Klare Pfade: Xpath



Nr. 13: Die SEO Campixx 2021 als Hybrid – websiteboosting.com

So langsam kommt die Konferenzszene wieder in Schwung, wenn auch meist mit einer reduzierten Anzahl an Teilnehmenden und einer gleichzeitigen Übertragung von Inhalten an Zuhausegebliebene. Neben den fachlichen Inputs in den über 70 Vorträgen in sieben parallelen Slots war es erfrischend und…

Klare Pfade: Xpath

Sebastian Adler erklärte, was ein Xpath ist und wie man ihn für SEO-Zwecke einsetzen kann. Im Prinzip ist der Xpath die Adresse eines einzelnen Teils eines HMTL-Dokuments. Die URL selbst stellt die Adresse dar, wo ein Dokument zu finden ist im Web, und der Xpath zeigt die genaue Adresse, mit der man eine Information im sog. DOM, dem Objektbaum des Dokuments, ansprechen kann. Stellt man sich ein Worddokument vor, bei dem vereinfacht erklärt jeder Satz und jedes Objekt eine eigene Gliederungsnummer hätte, könnte man mit dem Dateinamen (wo liegt die Datei) und dem Gliederungspunkt (wo im Dokument liegt das Objekt, das man haben möchte) eben nur dieses eine Objekt „herauskopieren“. Nichts anderes macht man mit dem Xpath. Er stellt quasi die Wegbeschreibung zu einem gesuchten Element dar. Da der Xpath Groß- und Kleinschreibung unterscheidet, liegt hier manchmal die erste Stolperfalle bei der Anwendung. ist eben nicht gleich <Div> und so empfahl Adler dringend, alles jeweils sauber zu testen, bevor man großflächig Daten einsammeln möchte.

Typische Anwendungen für SEO wären Fragen wie: Was steht im Title, welche Bilder gibt es auf einer Seite, welche externen Links gehen von einer Seite weg, wann wurde der Beitrag

publiziert und einige andere mehr. So liefert zum Beispiel der Xpath „//title“ oder „//H1“ zusammen mit der URL eines HTML-Dokuments alle Title- und H1-Einträge zurück. Für den Chrome-Browser gibt es das kostenlose Add-in „Xpather“, für Firefox „Try Xpath“, mit denen man gut testen kann, ob man die richtigen Adressen ermittelt hat. Toolgestützt, etwa via SEO-Tools für Excel, Google Spreadsheets oder den Screaming Frog kann man beim Crawlen von Dokumenten unter Angabe es Xpath gleichzeitig ohne Zusatzaufwand die entsprechenden Informationen mit wegschreiben. Natürlich lassen sich damit auch Preise, Verfügbarkeiten, die Anzahl Bewertungssterne etc. adressieren.

Sebastian Adler zeigte einige praktikable Beispiele, die relativ leicht umsetzbar sind und mit denen man via Xpath an Daten bzw. Erkenntnisse kommen kann. Eines der Beispiele bezog sich auf Featured Snippets in den Google Suchergebnissen. Im konkreten Fall ging es darum, den Text von Featured Snippets für verschiedene Suchbegriffe zu extrahieren. Dazu sucht man im Quelltext der Suchergebnisseite von Google das Element, das den Textinhalt trägt:

```
<div>
<div> hier steht viel Zeug (u.a. das Bild)</div>
<div class="mod"> Featured Snippet Text</div>
<div class="g">Title + URL mit Link</div>
</div>
```

Über den Xpath

```
//*[@class='mod']/..
```

und die zugehörige URL wie z. B. www.google.de/search?q=“Suchwort“&hl=de lässt sich dann eine automatische Abfrage zusammenbauen. Statt „Suchwort“ muss natürlich das jeweilige Keyword verwendet werden.

So erzeugt z. B. www.google.de/search?q=“geldmarktzins“&hl=de eine Suchergebnisseite mit einem Featured Snippet, wie in Abbildung 7 zu sehen ist. Den Text „Der Geldmarktzins ist der

Zins…“ kann man dann mit dem Xpath //*[@class=’mod‘]/.. extrahieren. Besonders leicht lässt sich das dann z. B. mit den SEO Tools für Excel lösen. Man hinterlegt in der Spalte A alle Keywords in Zeilen und fügt diese dann über eine Formel rechts daneben in Spalte B in die Google-Such-URL ein bzw. „baut diese zusammen“. Das sieht dann z. B, so aus:

https://www.google.de/search?q="&A1&"&hl=de

Erzeugt wird durch diese Formel die vorherige URL www.google.de/search?q=“geldmarktzins“&hl=de. In der Spalte daneben an den Xpath ein und über das automatische Ausfüllen nach unten über alle befüllten Zeilen wird das Abholen aller Daten ausgelöst. Allerdings sollte man hier nicht zu viele Abfragen auf einmal machen, damit Google den (eigenen) Client nicht kurzzeitig sperrt.

Abbildung 7: Per Xpath Texte aus Suchergebnissen holen

Abbildung 8: Mit Tools wie Xpather zieht man Elemente direkt und einzeln von jeder HTML-Seite (Quelle: Sebastian Adler)

Adlers besonderer Tipp galt der Dokumentation. Man solle lieber ausführlicher als zu knapp dokumentieren, was man mit einem oder mehreren Xpath-Anweisungen beabsichtigt hat. Das hilft später enorm bei erneuter Verwendung.