
FigureCaption	sub-figure caption of figure text	Image	figure
Equation-Block	numbered math equation	Equation-Block	digital object identifier
Equation-Block	un-numbered equation	Equation-Block	label
		Equation-Block	title

**Kursivschrift**

**Eingebettete URL**

WWW '19, 2019-04-23, 2019, Lyon

First A. Author and Second B. Author

An additional step is necessary to ensure that the proper CCS terms are added to the Digital Library citation page: from the "view CCS TeX Code" listing, click on "Show the XML Only." Highlight and copy the XML code from the window. You must insert the XML code into your Word document's metadata from the "Info" pane of the Word ribbon.

**Unnummerierte Abschnitte**

**Literaturverweise**

**CCS CONCEPTS**

Insert your first CCS term here. Insert your second CCS term here. Insert your third CCS term here.

**KEYWORDS**

Insert comma delimited author-supplied keyword list. Keyword number 1. Keyword number 2. Keyword number 3.

**ACM KEYWORDS**

Insert comma delimited author-supplied keyword list. Keyword number 1. Keyword number 2. Keyword number 3.

**1 Introduction: About this Template**

All style elements are specified in this template to facilitate the production of your paper and to have the styles consistent throughout. The paragraph styles are built-in and examples of the styles are provided throughout this document. Save as you go and backup your work regularly!

**1.1 More about the submission template**

This submission template is designed to help authors create headers and footers that are consistent with the ACM Master Article Template. The following sections describe the format and the style elements of the template.

**1.1.1 Figure Caption and Image**

Figure 1: Figure caption and image above the caption [in draft mode the image will not appear on the screen.]

**Eingebettete Formeln**

$$\sqrt{b^2 - 4ac} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad (1)$$

The **DisplayFormula** style (above) is applied in the numbered math equation. A numbered display equation always has an equation number (label) on the right.

$$\sqrt{b^2 - 4ac} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

The **DisplayFormulaUnnum** style (above) is applied only in unnumbered equations. An unnumbered display equation never has an equation number (label) on the right.

**Fußnoten**

**Eingebettete Bilder**

**Nummerierte Aufzählungen**

- (i) If the author wants to insert two multi-part images, they must draw a one-row and one column table and insert the images one-by-one in the cells.
- (ii) If the author wants to insert three multi-part images, they must draw a one-row and three-column table and insert the images one by one in all three cells.
- (iii) If the author wants to insert four multi-part images, they must draw a two-row and two-column table and insert the images one-by-one in all four cells. (see the following example).

# Markups

- Vordefinierte Notation für Textauszeichnungen
- Ursprünglich für den Schriftsetzer handschriftlich auf Papier
- Text mit und ohne Markups lesbar

```
...
\subsection{Template Styles}
The primary parameter given to the
``\verb|acmart|'' document class is the
{\itshape template style} which corresponds
to the kind of publication or SIG
publishing the work. This parameter is
enclosed in square brackets and is a part
of the {\verb|documentclass|} command.
\begin{itemize}
\item acmsmall: The default journal
template style.
\item acmlarge: Used by JOCCH and TAP.
\item acmtog: Used by TOG.
\end{itemize}
...
```

**Beispiel: Textauszeichnungen in Latex**

# Bedeutung von Markups

## 1. Layout (Formatierungsanweisungen):

Beispiel: Ein Textstück soll fett/kursiv/... angezeigt werden, Verarbeitung z.B. durch Renderer

## 2. Inhaltliche Struktur:

Beispiel: Ein Textstück soll Referenz/Bild-unterschrift/... sein, Verarbeitung z.B. durch Verzeichnis-Generatoren

## Diskussion: Markups in textuellen Dokumenten

- Markups sind meistens nicht eindeutig den Kategorien 1. und 2. zuzuordnen.
- Markups der 2. Kategorie liefern bei Weitem keine verlässlichen/vollständigen Informationen zur inhaltlichen Strukturierung von Dokumenten, sondern werden zumeist „willkürlich“ eingesetzt.

## SGML-Entitäten

SGML-Entitäten unterstützen die lexikalische Analyse der Inhalte einer Transportdatei.

## Deklaration von Entitäten

```
<!ENTITY entityname "ersetzungzeichenkette">  
<!ENTITY entityname SYSTEM "Dateiname">
```

- Auftreten von “*entityname*“ werden ersetzt durch “*ersetzungzeichenkette*“ bzw. dem Inhalt von “*Dateiname*“.
- Handhabung in Dokumenttypdefinitionen anders als in Nutzdaten.

## Einsatzgebiete von Entitäten

- Ersatzdarstellungen für nicht direkt darstellbare (Sonder-) Zeichen
- Makros für wiederholt auftretende Textteile
- Aufteilung eines Gesamttextes auf verschiedene Dateien
- Verschiedene Ausgabevarianten für einen einzigen Quelltext

# Tags

SGML-Tags repräsentieren benannte  
Klammerungen der Form:

`<tagname attributliste>`

...

`</tagname>`

öffnendes *tag*

Inhalt des tags

schließendes *tag*

`<tagname attributliste />`

*tag* ohne Inhalt



## Wohlgeformtheit

- Tags können wie üblich in Eingabetexten geschachtelt werden.
- Ein Eingabetext ist **wohlgeformt**, wenn alle öffnenden und schließenden tags paarweise „korrekt geklammert“ sind.

## Unterschiede zwischen SGML, HTML und XML

- Schließende *tags* können in der SGML und HTML oft fehlen - in der XML nicht.
- Tags sind in der HTML fest vordefiniert - in SGML und XML frei definierbar.
- Tags betreffen in der HTML vor allem Layout - in SGML und XML müssen dafür separate Visualisierungsvorschriften definiert werden.

## Elemente

SGML-Elemente sind Paare zusammengehöriger tags und der dazwischenstehende Inhalt der Form:

`<tagname attributliste>`

`... Inhalt...`

`</tagname>`

öffnendes *tag eines Elements*

Inhalt eines Elements

schließendes *tag eines Elements*

Ein Element entspricht somit einem kompletten Unterbaum des Syntaxbaums eines Dokuments

# Elementtypen

SGML-Elementtypen definieren für Elemente

- deren Menge von Attributen (im öffnenden tag) und
- die zulässigen inneren Elemente sowie ggf. deren Anzahl und Anordnung.

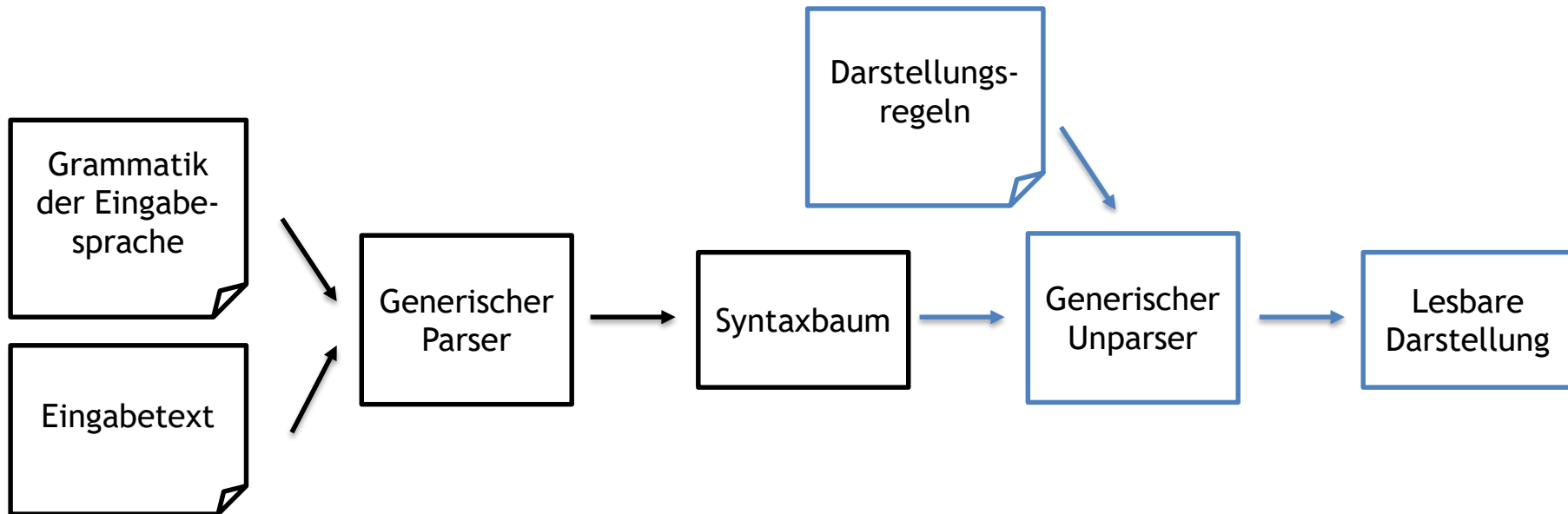
# Dokumenttypdefinitionen

SGML-Dokumenttypdefinition (DTD) umfassen

- eine Menge von Elementtyp-Definitionen und
- den Startelementtyp

# Visualisierung strukturierter Daten

# Layout-Spezifikationen



- Ausführung Server- oder Client-seitig möglich
- Beispiele: XSL (siehe später) und CSS

## Prüfungsstoff

- Motivation von Transportdateien erklären
- Gemeinsamkeiten und Unterschiede von Transportdateien und Datenbanken diskutieren
- Grundprinzipien von SGML kennen