

Aufgabe 11b - Benutzung von XPath Ausdrücken

Christof Schöch

DARIAH-DE

Lizenz: Creative Commons Attribution 4.0 (CC-BY).

Das vorliegende Dokument wurde für DARIAH-DE neu erstellt.

Aufgabe: Benutzung von XPath-Ausdrücken

1. Lernziele

Nach dieser Aufgabe sollten Sie Folgendes erreicht haben:

- * Sie können einfache XPath-Ausdrücke definieren, um spezifische Teile eines XML-Dokuments zu identifizieren.
- * Sie können in Oxygen (oder in einem anderen, geeigneten Editor) XPath-Ausdrücke in dieser Weise verwenden.
- * Außerdem veranschaulicht die Aufgabe noch einmal die Baumstruktur von XML und ruft einige wichtige TEI-Elemente in Erinnerung.

2. Überblick

Grundlage dieser Aufgabe ist die digitale Edition der gesammelten Werke von Shakespeare in der Ausgabe des University of Oxford Text Archive (Datei `Shakespeare.xml`). Es handelt sich um ein ziemlich umfangreiches Dokument: möchte man seine Struktur verstehen, ist die Baumstruktur hilfreich; möchte man Teile davon suchen und ansehen, ist XPath dafür nützlich. Beides ist Gegenstand dieser Aufgabe.

3. Vorbereitung

Öffnen Sie die Datei `Shakespeare.xml` mit Oxygen.

- * In dem Bereich von Oxygen auf der linken Seite (mit dem Titel Outline), können Sie sich über die hierarchische Struktur des Dokuments orientieren.
- * (Wird dieser Bereich bei Ihnen nicht angezeigt? Sie können ihn über das Menü `Windows > Show View > Outline` aktivieren; dort können Sie auch weitere Ansichten aktivieren.)

1. Die Struktur der Datei verstehen

Oxygen generiert automatisch eine interaktive Ansicht der hierarchischen Baumstruktur einer geöffneten XML-Datei. Diese Ansicht wird im Bereich `Outline` angezeigt. Beantworten Sie mit Hilfe der Outline-Ansicht die folgenden Fragen bezüglich der Shakespeare-Datei:

- * Wie lautet der Wurzel-Knoten?
- * Wie viele (unmittelbare) "Kinder"-Knoten hat der Wurzel-Knoten?

- * Wie lautet der (unmittelbare) "Eltern"-Knoten des Knotens mit dem Element `text` mit Attribut `xml:id="fol-MM"`?
- * Werden Akte und Szenen strukturell unterschieden, und wenn ja, wie?
- * Wieviele Teile hat das Stück A Midsommer Night's Dreame (Kürzel: fol-MND)?
- * Welches Element beinhaltet die Titel der einzelnen Stücke, und wie lauten ihre Geschwister-Knoten („siblings“)?

2. Vorgegebene XPath-Ausdrücke

Jetzt geht es um die eigentlichen XPath-Ausdrücke. Oxygen hat eine Suchfunktion, die nicht Volltext oder Reguläre Ausdrücke nutzt, sondern XPath-Ausdrücke versteht. Das entsprechende Suchfeld befindet sich links oben und ist mit "XPath 2.0" gekennzeichnet. Setzen Sie jeweils geeignete XPath-Ausdrücke in das Suchfenster ein und bestätigen Sie mit `return`. Das Ergebnis der jeweiligen Suche erscheint in einem Fenster unter dem Textfenster. Kontrollieren Sie, ob ihr XPath-Ausdruck die gewünschten Teile der Datei angesprochen hat. Setzen Sie die folgenden Ausdrücke ein und notieren Sie das Ergebnis: getroffene Zeile(n), und/oder Anzahl von Treffern, und/oder sonstiges Ergebnis (Text):

- * `//titlePage`
- * `/TEI/teiHeader/fileDesc/titleStmt/author/text()`
- * `/TEI/teiHeader/fileDesc/titleStmt/author`
- * `/TEI/teiHeader/fileDesc/publicationStmt`
- * `//idno/attribute()`
- * `//idno[attribute()="isbn13"]`
- * `//text/body/div/stage`
- * `/TEI/text/group/text/body/div[2]/stage[2]`

Welche der obigen Ausdrücke sind relativ, welche sind absolut?

3. Einfache XPath-Ausdrücke finden

Finden Sie für die folgenden gewünschten Ergebnisse passende XPath-Ausdrücke; wenn Sie den XPath-Ausdruck für ein Ergebnis gefunden haben, schreiben Sie ihn hier in die Liste. Manchmal gibt es vielleicht zwei mögliche Lösungen; notieren Sie beide und beschreiben Sie kurz, welche Variante Sie aus welchen Gründen besser finden.

- * Der Knoten, der die vollständige Post-Adresse des University of Oxford Text Archive enthält
- * Genau den Text der Email-Adresse des University of Oxford Text Archive
- * Alle Abschnitte (`divs`) des gesamten Dokuments; wie viele sind es?

4. Komplexere Suchanfragen

Jetzt geht es darum, etwas komplexere Suchanfragen zu gestalten oder Fragen über das Dokument zu klären. Notieren Sie jeweils den XPath-Ausdruck und ggfs.

die inhaltliche Antwort. Gehen Sie ansonsten vor wie im vorigen Abschnitt.

- * Die 13-stellige ISBN der digitalen Ausgabe
- * Wieviele Bühnenanweisungen enthält die Shakespeare-Ausgabe?
- * Wieviele dieser Bühnenanweisungen lauten nur Exit.?
- * Wieviele Bühnenanweisungen enthält das Stück Much Adoe About Nothing?
(id: fol-MAN)
- * Den Text aller Bühnenanweisungen in The Tempest (id: fol-Tem)
- * Wie oft ergreift Hamlet („Ham.“) das Wort in Hamlet (in fol-HAM)?
- * Wer spricht als letztes in Hamlet?
- * Und als Sternchen-Frage: Was sind Hamlets letzte Worte?

Das ist das Ende der Übung!