

Synchronisation auf XML-Dokumenten

Michael P. Haustein

haustein@informatik.uni-kl.de

Datenbankarbeitsgruppentreffen
Rathen, 29./30. September 2003

Gliederung

- Document Object Model
 - DOM-Baum und -API zum Zugriff auf XML-Dokumente
- taDOM-Baum
 - Erweiterung des DOM-Baums zur feingranularen Synchronisation
- Sperrkonzept
 - Knotensperren
 - Navigationssperren
 - Logische Sperren
- Realisierungsaspekte
 - Implementierung einer Client/Server-Testumgebung
 - Auswertungsergebnisse
- Zusammenfassung und Ausblick

Document Object Model (1)

- Document Object Model
 - Verwaltung des DOM-Standards durch W3C seit Oktober 1998
 - Darstellung von XML-Dokumenten in Baumstruktur
 - Parsen und Aufbau des gesamten Dokuments beim ersten Zugriff im Hauptspeicher
 - Navigation, einfache Suche und Modifikation
 - Normalerweise: "Datei-basierter Einbenutzerbetrieb"

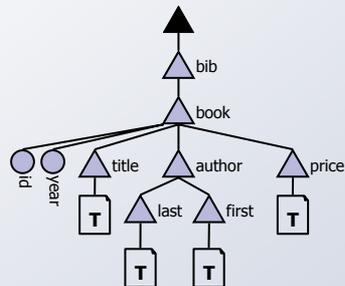


Document Object Model (2)

■ XML-Fragment

```
<bib>
  <book year="1994" id="1">
    <title>TCP/IP Illustrated</title>
    <author>
      <last>Stevens</last>
      <first>W.</first>
    </author>
    <price>65.95</price>
  </book>
</bib>
```

■ Darstellung als DOM-Baum



■ DOM-API

- | | | |
|----------------------|--------------------|----------------------------|
| • Navigation | • Modifikation | • Suche |
| getFirstChild() | appendChild (...) | getElementById (...) |
| getLastChild() | insertBefore (...) | getElementsByTagName (...) |
| getPreviousSibling() | removeChild (...) | hasAttribute (...) |
| getNextSibling() | setNodeValue (...) | |
| getAttributes() | setAttribute (...) | |
| getNodeValue() | | |



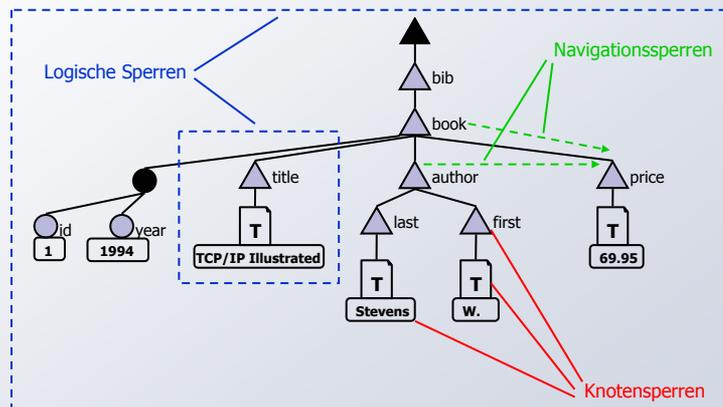
taDOM-Baum (1)

- Motivation und Anforderungen
 - Speicherung von XML-Dokumenten in Datenbanksystemen
 - (O)RDBMS mit XML-Erweiterung vs.
 - Natives XML-Datenbanksystem
 - "Direkter" knotenorientierter Zugriff auf gespeicherte XML-Dokumente mit der DOM-API
 - Client/Server-Umgebung
 - Paralleler Zugriff mehrerer Transaktionen auf dasselbe Dokument erfordert Synchronisation
 - Nur Aufbau benötigter XML-Fragmente im Hauptspeicher des Clients (*deferred DOM*)



taDOM-Baum (2)

- Erweiterung des DOM-Baums zum taDOM-Baum
 - neue Knotentypen *Attributwurzel* und *String*
 - feingranulare Synchronisation



Knotensperren (1)

- Anforderung einer ausreichenden Knotensperren vor dem Zugriff auf den Knoten

Sperre	Wirkung
NR (node read)	Lesesperre auf dem Knoten
LR (level read)	Lesesperren auf allen Knoten der darunterliegenden Ebene
SR (subtree read)	Lesesperre auf gesamtem Teilbaum

Lesesperren

	-	IX	NR	CX	LR	SR	U	X
IX	+	+	+	+	+	-	-	-
NR	+	+	+	+	+	+	-	-
CX	+	+	+	+	-	-	-	-
LR	+	+	+	-	+	+	-	-
SR	+	-	+	-	+	+	-	-
U	+	+	+	+	+	+	-	-
X	+	-	-	-	-	-	-	-

Kompatibilitätsmatrix

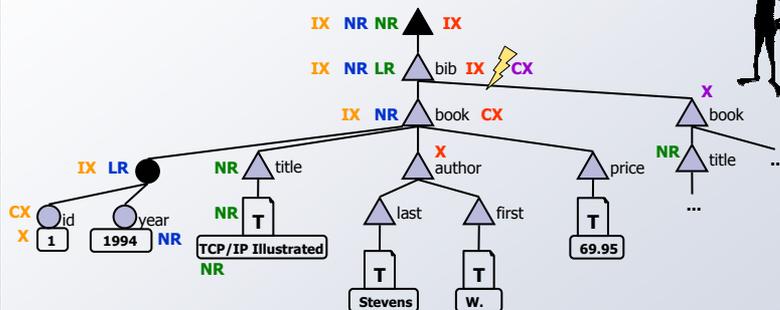
Sperre	Wirkung
X (exclusive)	Schreibsperre auf gesamtem Teilbaum
CX (child excl.)	Schreibsperre auf direktem Kindknoten
IX (intent. excl.)	Schreibsperre auf nicht-direktem Kindknoten
U (update)	Lesesperre mit beabsichtigtem Ändern auf gesamtem Teilbaum

Schreibsperren



Knotensperren (2)

- Beispiel für Sperranforderungen



- Vorteil: Sehr feingranulare Synchronisation
Hohe Parallelität in der Verarbeitung
- Nachteil: Große Anzahl zu verwaltender Sperren



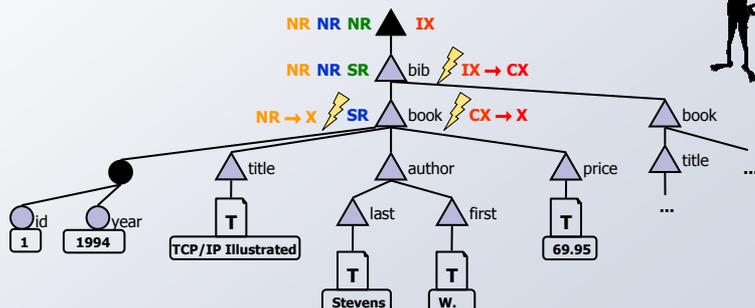
Knotensperren (4)

- Reduzieren der Anzahl gehaltener Sperren
 - Einstellbare Sperrgranularität
 - *Sperrtiefe*: Anzahl der Knotenebenen, auf denen Sperren angefordert werden (einschließlich der Dokumentwurzel)
 - Bei Überschreitung:
nur noch SR- bzw. X-Sperre auf Sperrtiefe anfordern
 - Sperreskalation
 - *Eskalationsgrenzwert* und *Eskalationstiefe*
 - Ausgehend von den taDOM-Blättern durchsuchen von Teilbäumen der Höhe *Eskalationstiefe*
 - Berechnung des prozentualen Anteils von Sperren auf den vorhandenen Knoten
 - Bei Überschreiten des Eskalationsgrenzwertes:
Ersetzen der Sperren im Teilbaum durch SR- bzw. X-Sperren auf Teilbaumwurzel



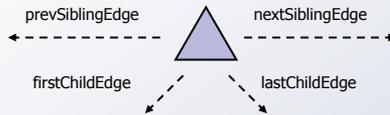
Knotensperren (5)

- Beispiel mit eingestellter Sperrtiefe 3
 - Anforderung von Sperren nur bis Ebene 3
 - Auf Ebene 3 nur Verwaltung von SR- bzw. X-Sperren
 - Knoten auf darunter liegenden Ebenen durch entsprechende Teilbäume gesperrt



Navigationssperren (1)

- Definition von virtuellen Navigationskanten für Knoten



- Anforderung einer ausreichenden Sperre
 - beim Navigieren auf der entsprechenden Kante
 - beim Hinzufügen oder Löschen von Knoten auf allen betroffenen Kanten
- R/U/X-Sperrverfahren für die Navigationssperren

	-	ER	EU	EX
ER	+	+	-	-
EU	+	+	-	-
EX	+	-	-	-

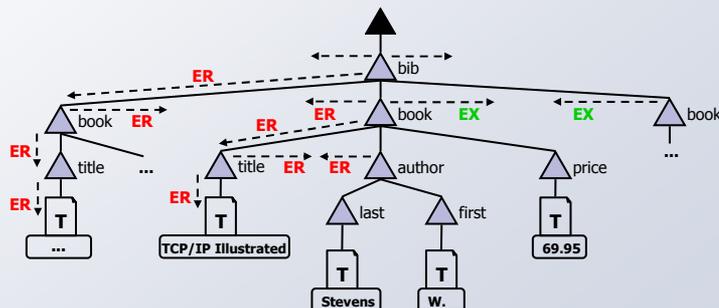
ER: edge read
EU: edge update
EX: edge exclusive



Navigationssperren (2)

- Beispiel: Paralleles Lesen und Einfügen von Büchern

- Transaktion 1**
 - navigiert zum Titel des ersten Buchs
 - navigiert zum Titel des zweiten Buchs
 - navigiert zum Autor des zweiten Buchs
- Transaktion 2**
 - fügt parallel zu Transaktion 1 ein neues Buch ein



Logische Sperren (1)

- Vermeidung von Phantomen
- Synchronisation der DOM-API Methoden
 - getElementByTagName (...)
 - getElementById (...)
 - hasAttribute (...)
- Verwaltung logischer Sperren:

LocksTagnameQuery			
TAID	lock	scopeNID	tagname
...	R/U/X

LocksIDQuery		
TAID	lock	ID
...	R/U/X	...

LocksAttributeQuery			
TAID	lock	NID	attribute
...	R/U/X



Logische Sperren (2)

- Beispiel
 - Transaktion 1
 - Anfrage auf gesamtem Dokument nach Knoten nach ID-Attributwert "4-711-0-815" (DTD erforderlich)
 - Transaktion 2
 - Anfrage auf Teilbaum mit Wurzelknoten (interne ID 815) nach Elementen mit Tagname "last"
 - Anfrage auf Knoten (interne ID 4711) nach Existenz von Attribut "year"



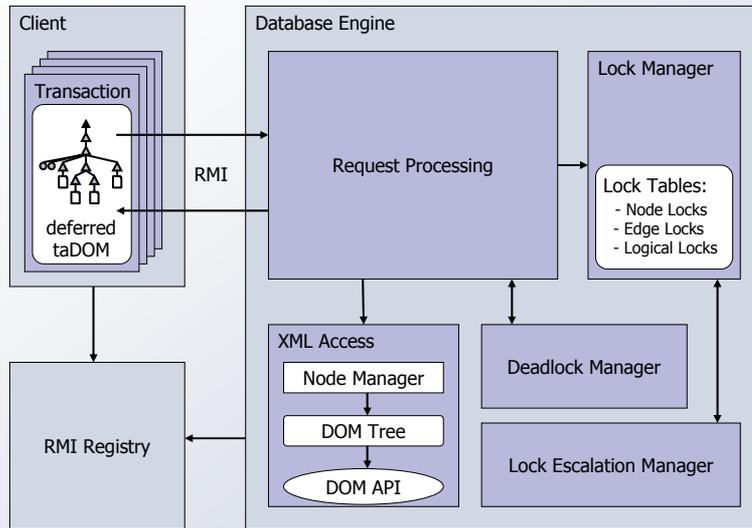
LocksTagnameQuery			
TAID	lock	scopeNID	tagname
2	R	815	last

LocksIDQuery		
TAID	lock	ID
1	R	4-711-0-815

LocksAttributeQuery			
TAID	lock	NID	attribute
2	R	4711	year

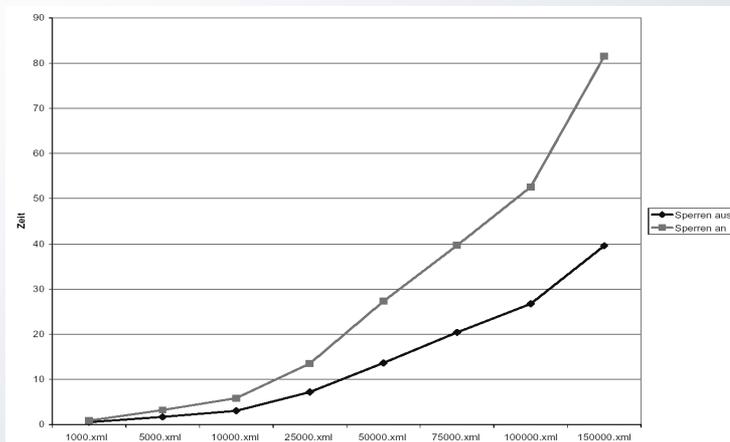


Realisierungsaspekte (1)



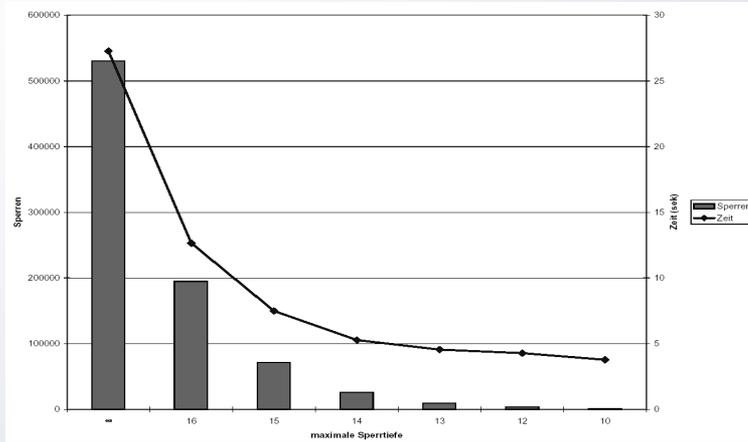
Realisierungsaspekte (2)

■ Aufwand für Sperrverwaltung



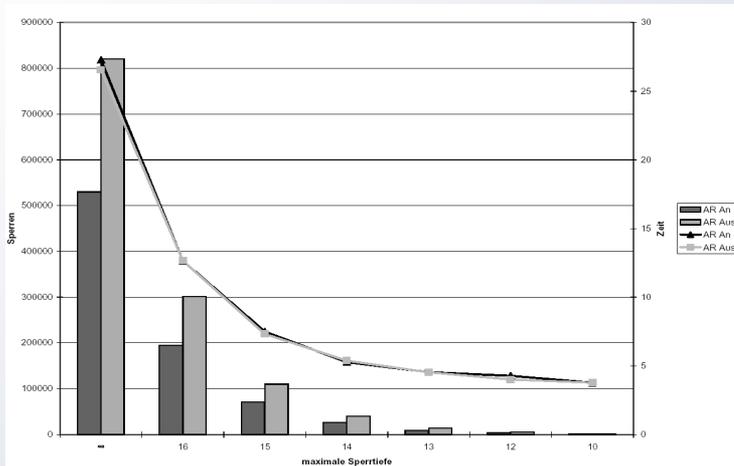
Realisierungsaspekte (3)

Einfluss der Sperrtiefe



Realisierungsaspekte (4)

Aufwand für Verwaltung einer Attributwurzel



Zusammenfassung

- Document Object Model
 - Baumdarstellung von XML-Dokumenten
 - Methoden zur Navigation, Modifikation und Suche
- Erweiterung des DOM-Baums zum taDOM-Baum
 - Feingranulare Synchronisation
 - Neue Knotentypen *Attributwurzel* und *String*
- Sperrkonzept
 - Knotensperren (variable Sperrgranularität und –eskalation)
 - Navigationssperren
 - Logische Sperren
- Realisierungsaspekte
 - Client/Server-Testumgebung
 - Auswertungsergebnisse



Ausblick

- Weitere Messungen
 - Ermittlung typischer Lastsituationen beim Zugriff auf zentral verwaltete XML-Dokumente
 - Bestimmung eines Parallelitätsgewinns bei der Verarbeitung aus Anwendungssicht
- Native Speicherkomponente für XML-Dokumente
 - Ersetzung der XML-Access-Komponente
 - Evaluierung des Konzepts in nativem XML-Datenbanksystem



Vielen Dank. Fragen?

