BESSER INTERN VERLINKEN MIT SCREAMING FROG UND KNIME

Michael Göpfert

Die Optimierung der internen Verlinkung gehört für viele SEOs zum Tagesgeschäft. Abseits von Metriken wie PageRank und CheiRank gibt es bei einer strategischen Optimierung der Ankertexte häufig low hanging fruits zu ernten. Um im Dschungel von Abertausenden Links und Ankertexten nicht den Überblick zu verlieren, lernen Sie in diesem Beitrag, wie mit Screaming Frog und KNIME auch große Mengen an Verlinkungen effizient analysiert werden können.

Ziel ist es, Widersprüchlichkeiten in der internen Verlinkung zu erkennen und zu prüfen, ob die Verlinkungsstrategie mit den Traffic-Daten aus der Search Console in Einklang ist.

kung wurde schon in der vergangenen Ausgabe berichtet. Ein wichtiges Learning ist: Versuchen Sie stets, eindeutige Signale an Suchmaschinen zu senden. Im Kontext der Verlinkung ist eines dieser Signale der Linktext (auch Ankertext genannt). Er gibt Suchmaschinen wichtige Hinweise darauf, worum es sich auf der verlinkten Seiten (der Zielseite) handelt, und ist damit ein wichtiger Faktor der On-Page-Optimierung.

Über die Wichtigkeit von (interner) Verlin-

Um in diesem Zusammenhang keine widersprüchlichen Signale an die Suchmaschinen zu senden, sollte nicht mit demselben Ankertext auf unterschiedliche Zielseiten verlinkt werden. Im Laufe der Zeit werden Websites immer größer und unübersichtlicher. So kann es leicht passieren, dass unbewusst gegen dieses Prinzip verstoßen wird. Leider ist das Aufdecken solcher Fehler recht aufwendig. Vor allem große Websites verfügen häufig über so viele interne Verlinkungen, dass ein Audit mit herkömmlichen Mitteln nur schwer durchführbar ist.

Mit der Hilfe des Website-Crawlers Screaming Frog und dem Datenanalysetool KNIME lassen sich auch bei großen Websites Potenziale in der internen Verlinkung schnell und einfach aufdecken.

In diesem Beispiel werden folgende Fragestellungen beantwortet:







gerne mit Rohdaten, um

diese in maßgeschneiderten Analysen für seine Kun-

den aufzubereiten

TIPP

Wie Sie an das kostenlose Tool KNIME kommen und wie es prinzipiell funktioniert, finden Sie in der Ausgabe 53 oder online frei als HTML oder PDF unter http://einfach.st/knime53.



- » Wird mit einem Keyword als Ankertext auf Zielseiten verlinkt, die nicht der für dieses Keyword vorgesehenen Seite entsprechen?
- » Geben die Daten aus der Search Console Hinweise auf Potenziale und Anomalien in der internen Verlinkung?

Missmatch zwischen Ankertext und Zielseiten aufdecken

Für die erste Fragestellung benötigt es zunächst einen Crawl der Website mit dem Screaming Frog SEO Spider. Anschließend muss unter Bulk Exports > Links > All Inlinks ein Export als CSV-Datei gestartet werden. Dieser enthält sämtliche gecrawlten Verlinkungen der Website inklusive ihrer Ankertexte (Abb. 1).

Dieser Export kann bei großen Websites leicht zu einer sehr großen Datei führen. Davon sollten Sie sich aber nicht abschrecken lassen, denn mit KNIME lassen sich auch mehrere Gigabyte große Dateien performant verarbeiten.

Im nächsten Schritt sollten Sie nun in KNIME einen neuen Workflow anlegen und die aus Screaming Frog exportierte Datei mit der Node CSV Reader importieren.

Dazu muss lediglich die Node per

Bulk Export	Reports	Sitemaps	Visua	lisatio	ons Cra	awl Ana
Queued URLs		weg-com	- Screa	ming	Frog SE	O Spide
Links		All Inlinks				
Web	>	All Outlin	ks			
Path Type	>	All Ancho	r Text	ent	Images	Canon
Security	>	External L	inks			
Response Coo	des >					
Content	>					
Images	>	28155ea4bo	d69e.css			
Canonicals	>					

Abb. 1: Export aller internen Links aus Screaming Frog



Abb. 2: Verbundene Nodes in KNIME



Abb. 3: Herausfiltern einer URL mit dem Row Filter

Drag-and-drop aus dem Node Repository auf die Arbeitsfläche gezogen werden und anschließend muss das Konfigurationsmenü mit einem Doppelklick geöffnet werden.

Nun kann die aus Screaming Frog exportierte CSV-Datei ausgewählt werden. Im unteren Teil des Konfigurationsfensters wird eine Vorschau der zu importierenden Datei angezeigt. Sieht diese korrekt aus, kann das Fenster mit Ok geschlossen werden und die Node kann mit einem Rechtsklick und dem Befehl Execute ausgeführt werden.

Springt nun die "Ampel" der Node auf Grün sind die Daten erfolgreich importiert worden. Das Ergebnis des Imports lässt sich mit einem Rechtsklick auf die Node und der Auswahl des Menüpunkts File Table begutachten.

TIPP

Die Node Row Filter ermöglicht es Ihnen, auch mithilfe von regulären Ausdrücken oder Wildcards zu filtern. So lassen sich auch Linktexte filtern, die diesem nicht exakt entsprechen, sondern ihn lediglich enthalten.

Search Analyti	cs for Sheets		×
Requests	Backups	Abou	ıt
Verified Site:			0
https://www.	federweg.com/		\$
Date Range:			0
2021-04-	06 to	2022-08-16	
Search Type:			0
Default (Web)			\$
Group By:			0
Query X P	age 🗙		
Filter By:			0
Add Filter			
Aggregation T	ype:		0
Default (auto)			\$
Rows returned	1:		0
Everything			\$
Results Sheet	:		0
Create New S	heet		\$
Include fre	esh data		0
URL Inspe	ection (beta)		0
Request Data			

Abb. 4: Daten aus der Search Console via Search Analytics for Sheets exportieren

Zielseiten und Ankertexte filtern

Die nächste Aufgabe besteht darin, diesen unübersichtlichen Datensatz so aufzubereiten, dass er Ihnen hilft, die Seiten zu finden, die mit einem bestimmten Ankertext auf Seiten verlinken, die nicht der gewünschten Zielseite entsprechen.

Dazu wird nun mit der Node Row Filter die gewünschte Zielseite aus den Daten herausgefiltert. Auch diese Node kann aus dem Node Repository auf die Arbeitsfläche gezogen werden. Anschließend kann das rechte schwarze Dreieck (der sogenannte Exit Port) des CSV Readers mit dem linken schwarzen Dreieck des Row Filters verbunden werden (Abb. 2).

Um eine bestimmte Seite aus dem Datensatz zu filtern, werden in der Konfiguration der Node Row Filter folgende Einstellungen vorgenommen (Abb. 3):

- » Column to Test: Destination (diese Spalte im Export enthält die Zielseiten)
- » Matching Criteria: URL der Zielseite, die aus dem Datensatz gefiltert werden soll
- » Ganz links muss noch der Wert Exclude rows by attribute value aktiv sein. Dies stellt sicher, dass die eingetragene URL aus dem Datensatz herausgefiltert wird.

Anschließend wird dem Workflow eine weitere Node Row Filter hinzugefügt. Diese sorgt dafür. dass nur noch Seiten im Datensatz enthalten sind, die einen bestimmten Ankertext beinhalten. In der Konfiguration wird nun Folgendes angegeben:

» Column to Test: Anchor (die Spalte

mit den Ankertexten)

» Matching Criteria: der Ankertext bzw. das Keyword, für das die gewünschte Zielseite ranken soll Nun enthält der Datensatz aus-

schließlich Verlinkungen, die mit dem definierten Ankertext auf eine andere URL verlinken als die dafür vorgesehene Ziel-URL. Sie können nun die Verlinkungen auf Ihrer Website so anpassen, dass jeder wichtige Ankertext auf die dafür vorgesehene Zielseite verlinkt.

Doch welche Zielseite passt zu welchem Ankertext? Das Matching von Ankertext und Zielseite ist nicht immer einfach. Für die wichtigsten Hauptkeywords lässt sich das häufig noch mithilfe einer Keyword-Tabelle überblicken, doch was ist mit dem Rest? Mit der Kombination von Search-Console-Daten und dem Link-Report von Screaming Frog lassen sich völlig neue Erkenntnisse gewinnen.

Linkdaten mit Daten aus der Search Console anreichern

Ziel des nächsten Schritts ist es, zu prüfen, ob es Seiten gibt, die über ein Keyword viele Impressionen bekommen, während dieses Keyword jedoch

	String Manipulation	n Flow Variables Memory Policy
Column List ROWID ROWINDEX ROWCOUNT IS Type Source Destination Size (Bytes) S Alt Text S Antchor S Status S Follow S Target S Rel S I Ink Path S I Ink Path	Category All Function capitalize(str, chars) compare(str, stars) count(str, toCount) count(str, toCount) count(hars(str, chars) countChars(str, chars, modifiers) index/Offstr, toSearch index/Offstr, toSearch, start) Expression 1 lowerCase(\$Anchor\$)	Description
O Replace	d Column: e Column: S Anchor S	□ Insert Missing As Null ☑ Syntax check on close
		OK Apply Cancel

Abb. 5: Alle Wörter in einer Spalte kleinschreiben mit der Node String Manipulation

als Ankertext für Links auf andere Seiten dient.

Um diese Anomalie aufzudecken, werden zwei Listen benötigt:

- 1. Eine Liste aller Ankertexte und die dazugehörigen Quell- und Ziel-URLs
- Eine Liste aller Keywords, die Impressionen bringen, und die dazugehörigen URLs

Für die erste Liste kann auf den Import aus dem ersten Teil dieses Workflows zurückgegriffen werden (all_ inlinks.csv aus Screaming Frog). Für die zweite Liste wird ein Export der Search-Console-Daten (idealerweise als CSV) mit den Suchbegriffen und den dazugehörigen URLs benötigt. Die Daten können in dieser Form leider nicht direkt aus der Google Search Console (GSC) exportiert werden, deshalb kann auf die kostenlose Google-Sheets-Erweiterung Search Analytics for Sheets zurückgegriffen werden (Abb. 4). Diese kann die Daten schnell und einfach aus der GSC in ein Google-Sheet importieren. Das Sheet wiederum kann als CSV-Datei exportiert und per Node CSV Reader in KNIME importiert werden.

Bevor beide Importe miteinander verbunden werden, müssen die Daten noch etwas aufbereitet werden.

Der CSV Reader für den Import der Ankertexte muss nun mit einer Node String Manipulation verbunden werden. Diese wird benötigt, um die Ankertexte in Kleinschreibweise zu verändern, da die Daten aus dem Search-Console-Export ebenfalls kleingeschrieben sind. Dazu muss die Node wie in Abbildung 5 konfiguriert werden. Zusätzlich kann im Bereich Function die Funktion lowerCase(str) ausgewählt werden. Sie sorgt dafür, dass alle Wörter in einer bestimmten Spalte (in diesem Fall die Spalte Anchor) kleingeschrieben werden. Anschließend muss lediglich festgelegt werden, dass die bestehende Spalte ersetzt werden soll.

Im nächsten Schritt muss die Tabelle so aufbereitet werden, dass es für jeden Ankertext nur noch einen Eintrag gibt. Dies lässt sich mithilfe der Node GroupBy umsetzen. Ist die Node auf der Arbeitsfläche und mit der vorherigen Node (String Manipulation) verbunden, wählen Sie im Tab Groups die Spalte Anchor aus, sodass diese im Bereich Group column(s) steht (Abb. 6).

Damit erreichen Sie, dass pro Eintrag in der Spalte Anchor nur noch eine Zeile in der Tabelle existiert.

Nun haben Sie die Möglichkeit, für jeden Ankertext Werte aus den anderen

Spalten zu aggregieren. Für diese Auswertung ist Folgendes relevant:

- » Eine Liste aller URLs, die mit dem jeweiligen Ankertext verlinkt sind
- » Eine Liste aller URLs, die mit dem jeweiligen Ankertext auf eine andere Seite verlinken
- » Die Anzahl der URLs, die mit dem jeweiligen Ankertext verlinkt sind
- » Die Anzahl der URLs, die mit dem jeweiligen Ankertext auf eine andere Seite verlinken

Konfigurieren Sie dazu die Node GroupBy wie in Abbildung 7.

Group settings	Groups Manual Aggre	gation Pattern Based	Aggregation Type	Based Aggregation	
Available column(s) T Filter S Type S Source S Destination I Size (Bytes) S Alt Text		> >	Group column(s) - <i>Filter</i> S Anchor		
Status Code Status Status Follow S Target Rel Path Type Link Path Link Position Link Origin		 × × 			
vanced settings					

Abb. 6: Tabelle nach Ankertexten gruppieren

Aggregation settings Available columns S Type S Source Destination I Size (Bytes) S Alt Text	Select	To change multi Column S Destination S Source S Source Destination	ple columns use right mouse clic Aggregation (click to cha Unique concatenate Unique concatenate Unique count	k for context m :nu. inge) Missi F irame V V V
Status Code Status S Follow S Target S Rel	add >> add all >>			_
S Path Type S Link Path S Link Position S Link Origin	<< remove			

Abb. 7: Aggregation der Spalten

	Dialog - 4:2	1 - Joiner ((Tabellen verbinden)		
	Joiner Settings Column Selectio	n Perfo	rmance Flow Variables	Memory Polic	y
loin columns					
Match O all o	of the following any of the following	1			
	Top Input ('left' table)		Bottom Input ('right'	table)	
S Anchor		0	S Query		+ -
Compare value Include in outp	s in join columns by 🧿 value and type ut	string	representation O mak	ing integer types	compatible
🗹 Matching ro	ws		Inner join		
Left unmato	hed rows		(\bigcirc)		
🗌 Right unma	tched rows				
Output options Split join re Merge join Hiliting ena	sult into multiple tables (top = matching columns bled	g rows, mic	ddle = left unmatched rov	vs, bottom = rigł	It unmatched rows)
Output options Split join re Merge join Hiliting ena	sult into multiple tables (top = matching columns bled	g rows, mio	ddle = left unmatched rov	vs, bottom = rigł	it unmatched rows)
Output options Split join re Merge join Hiliting ena Row Keys Concatenat	sult into multiple tables (top = matching columns bled e original row keys with separator _	g rows, mic	dde = left unmatched rov	vs, bottom = rigł	It unmatched rows)
Output options Split join re Merge join Hiliting ena Row Keys Concatenat Assign new	sult into multiple tables (top = matching columns bled e original row keys with separator row keys sequentially	g rows, mic	idle = left unmatched rov	vs, bottom = rigł	nt unmatched rows)

Abb. 8: Zusammenführen der Tabellen mit dem Joiner

Verbinden der beiden Datensätze

Ziehen Sie nun die Node Joiner auf Ihre Arbeitsfläche und verbinden Sie die Node mit der Node GroupBy mit den Ankertexten und den GSC-Daten. Um die Node zu konfigurieren, müssen zwei Einstellungen getroffen werden (Abb. 8):

- » Im Tab Joiner Settings werden die Spalten Anchor und Query ausgewählt.
- » Im Bereich Include in output wählen Sie Matching Rows.

Diese Einstellungen führen zu einem sogenannten Inner Join beider Tabellen. In diesem Fall bedeutet das, dass nur die Zeilen ausgegeben werden, bei denen die Suchanfrage (Query) und der Ankertext identisch sind. Alle anderen Zeilen werden nicht übernommen. Die Werte der restlichen Spalten bleiben in der Tabelle enthalten.

Die zusammengeführte Tabelle enthält nun folgende Spalten:

- » Ankertext (identisch mit der Suchanfrage)
- » Destination (Unique concatenate): die mit diesem Ankertext verlinkten Zielseiten
- » Source (Unique concatenate): die mit diesem Ankertext verlinkenden

Seiten

» Source (Unique Count): Anzahl der verlinkenden Seiten mit diesem Ankertext

- » Destination (Unique Count): Die Anzahl an URLs, die mit diesem Ankertext verlinkt sind
- » Query (identisch mit dem Ankertext)
- » Page: die URL, die laut den GSC-Daten die meisten Klicks über diese Query (=Ankertext) bekommen hat
- » Impressions: Anzahl der Impressionen, die die URL über diese Query bekommen hat
- » Clicks: Anzahl der Klicks, die die URL über diese Query bekommen hat

Erste Erkenntnisse

Sie haben es sicher schon selbst bemerkt: Ist in der Spalte Destination (Unique concatenate) mehr als eine URL enthalten, bedeutet dies, es wird auf der Website mit demselben Ankertext auf mehr als eine Seite verlinkt und damit wird ein mehrdeutiges Signal an Google gesendet.

•••	Dialog - 4	29 - Column Rename (Spalten)			
(Change colum	ns Flow Variables Memory Polic	у		
Column Search	Anchor			Remo	ove
Filter Options One	🗹 Change:	Ankertext / Query	5	StringValue	٢
Anchor Destination (Unique concat Source (Unique concatenate) Source (Unique count) Destination (Unique count) Query S Page	Destination (l	Jnique conca		Remo	ove
	🗹 Change:	StringValue	0		
	Source (Uniqu	Remo	ove		
D Impressions I Clicks	Change: Verlinkende Seite(n)				٢
	Source (Uniqu	e count)		Remo	ove
	Change: Anzahl Verlinkender Seiten				٢
	Destination (l	Jnique count		Remo	ove
	🗹 Change:	Anzahl Zielseiten	I IntValu	e	0
	Page			Remo	ove
	🗹 Change:	Page mit den meisten Impressionen	S	StringValue	٢
	Impressions			Remo	ove
	🗹 Change:	Impressionen	D Double	Value	0
	Clicks			Remo	ove
	🗹 Change:	Klicks	1 IntValu	e	٢
		OK Apply	Canc	el 🕜	

Abb. 9: Umbenennen von Spalten mit der Node Column Rename



Abb. 10: Abgleich der URLs via Rule Engine

Die Ergänzung der Daten aus Screaming Frog mit den Daten aus der Search Console ermöglicht es Ihnen nun, darüber hinaus zu sehen, ob die mit einem bestimmten Ankertext verlinkte Zielseite auch die ist, die laut GSC die meisten Klicks über das Keyword bekommt, das mit dem Ankertext identisch ist.

Um nicht den Überblick zu verlieren, sollten die Spalten etwas eindeutiger benannt werden. Dazu können Sie auf die Node Column Rename zurückgreifen. Einen Vorschlag für die Benennung der Spalten sehen Sie in Abbildung 9.

Mit der Node Rule Engine lässt sich anschließend abgleichen, ob die URL(s) in der Spalte Zielseite(n) identisch mit der URL in der Spalte Page mit den meisten Impressionen ist. Ist dies nicht der Fall, rankt Google nicht die Seite, die mit diesem Keyword als Ankertext am häufigsten verlinkt ist.

Dazu wird in der Rule Engine eine neue Spalte mit dem Namen Zielseite = Page mit den meisten Impressionen? erzeugt, die das Ergebnis dieser Prüfung enthält (Abb. 10).

Um die Fälle mit dem größten Potenzial sofort zu erkennen, können Sie mit der bereits bekannten Node Sorter die Tabelle so sortieren, dass sie nach der Anzahl an Klicks absteigend sortiert ist.

Um die Fälle, bei denen die Zielseite und die Seite mit den meisten Impressionen nicht übereinstimmen, sofort zu erkennen, können Sie diese mit der Node Color Manager farbig markieren (Abb. 11).

Anschließend können Sie die Daten in eine CSV-Datei (Node: CSV Writer) oder Excel-Datei (Node: Excel-Writer) exportieren.

Fazit

Zugegeben, dieser Workflow ist recht anspruchsvoll. Nicht im Hinblick auf den Aufbau in KNIME, sondern in Bezug auf diese spezielle Herangehensweise für die Optimierung der internen Verlinkung. Gleichwohl können so mit recht einfachen Mitteln Erkenntnisse gewonnen werden, die sich leicht auf der eigenen Website umsetzen und widersprüchliche Signale durch interne Verlinkung auflösen lassen. Wie immer gibt es den fertigen Workflow zum Herunterladen unter *http://einfach.st/ knime76*.

Viel Spaß beim Ausprobieren!



Abb. 10: Abgleich der URLs via Rule Engine