

Querying mit XPath



Prof. Dr. Christof Schöch

Modul Auszeichnungssprachen
MSc. Digital Humanities, Universität Trier



Überblick



1. Motivation: Wozu XPath?
2. Das XML-Datenmodell von XPath
3. Lokalisierungsschritte
 1. Achsen
 2. Knotentest
 3. Prädikate
4. Operatoren und Funktionen

(1) Motivation: Wozu XPath?

Motivation

- Wir können XML-Dokumente erstellen
- Wir können XML-Dokumente validieren
- Aber was dann?
 - Wie bekommen wir Informationen wieder heraus?
 - Wie können wir die Informationen weiterverarbeiten?

Motivation

- Wir können XML-Dokumente erstellen
- Wir können XML-Dokumente validieren
- Aber was dann?
 - Wie bekommen wir Informationen wieder heraus?
 - Wie können wir die Informationen weiterverarbeiten?

Antwort: XPath!

Was ist XPath?

- XPath bedeutet XML Path Language
- Mit XPath können verschiedenen Teile eines XML-Dokuments präzise adressiert werden
- XPath wird vom W3C entwickelt und gepflegt
- XPath wird selbst nicht in XML formuliert
- Die aktuellste Version von XPath ist 3.1 (2017), aber oft wird nur 1.0 unterstützt
- XPath kommt meist in anderen Sprachen eingebettet vor

XPath im Kontext

- XPath: Navigation in XML-Dokumenten
- XQuery: Abfragen auf XML-Dokumenten (verwendet XPath)
- XSLT: Transformation von XML-Dokumenten (verwendet XPath)

(2) Das XML-Datenmodell von XPath

XML in XPath: Baum

- Dokument-Knoten (nur konzeptuell)
- Wurzel-Element (oberstes Element in der Hierarchie)
- Element-Knoten (Name und Inhalt)
- Attribut-Knoten (Name und Wert)
- Text-Knoten

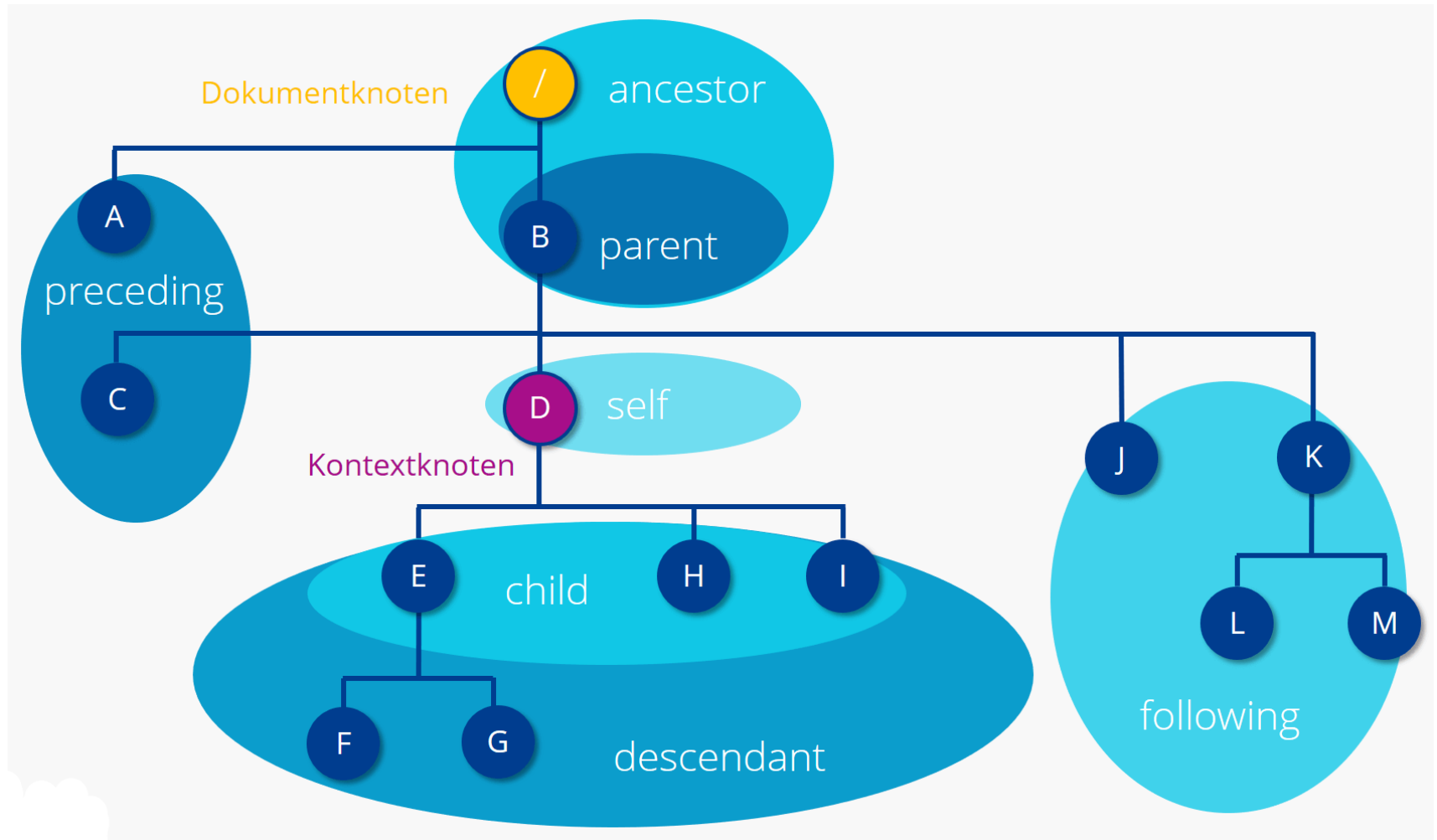
Dokumentknoten vs. Kontext-Knoten

- Jeder XPath geht von einem Punkt im Baum aus
 - Entweder vom Dokumentknoten aus (absoluter Pfad)
 - Oder vom Kontext-Knoten aus (relativer Pfad)
- Die "Achsen" bestimmten die Richtung / Orientierung

Die "Achsen" in XPath

Achse	Deutsch	Beschreibung
child	Kind	Alle direkt untergeordneten Kind-Knoten
descendant	Nachkommen	Alle untergeordnete Knoten
following	Nachfolgende Knoten	Alle im Dokument nachfolgenden Knoten (ohne Nachkommen)
following-sibling	Nachfolgende Geschwisterknoten	Alle im Dokument nachfolgenden Knoten des selben Elternknotens
attribute	Attribut	Alle Attributknoten eines Elementknotens
ancestor	Vorfahr	Alle übergeordnete Knoten
parent	Elternknoten	Der direkt übergeordnete Elternknoten
preceding	Vorhergehende Knoten	Alle im Dokument vorangehenden Knoten (ohne Vorfahren)
preceding-sibling*	Vorhergehende Geschwisterknoten	Alle im Dokument vorangehenden Knoten des selben Elternknotens

XPath-Achsen visualisiert



(Quelle: "XPath-Tutorial für Einsteiger", *Digital-Guide*, IONOS, 28.9.2020, [URL](#).)

(3) Die Syntax der Lokalisierungspfade

Lokalisierungspfade: Drei Teile

1. **Achse:** Richtung der weiteren Navigation (vom Dokumentknoten oder Kontextknoten aus)
2. **Knotentest:** Filter auf den in der Achse liegenden Knoten (bspw. Element oder Attribut)
3. **Prädikate** (optional): Weiterer Filter (bspw. Attribut-Wert)

Ausgangspunkt

- Dokumentknoten (absoluter Pfad): /
- Kontextknoten (relativer Pfad): . / ("aktueller Punkt", optional)
- beliebige untergeordnete Position: / /

(A) Achsen

Achse	ausführlich	abgekürzt
child	child::	/
parent	parent::	..
attribute	attribute::	@

(B) Knotentest

- Name eines Elements
 - `//figure`
 - `/wlv/metadata/curator`
- Name eines Attributs
 - `//@ref`
 - `//figure/@figureType`
 - `//@*` (beliebiges Attribut)

(C) Prädikate

- Wert eines Attributs
 - `//figure[@figureType="symbolic"]`
- Position
 - `div[1]`
 - `div[last()]`
- Ausgabe
 - `text()` (default)
 - `name()`

Operatoren und Funktionen

Operatoren

- kleiner / größer als: `<` / `>`
 - `figure[@figureNum < 3]`
 - `figure[@figureNum > 1]`
- und, oder: `and` bzw. `or`
 - `//figure[@figureType="symbolic" or @figureType="realistic"]`
- `uvm.`

Funktionen

- `count()`
 - `count (//figure)`
 - `count (//wineOrigin)`
- `sum()`
 - `sum (//figure/@figureNum)`
- `string-length()`
- `uvm.`

XPath in der Praxis

Beispiel Weinetiketten

XML-Kodierung

```
<!--?xml version="1.0" encoding="UTF-8"?-->
<wlv xmlns="https://tcdh.uni-trier.de/wlv/0.1">
  <metadata labelid="scs-piesport-0001" collectionid="scs">
    <curator curatorid="#cs">
      <curationdate>2020-11-18</curationdate>
    </curator></metadata>
  <label labeltype="Lagenetikett">
    <labelpart parttype="front" partnum="1">
      <physical sizeh="158" sizev="89" shape="rectangle" material="paper">
        <visual>
          <figure figuretype="realistic" figurenum="1" figureposition="upper-across">Landscape: river, vineyards, villa
          <figure figuretype="coat-of-arms" figurenum="2">Red coat of arms.</figure>
          <figure figuretype="symbolic" figurenum="3">wine leaves</figure>
        </visual>
        <textual>
          <wineorigin wineorigintype="Bereich">Mosel - Saar - Ruwer</wineorigin>
          <wineorigin wineorigintype="Lage" wineoriginlocality="Piesport" ref="wikidata:Q375762">Piesporter Lay</wineor
          <qualitygrapes qualitygrapesnorm="Auslese">Auslese</qualitygrapes>
          <agent agentrole="producer/bottler">Joh. Koenen-Veit, Weinbau - Weinversand</agent>
        </textual>
      </physical></labelpart>
    </label>
  </wlv>
```


Ausprobieren!

- XPath-Tester
- Oder
 - Aufrufen: <http://videlibri.sourceforge.net/cgi-bin/xidelcgi>
 - XML-Dokument reinkopieren
 - Dann XPath-Ausdrücke ausprobieren

Abschluss

Lektürehinweise

Grundlagen / Referenz

- "Einführung in XPath", *SelfHTML Wiki*, Version vom 13. Mai 2020, URL: <https://wiki.selfhtml.org/wiki/XML/XSL/XPath>.

Weitere Empfehlungen zur Vertiefung

- David Hunter et al.: "Chapter 7: XPath" in: *Beginning XML*, 4th edition. Wiley, 2007.
- Doug Tidwell, "XPath: A Syntax for Describing Needles and Haystacks", in: *XSLT*, 2nd edition. O'Reilly, 2008.

Danke!

Lizenz: [Creative Commons Attribution \(CC BY\)](#), 2020.
