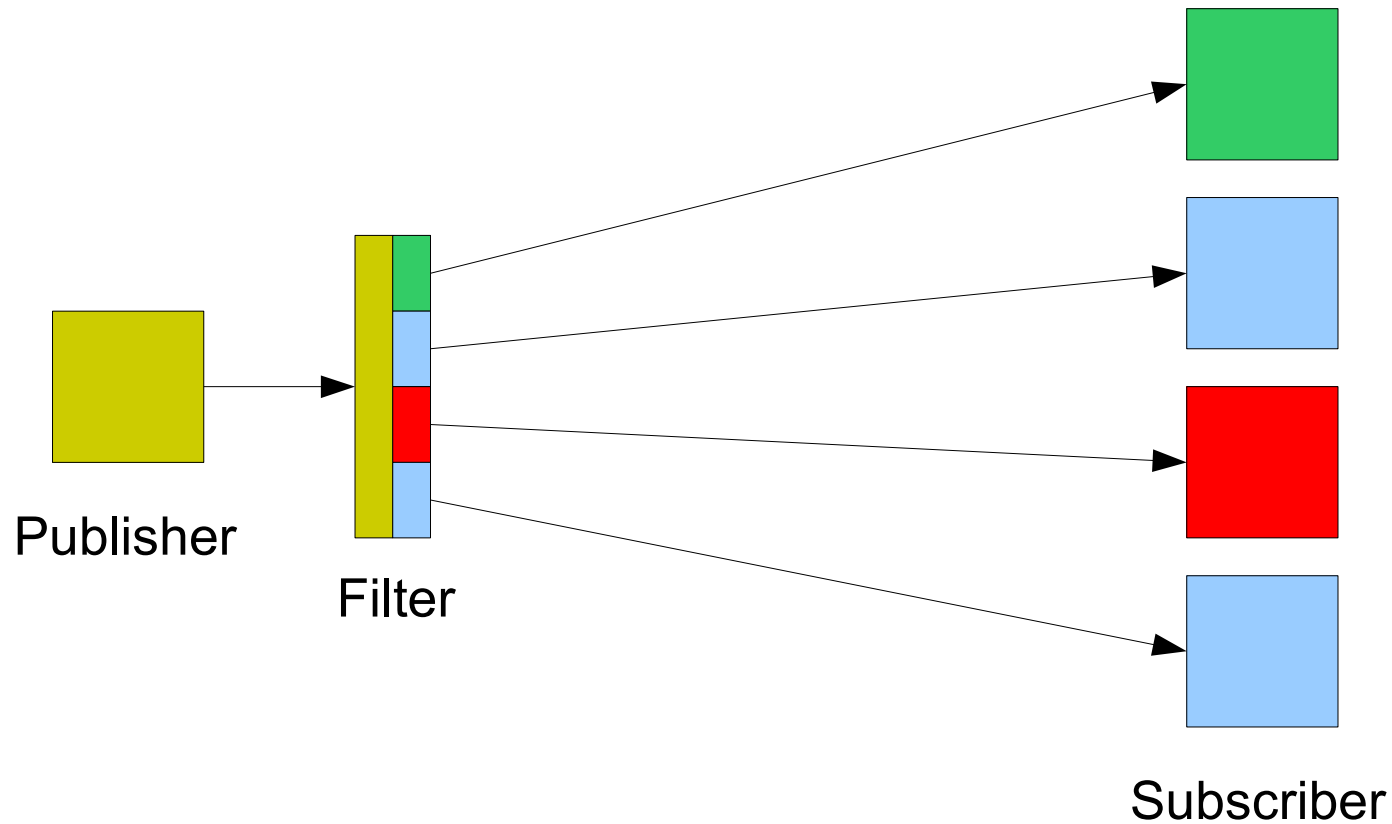


Fast XPath Matching: XTrie

- Einführung
- XPE-Bäume
- XTrie-Index
- XTrie-Matching
- Fazit

Publish/Subscribe-Dienste

→ asynchrone Nachrichtendienste:



Das „XPath-Expression (XPE) Retrieval Problem“

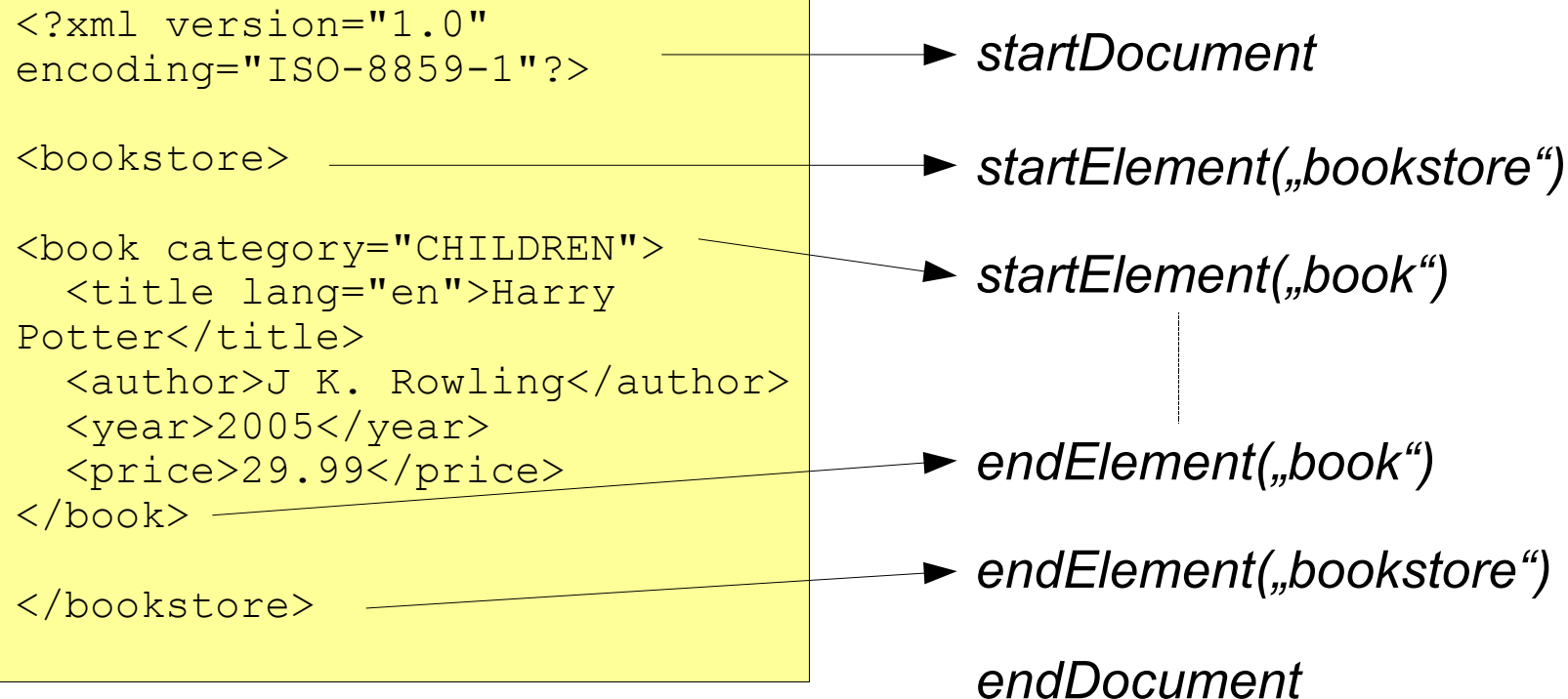
„Aus einer großen Menge P von XPath-Expressions (XPEs) und einem XML-Dokument D , finde die Teilmenge von XPEs aus P die in D eine Übereinstimmung haben.“

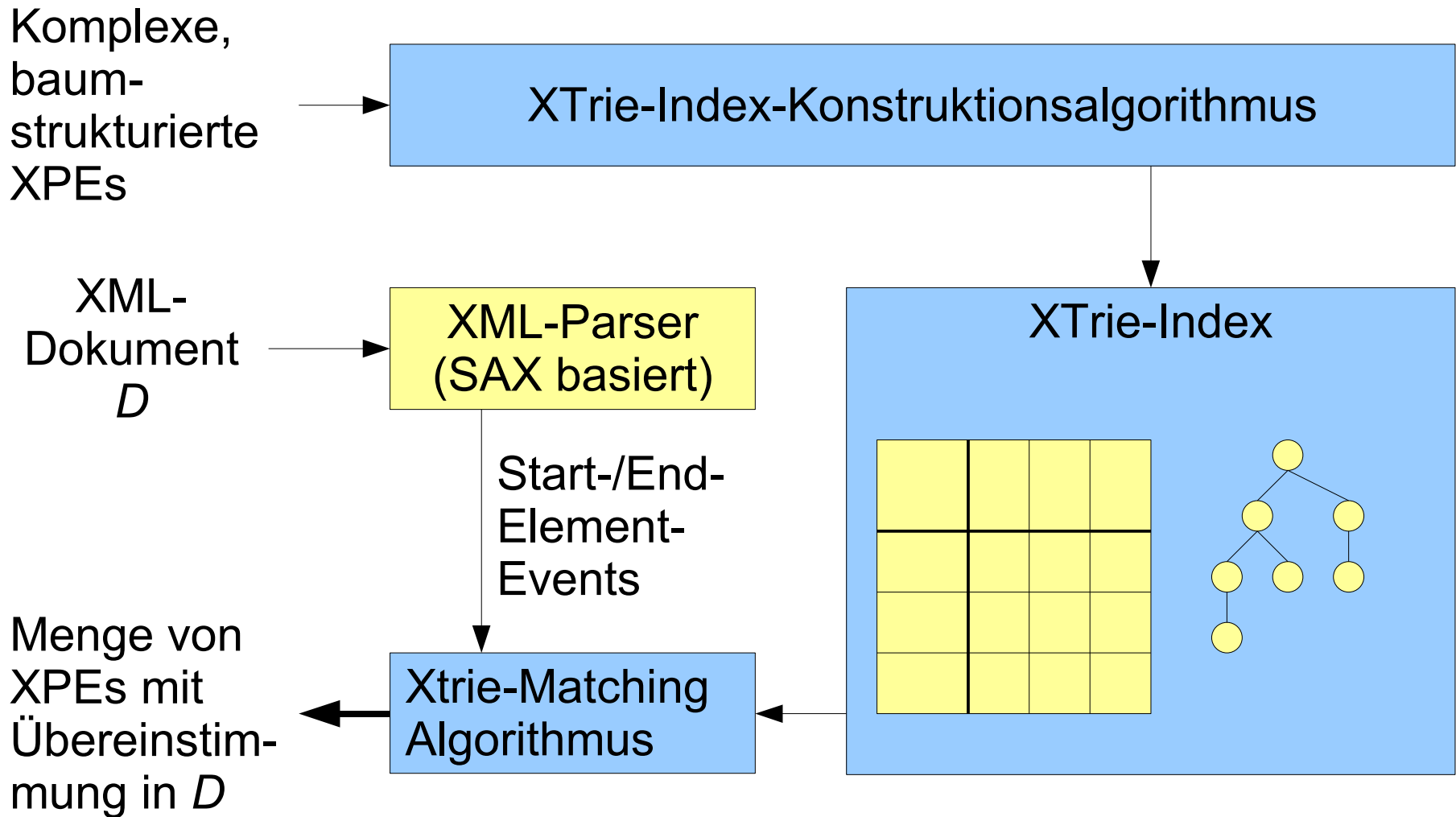
XTrie: Gleichzeitigkeit

→ Mehrere komplexe XPEs werden gleichzeitig verarbeitet

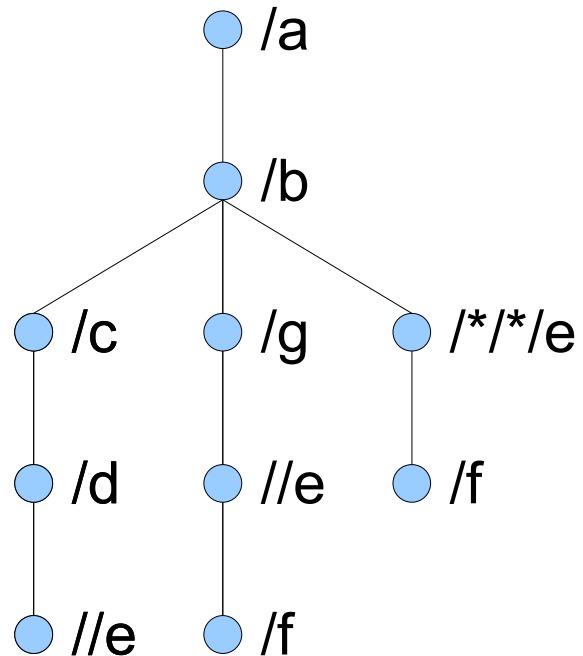
Der SAX-Parser (Simple API for XML)

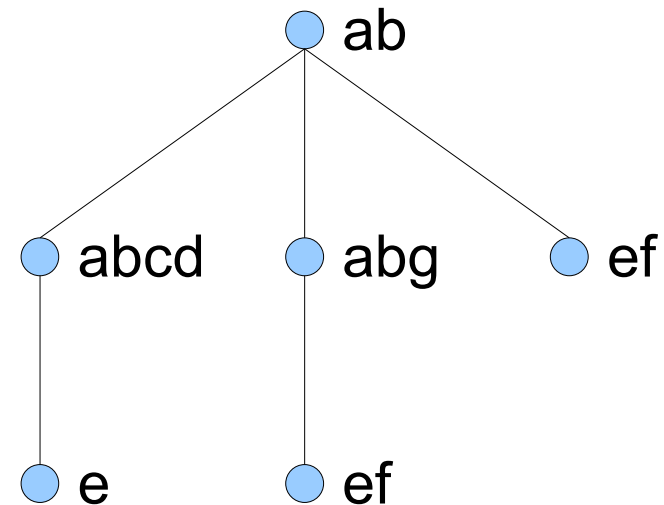
- Event-basiert:





XPE (XPath Expression)

$$p = /a/b[c/d//e][g//e/f]/**/*e/f$$


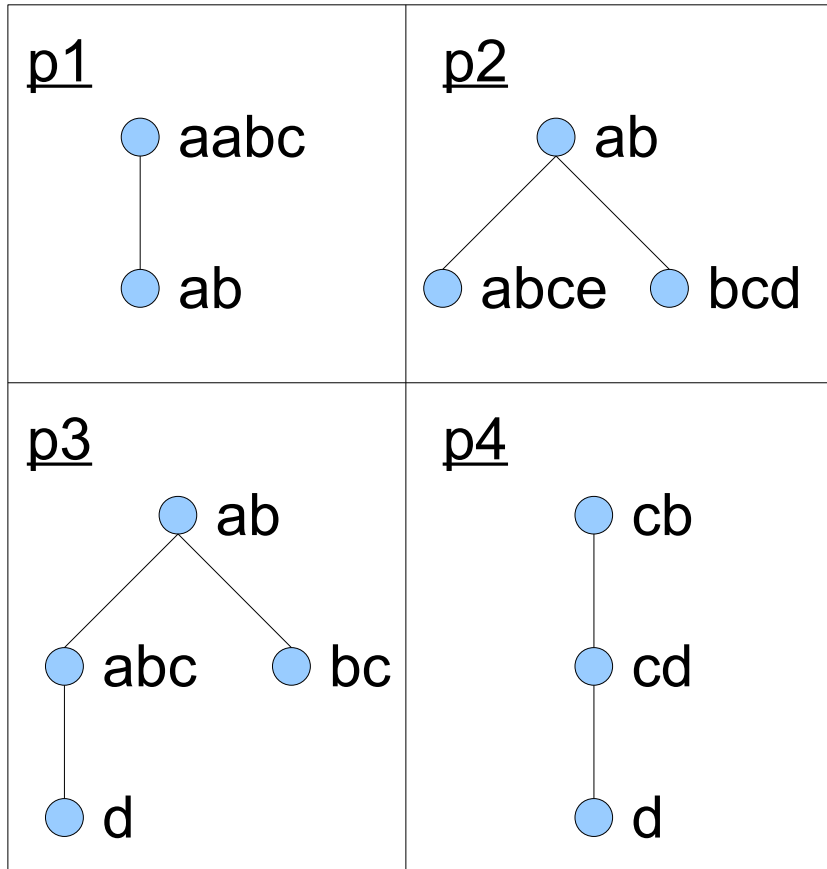
$$S = \{ ab, abcd, e, abg, ef, ef \}$$


p1 = //a/a/b/c/*/a/b

p3 = /a/b[c/*/d]//b/c

p2 = /a/b[c/e]/*/b/c/d

p4 = //c/b//c/d/*/*/d

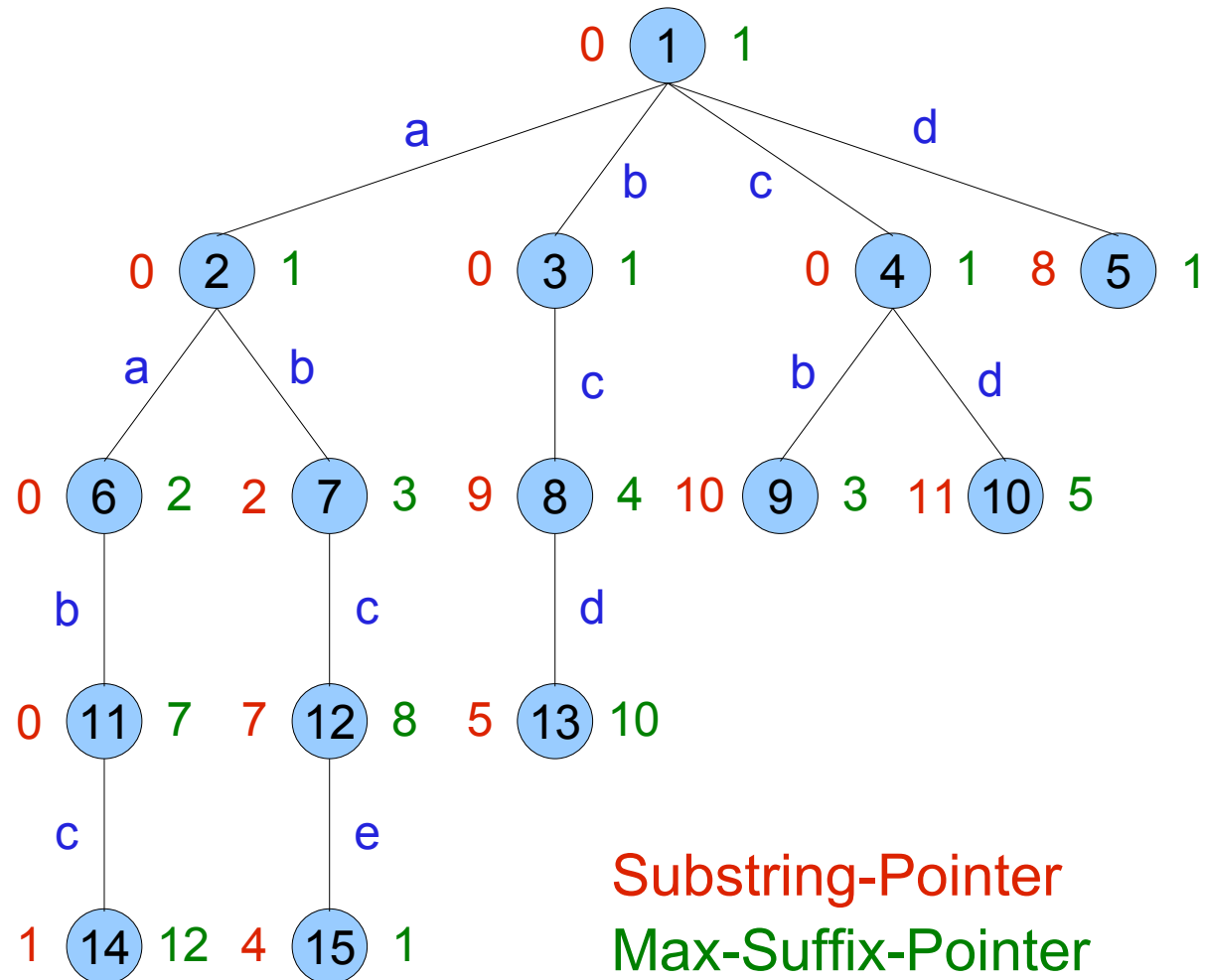


→ **Substring-Tabelle**

		Parent Row	Rel Level	Rank	Num Child	Next
aabc	1	0	[4,∞]	1	1	0
ab	2	1	[3,3]	1	0	3
ab	3	0	[2,2]	1	2	6
abce	4	3	[2,2]	1	0	0
bcd	5	3	[4,4]	2	0	0
ab	6	0	[2,2]	1	2	0
abc	7	6	[1,1]	1	1	0
d	8	7	[2,2]	1	0	12
bc	9	6	[2,∞]	2	0	0
cb	10	0	[2,∞]	1	1	0
cd	11	10	[2,∞]	1	1	0
d	12	11	[3,3]	1	0	0

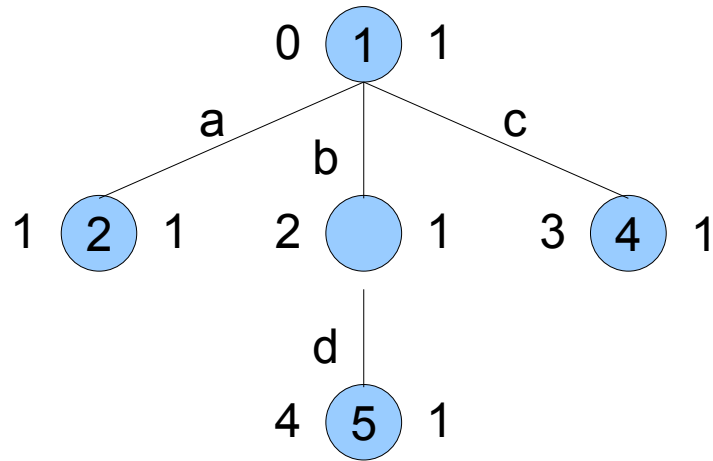
		Parent Row	Rel Level	Rank	Num Child	Next
aabc	1	0	[4,∞]	1	1	0
ab	2	1	[3,3]	1	0	3
ab	3	0	[2,2]	1	2	6
abce	4	3	[2,2]	1	0	0
bcd	5	3	[4,4]	2	0	0
ab	6	0	[2,2]	1	2	0
abc	7	6	[1,1]	1	1	0
d	8	7	[2,2]	1	0	12
bc	9	6	[2,∞]	2	0	0
cb	10	0	[2,∞]	1	1	0
cd	11	10	[2,∞]	1	1	0
d	12	11	[3,3]	1	0	0

Der XTrie

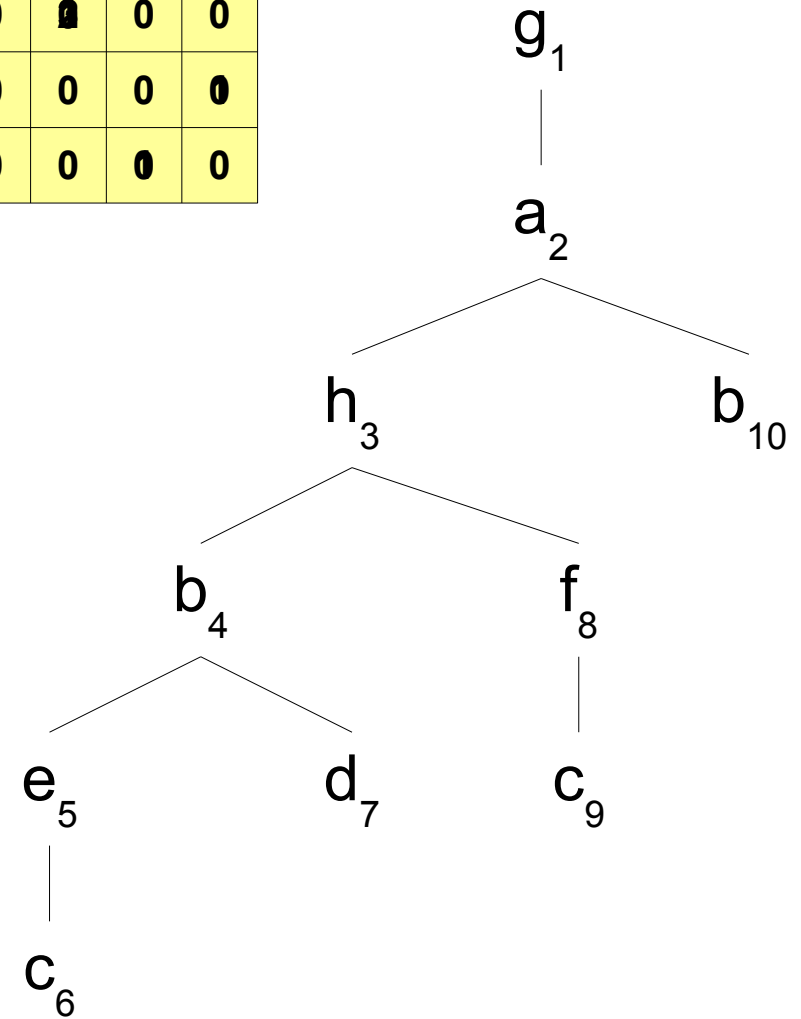
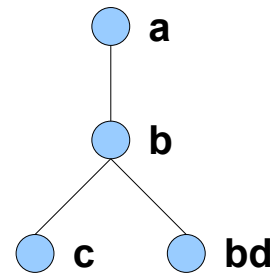


Der XTrie-Algorithmus

$p = //a//b[*]/c//d$



B	L1	L2	L3	L4	L5	L6
a	0	1	0	0	0	0
b	0	0	0	1	0	0
c	0	0	0	0	0	0
bd	0	0	0	0	0	0



		Parent Row	Rel Level	Rank	Num Child	Next
a	1	0	[1,∞]	1	1	0
b	2	1	[1,∞]	1	2	0
c	3	2	[2,2]	1	0	0
bd	4	2	[1,1]	2	0	0

XTrie - Fazit

Pro:

- Entwickelt für das Durchsuchen großer Datenmengen mit sehr vielen Suchausdrücken
- Streaming-fähig durch SAX

Contra:

- schwierig zu verstehendes Paper
- ?

**Vielen Dank
für die Aufmerksamkeit!**

Quellen:

- „Efficient filtering of XML documents with XPath expressions“,
C. Chan, P. Felber, M. Garofalakis, R. Rastogi
- Google