

Präsenzveranstaltung zur E-Learning-Veranstaltung

Einführung in XML

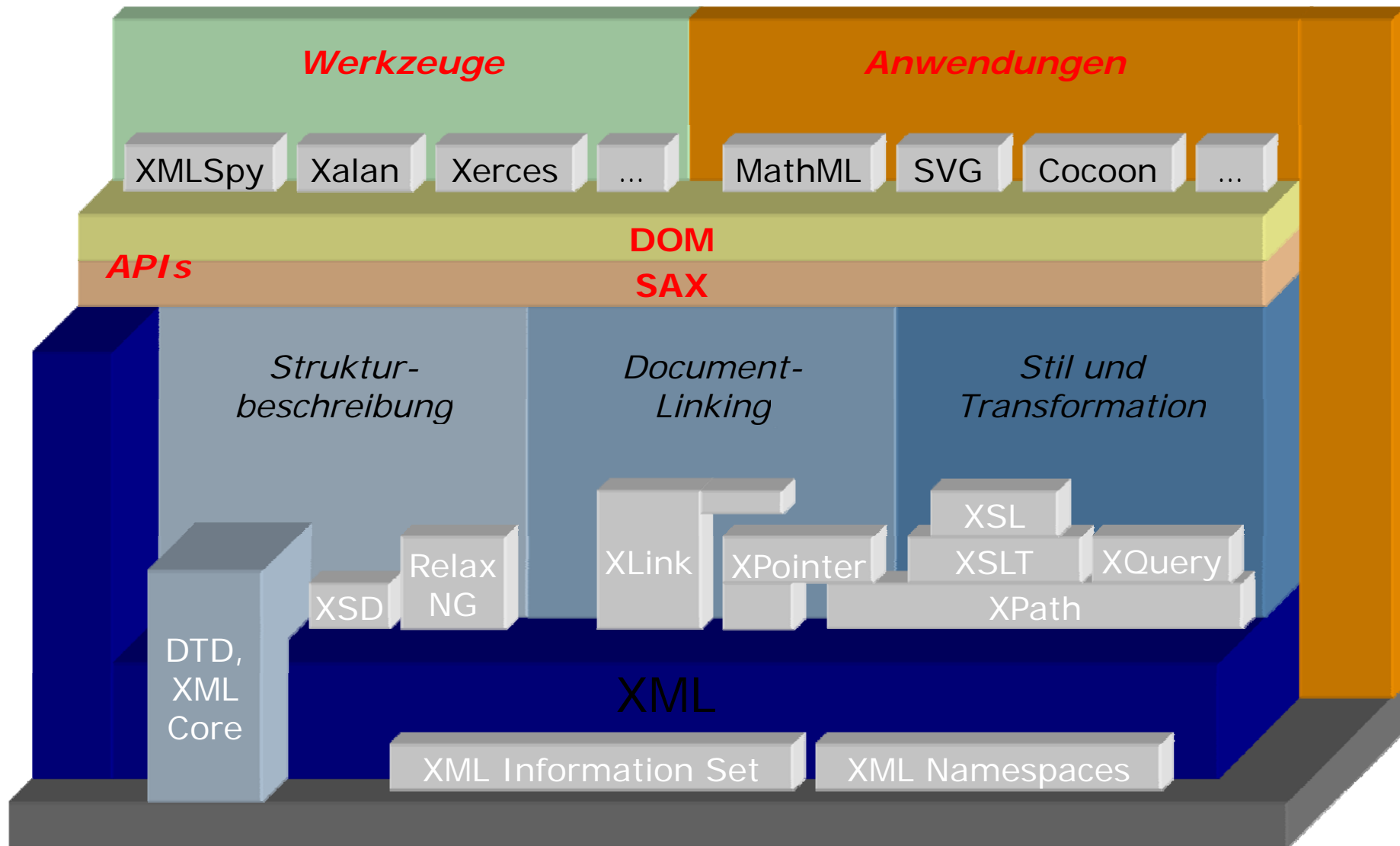
Sommersemester 2009

Prof. Dr. Klaus-Peter Fährnich
Martin Gebauer

Agenda

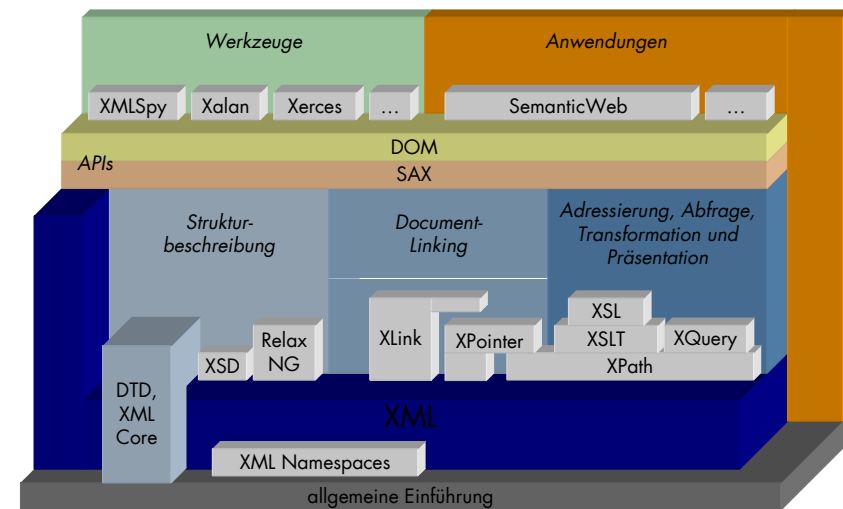
- **Zusammenfassung des WB Programmierschnittstellen und Werkzeuge**
- Diskussion
- Weiterer Ablauf

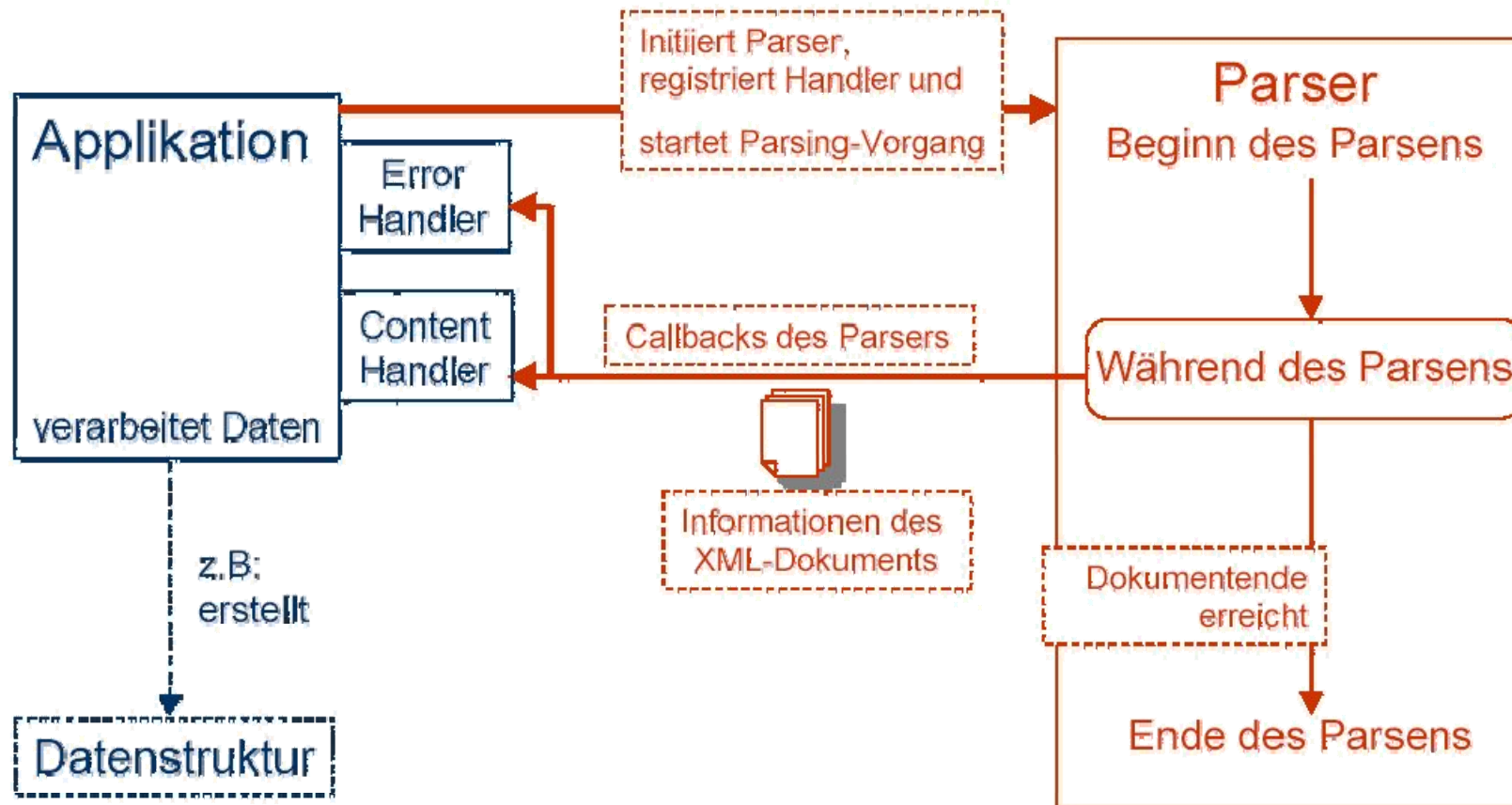
Gliederung der Vorlesung



6. Programmierschnittstellen und Werkzeuge

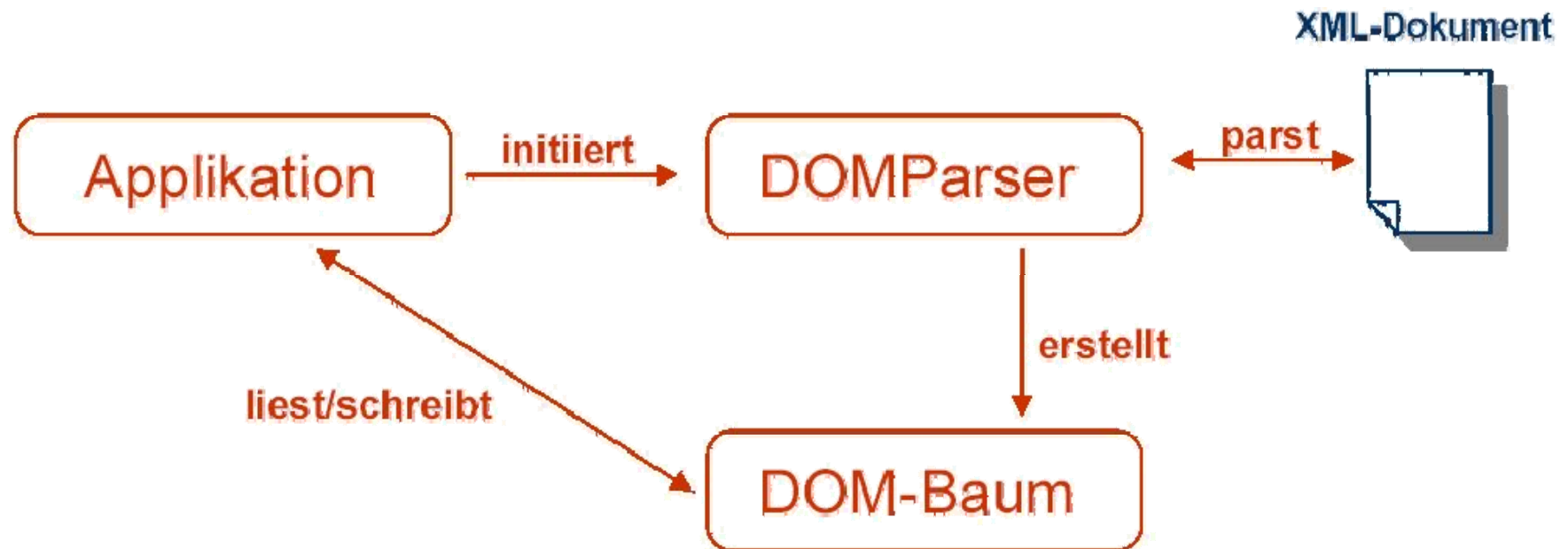
- Programmierschnittstellen
 - SAX
 - DOM
- Werkzeuge zur Bearbeitung von XML-Dokumenten
 - Strukturbeschreibung
 - Dokumentenerstellung
 - Dokumentenbetrachtung
 - Datenhaltung





rot: SAX; blau: Anwendung

Quelle: TU-Dresden, Informatik

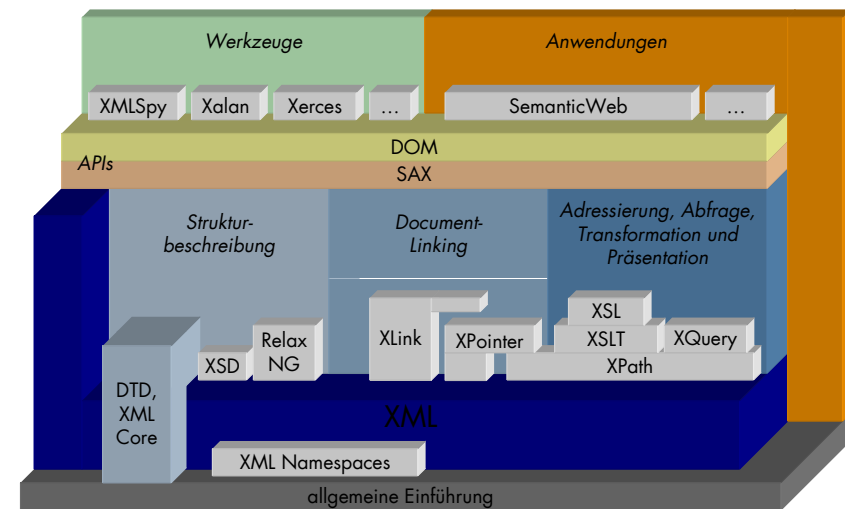


Quelle: TU-Dresden, Informatik

SAX vs. DOM

	SAX	DOM
Pro	<ul style="list-style-type: none">• zeit- und speichereffizient• Parsen jederzeit abbrechbar• gezielte Extraktion von Informationen	<ul style="list-style-type: none">• einfacher• gezielter Zugriff auf Elemente• einfache Suche• flexiblere Verarbeitung
Contra	<ul style="list-style-type: none">• keine Nachbearbeitungsmöglichkeit• schwierige Navigation• wenig Funktionalität	<ul style="list-style-type: none">• Zeit- und Speicherverbrauch• Probleme bei sehr großen Dokumenten

- Programmierschnittstellen
 - SAX
 - DOM
- Werkzeuge zur Bearbeitung von XML-Dokumenten
 - Strukturbeschreibung
 - Dokumentenerstellung
 - Dokumentenbetrachtung
 - Datenhaltung



Strukturbeschreibung, Dokumentenerstellung ...

- Strukturbeschreibung – DTD bzw. Schemaeditoren
 - ASCII Editor vs. DTD-/ Schemaeditor
- Dokumenterstellung – Instanzeditoren
 - Texteditor
 - Texteditor mit eigenen Erweiterungen
 - Texteditor mit XML spezifischen Erweiterungen vom Hersteller
 - XML Editor
- Dokumentenbetrachtung – Viewer, Browser
 - Struktursicht
 - Direkte Anzeige
 - Serverseitiges Publishing

Datenhaltung

- Manuelle Verwaltung
 - + gewohnte Bedienung, Kosten, Erweiterung mit Betriebssystem
 - Keine XML-Unterstützung, aufwändig (große Datenmengen)
- Herkömmliche Datenbank
 - Speicherung als flache Zeichenkette
 - CLOB oder Textdatei
 - Originalgetreu gespeichert, Performanz für ganze Dokumente, Suchen langsam da keine DB Funktionen anwendbar
 - Modellbasierte Verfahren
 - Analyse Abbildung auf Tabellen
 - Anfragen greifen auf Strukturinformationen zu
 - Schnell bei Teilen , langsam bei gesamten Dokumenten

Datenhaltung

- Modellbasierte Schemaabhängige Verfahren
 - Basis Relationenschema – Dokumente müssen diesem Schema entsprechen
 - Datenanfrage ohne aufwändige Navigation durch XML Struktur
 - Änderungen mit Relationenschemaänderung verbunden
 - Modellbasierte Schemaunabhängige Verfahren
 - Relationenschema für beliebige XML Dokumente
 - Flexibel, hoher Aufwand bei Navigation (Baumstruktur)
- + im Einsatz, kein Schulungsaufwand, indiv. Anpassungen, XML Erweiterungen
- Keine Native XML Unterstützung, Erweiterung teuer, Versionsverwaltung indiv. Geregelt
- Herkömmliches Dokumenten-Managementsystem

Datenhaltung

- XML-Datenbank:
 - Vorteile:
 - Bieten spezielle XML-Unterstützung:
 - Parsing beim Check-In,
 - Check-Out von Dokumententeilen,
 - Herunterbrechen der Speicher-/Sperrgranulate auf ein definiertes Elementniveau,
 - Suche nach Elementinhalten/Attributwerten mit XQuery/XQL
 - Unterstützt die Wiederverwendung von Dokumentteilen,
 - Unterstützt die Parallelbearbeitung von Dokumenten
 - Nachteile:
 - Noch relativ neue Produkte
 - Beispiele:
 - Tamino (Software AG),
 - eXcelon (eXcelon Corp.),
 - X-Hive/DB (X-Hive Corp.)
 - Xindice (Apache XML Project)

Fragen ?