



Rebecca Schwarz

Low-Hanging-Fruit-Potenziale erkennen: SEO-Analyse mit KNIME

Nachdem der Fokus der KNIME-Serie in der letzten Ausgabe auf der Beschaffung und Aufbereitung von Daten aus der Google Search Console (kurz: GSC) lag, geht es nun um die Durchführung einer tatsächlichen Analyse für die Suchmaschinenoptimierung. Dazu wird ein Use-Case aus dem großartigen Vortrag „Das Low-Hanging-Fruit Framework“ von Mario Fischer, vorgestellt auf der SEOkomm 2023, aufgegriffen. Konkret geht es dabei um Folgendes: Low-Hanging-Fruit-Potenziale von URLs finden, indem untersucht wird, ob die Top-Suchanfrage in wichtigen SEO-Feldern (hier: Title, Meta-Description und H1-Überschrift) auftritt. Im nachfolgenden Workflow wird deshalb Schritt für Schritt gezeigt, wie eine solche Überprüfung aufgebaut werden kann. Der Beitrag zeigt, welches enorme Potenzial in einer eigenen Analyse von Search-Console-Daten liegt – die man über Tools von der Stange eben nicht einfach mal so und noch dazu kostenlos bekommt.

DIE AUTORIN



Rebecca Schwarz ist SEO-Consultant bei der get:traction GmbH. Ihr Arbeitsalltag dreht sich um die Konzeption von SEO-Strategien und die Unterstützung von Kunden in der redaktionellen SEO. Um größere Datenmengen effizient zu verarbeiten und um bei wiederkehrenden SEO-Tasks Zeit zu sparen, nutzt sie die Open-Source-Software KNIME.

Ad-hoc-Potenziale finden und wirklich ins Arbeiten kommen! Das ist das Ziel bei der Arbeit in KNIME. Denn die Arbeit in der SEO sollte sich vielmehr um die Analyse und das Ableiten konkreter Maßnahmen drehen als um die immer gleichen lästigen Prozesse der Datenbeschaffung und -aufbereitung. Wie die Arbeit mit KNIME effizienter wird, um Zeit für die wirklich wichtigen Dinge zu haben, soll dieser Workflow darstellen. Dazu ist Folgendes nötig:

- » KNIME-Umgebung
- » Zugang zur Google Search Console der ent-

sprechenden Website

- » Browser-Plug-in „Search Analytics for Sheets“

Vorbereitung der Daten

Für den Anwendungsfall wird die Top-Suchanfrage pro URL einer Website benötigt. Weil das nicht durch einen einfachen Export im Frontend der GSC möglich ist, hilft das Browser-Plug-in „Search Analytics for Sheets“. Dieses Plug-in kann im Chrome Web Store kostenfrei aktiviert werden. Die Konfiguration, sobald das Plug-in aktiviert ist, sieht so aus:

Foto: akiyoko / gettyimages.de

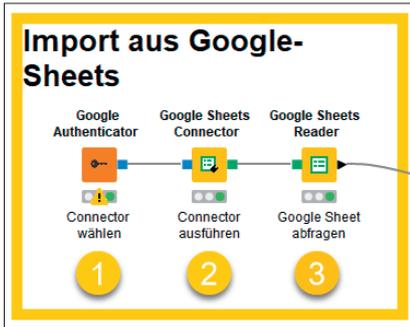


Abb. 1: Daten aus Google-Sheets in KNIME einlesen

1. Leeres Google-Sheet öffnen und Plug-in-Oberfläche über Extensions → Search Analytics for Sheets → Open Sidebar aufrufen
2. Gewünschten Zeitraum der Daten und den Suchtyp auswählen
3. Im Feld Group By die Felder „page“ und „query“ auswählen
4. Im Feld Aggregation Typ „by page“ auswählen
5. Auf „Request Data“ klicken, um die Abfrage zu senden und die Daten automatisch ins vorliegende Google-Sheet zu importieren

Beginn der Arbeit in KNIME

Mit den nun vorliegenden Daten wird mit der Arbeit in KNIME gestartet. In einem ersten Schritt wird in KNIME das Google-Sheet abgefragt, um damit in der grafischen Oberfläche weiterarbeiten zu können. In KNIME werden

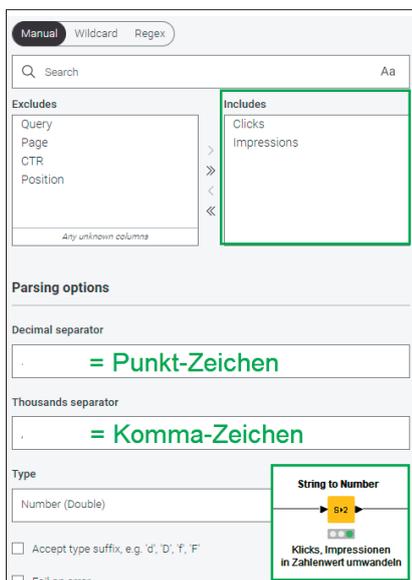


Abb. 2: Konfiguration im Knoten STRING-TO-NUMBER

dafür drei Knoten benötigt (Abbildung 1):

1. GOOGLE-AUTHENTICATOR:

Zur Konfiguration muss der Chrome-Browser, in dem das Google-Sheet erstellt wurde, geöffnet sein. Der Prozess startet mit Doppelklick auf den Knoten und Klick auf „Login“.

2. GOOGLE-SHEETS-CONNECTOR:

Dieser Knoten wird mit dem ersten verbunden. Hier ist keine Konfiguration notwendig. Der Knoten muss ausschließlich über einen Rechtsklick auf „EXECUTE“ ausgeführt werden. Ändert sich die Ampel unterhalb des Knotens von Gelb auf Grün, kann weitergearbeitet werden.

3. GOOGLE-SHEETS-READER:

Hier muss in der Konfiguration das entsprechende Google-Sheet ausgewählt werden. Nach erfolgreicher Ausführung des Knotens kann mit den Daten des Google-Sheets gearbeitet werden.

Letzte Anpassungen zur Aufbereitung der Daten

Die Daten aus dem Google-Sheet müssen nun noch in die korrekten Datenformate umgewandelt werden. Aktuell erkennt KNIME die Werte der Spalten Klicks und Impressions als

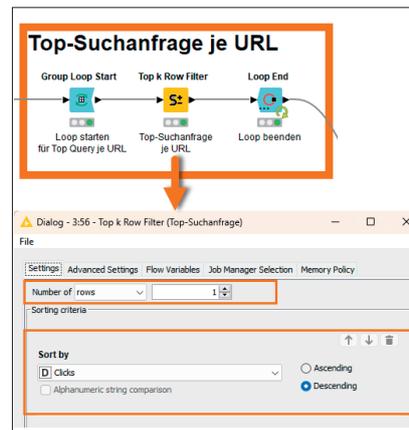


Abb.3: Konfiguration im Knoten TOP K ROW FILTER

TIPP

Ein kleiner Tipp zur Arbeit in KNIME für Einsteiger: Um einen Workflow zu erstellen, müssen die einzelnen Knoten immer über die Pfeile an der linken und rechten Seite verbunden werden. Solange die Knoten nicht verbunden sind, ist das durch einen roten Punkt unterhalb des Knotens gekennzeichnet. Sobald Knoten miteinander verbunden sind, ändert sich die Ampel auf Gelb. Sind die Knoten ausgeführt und können Daten hindurchfließen, ändert sich der Punkt auf die Farbe Grün. Nur so wird der Workflow ordnungsgemäß durchgeführt.

Zeichenketten (String). Für den weiteren Verlauf werden die Werte jedoch im Zahlenformat benötigt. Hierfür wird der Knoten STRING-TO-NUMBER in den Workflow integriert (Abbildung 2).

In der Konfiguration sind folgende Anpassungen durchzuführen:

- » *Includes*: Auswahl der Klicks und Impressions
- » *Decimal separator*: Punktzeichen eingeben
- » *Thousands separator*: Kommazichen eingeben
- » *Type*: Number (Double) auswählen

Mit der Auswahl „OK“ und dem Ausführen des Knotens über „EXECUTE“ ändert sich das Datenformat der Klicks und Impressions. Zur Überprüfung hilft die Vorschau der aktuellen Daten → Rechtsklick → „Transformed input“. In der Spalte Klicks und Impressions muss jetzt vor dem Namen der Spalten der Buchstabe „D“ für Double stehen. Zum Vergleich: Vor den Namen der anderen Spalten sollte weiterhin der Buchstabe „S“ für String stehen.

Erhebung der Top-Suchanfrage je URL

Um es später einfacher zu haben, wird auch hier wieder mit einem sogenannten Loop gestartet (Abbildung 3). Dazu sind folgende Knoten und Konfigurationen notwendig:

- » **GROUP LOOP START:** Dieser Knoten leitet immer einen Loop ein. Hier wird konfiguriert, mit welcher Spalte der Loop durchgeführt werden soll. Deshalb wird hier als Include „Page“ ausgewählt.
- » **TOP K ROW FILTER:** Mit diesem Knoten wird ein Ranking der Suchanfragen je Seite erzielt. Um die Top-Suchanfrage je Seite zu erhalten, wird als Number of rows „1“ ausgewählt. Im Feld Sort by wird „Clicks“ gewählt und mit der Auswahl „Descending“ absteigend sortiert (Abbildung 3).
- » **LOOP END:** Mit diesem Knoten wird der Loop beendet. Mit Ausführung des Knotens wird in der Oberfläche sichtbar, dass die drei Knoten so lange ausgeführt werden, bis der Vorgang für alle eindeutigen URLs abgearbeitet ist.



Abb.4: Zusammenführung der Daten im JOINER

hier kein vollständiger Crawl der Website notwendig ist, werden nur die URLs gecrawlt, die im letzten Knoten des Workflows, dem LOOP END, vorliegen.

Das funktioniert so:

1. Rechtsklick auf den Knoten LOOP END Auswahl „Collected results“
2. In der Vorschau der Tabelle auf die erste URL in der Page-Spalte klicken und mit der Tastenkombination „Strg + Shift“ und „ENDE“ alle URLs

Erhebung weiterer Daten zur Auswertung

Nun liegen alle Seiten vor, für die der Abgleich zwischen der Top-Suchanfrage und den Angaben in Title, Meta-Description und H1-Überschrift notwendig ist. Was fehlt, sind die Metadaten der Seiten. Um diese zu erhalten, wird auf den Screaming Frog als Crawling-Tool zurückgegriffen. Da

TIPP

Damit der Export nicht automatisch als Excel-Datei in eine unerwünschte Form gebracht wird, sollte als Type einfach „CSV“ gewählt werden. Somit läuft auch die Weiterarbeit in KNIME problemlos.

HINWEIS

Je nach Datensatz kann es hilfreich sein, mehr als nur die Top-Suchanfrage zu analysieren. Deshalb kann in der Konfiguration des TOP K ROW FILTER jederzeit die Number of rows erhöht und der Knoten neu ausgeführt werden. Zum Beispiel auf die Top-Three- oder Top-Five-Suchanfragen je Seite.

Ordnen & Aufräumen des Datensets

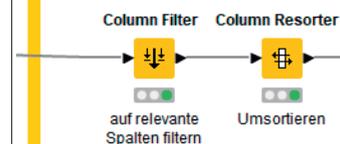


Abb.6: Datensatz filtern und Spalten umsortieren

- auswählen und mit „Strg + C“ in die Zwischenablage kopieren
3. Nun wird der Screaming Frog geöffnet.
4. Der Modus des Screaming Frog wird über Mode → List in den Listenmodus gewechselt. Über „Upload“ → Paste werden nun die URLs aus der Zwischenablage in das Tool importiert und gecrawlt.
5. Ist der Crawl beendet, wird das Ergebnis über Export als „internal_all“ exportiert.

Zusammenführung der Daten zum Abgleich

Um die GSC-Daten mit den Crawl-Daten zu vereinen, wird jetzt wieder in die KNIME-Umgebung gewechselt. Zunächst wird über den CSV READER der CSV-Export aus dem Screaming Frog eingelesen. Hierzu kann die CSV-Datei einfach mit Drag-and-drop auf die Oberfläche des Knotens CSV READER gezogen werden. Mit Rechtsklick auf „EXECUTE“ wird die Datei anschließend automatisch eingelesen.

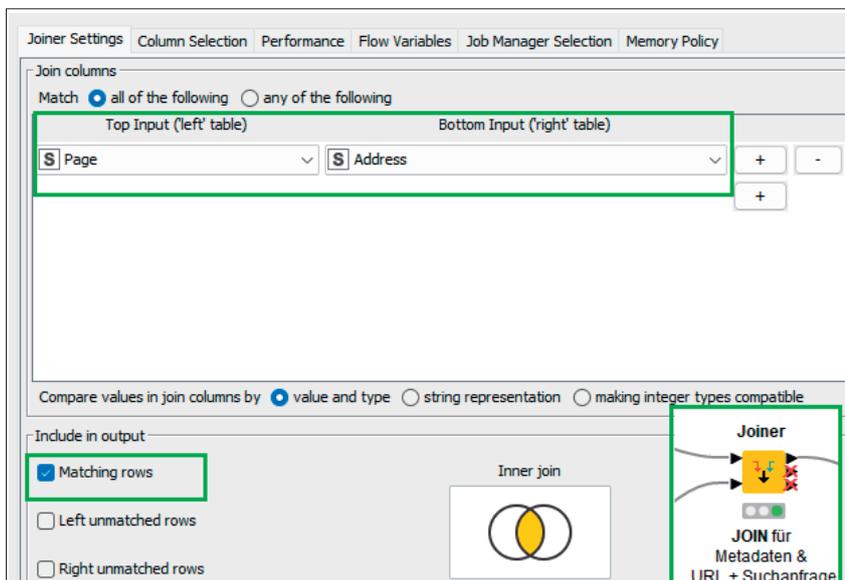
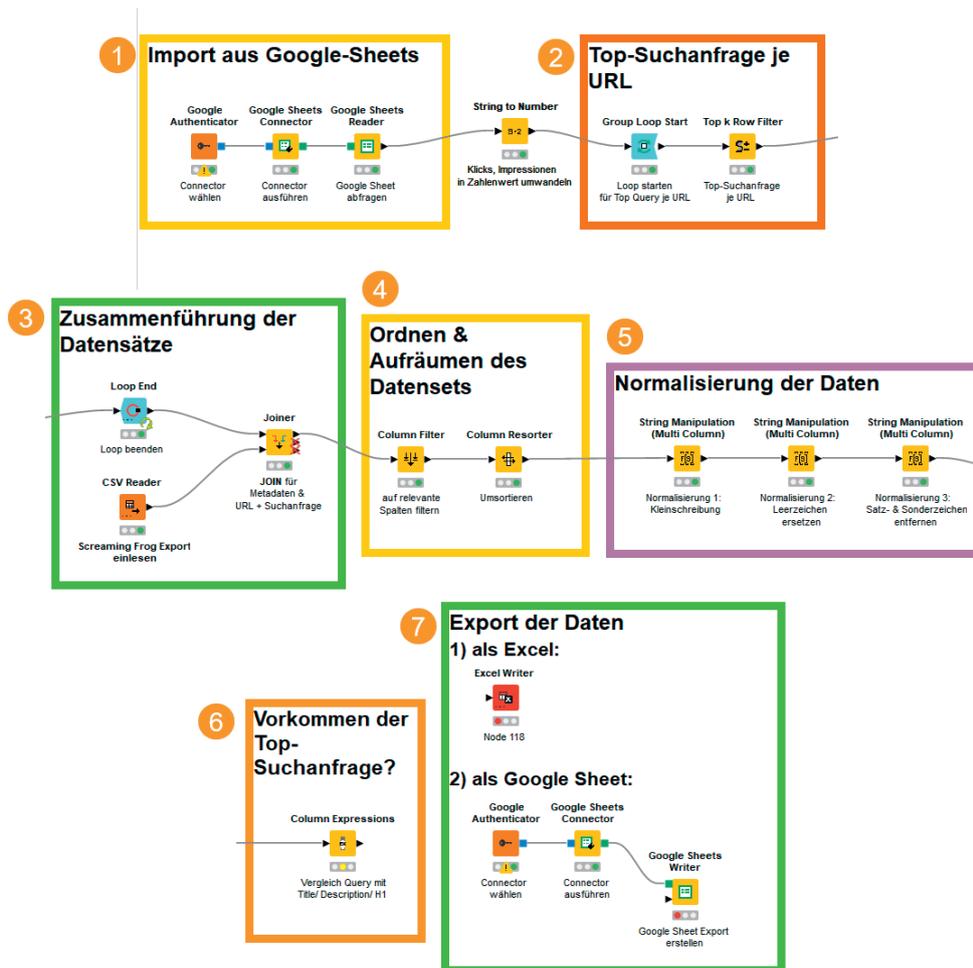


Abb.5: Konfiguration im Knoten JOINER

DER GESAMTE WORKFLOW SCHRITT FÜR SCHRITT IM ÜBERBLICK:



Zur Vereinigung der Daten kommt an dieser Stelle der JOINER zum Einsatz. Dieser Knoten funktioniert wie ein SVERWEIS in Excel. Weil der Knoten zum Zusammenführen der beiden Datensätze eine Referenzspalte braucht, hat der Knoten an der linken Seite zwei Einlese-Ports. Am oberen Port wird der Knoten LOOP END ange-bunden, am unteren Port der Knoten CSV READER (Abbildung 4).

In der Konfiguration werden nun die Spalten ausgewählt, anhand derer die Daten abgeglichen werden. Der Top-Input ist hier die „Page“-Spalte, der Bottom-Input ist die „Address“-Spalte. Für Include in Output wird „matching rows“ ausgewählt, um nur Spalten zu behalten, die im Crawl-Datenset und im GSC-Datenset vorkommen. Dabei handelt es sich um den sogenannten „Inner join“ (Abbildung 5).

Aufräumen als Zwischenschritt

Zur besseren Übersicht kann in den Workflow nun der Knoten COLUMN FILTER eingefügt werden, um nicht relevante neue Spalten, die aus dem Crawl-Export kommen, herauszufiltern. Denn für den weiteren Verlauf sind nur die folgenden Spalten notwendig: Query, Page, Clicks, Impressions, Title, Meta-Description und H1-Überschrift. Alle anderen Spalten können über Excludes ausgeschlossen werden. Optional kann außerdem der COLUMN RESORTER verwendet werden, um die Spalten der Tabelle zu vertauschen und damit eine bessere Übersicht zu erhalten (Abbildung 6).

Normalisierung der Daten zum Vergleich

Der Abgleich der Daten funktioniert über einen exakten Vergleich. Schon

Abweichungen in der Groß- und Kleinschreibung von Wörtern führen dazu, dass kein Matching stattfindet. Deshalb müssen die Spalten Title, Meta-Description und H1-Überschrift in die gleiche Form gebracht werden, in der die Suchanfragen vorliegen.

Durch den Prozess der sogenannten Normalisierung werden die Angaben wie folgt angepasst:

- » Kleinschreibung aller Wörter
- » Entfernung von Satz- und Sonderzeichen
- » Entfernung von mehreren Leerzeichen hintereinander

Für diese Schritte wird jeweils der Knoten STRING MANIPULATION verwendet. Weil die gleiche Normalisierung für mehrere Spalten notwendig ist, kommt hier der Knoten STRING MANIPULATION (MULTI COLUMN) zum Einsatz. Das spart Zeit und ist weniger fehleranfällig.

Allgemeine Konfiguration von STRING MANIPULATION (MULTI COLUMN)

Die Konfiguration des Knotens läuft für alle drei Normalisierungsschritte gleich ab. Nur die Formulierung der Expression unterscheidet sich. In Abbildung 7 ist zu sehen, welche Einstellungen vorgenommen werden müssen.

1. Auswahl aller Spalten in Includes, die normalisiert werden sollen. Dies sind hier: Title, Meta-Description und H1-Überschrift (Abbildung 7, Ziffer 1).
2. Auswahl der Function (Abbildung 7, Ziffer 2): Hier gibt es eine umfangreiche Liste mit verschiedenen Funktionen, um Spalteninhalte auf unterschiedliche Weise zu verändern. Auf der rechten Seite werden alle Funktionen mit einfachen Beispielen erklärt. Zudem wird angegeben, wie die Funktion konfiguriert werden muss: „lowerCase(str)“ heißt, dass sich in den Klammern ein String, also eine Zeichenkette, befinden muss.
3. Anschließend wird ausgewählt (Abbildung 7, Ziffer 3), auf welche Spalte die Funktion angewendet werden soll. Da es um mehrere Spalten geht, die gleichzeitig verändert werden sollen, wird aus der Column List „CURRENTCOLUMN“ mit Doppelklick ausgewählt, um alle Spalten auszuwählen, die in Includes gelistet sind.
4. In der Einstellung Result columns wird eingestellt, ob die Veränderung der Daten in eine neue Spalte geschrieben oder eine bestehende Spalte überschrieben werden sollen.
5. In einem letzten Schritt wird noch ein Suffix festgelegt, das an bestehende Spaltennamen angehängt wird.

Für den hier vorliegenden Use-Case wird, wie in Abbildung 8 zu sehen,

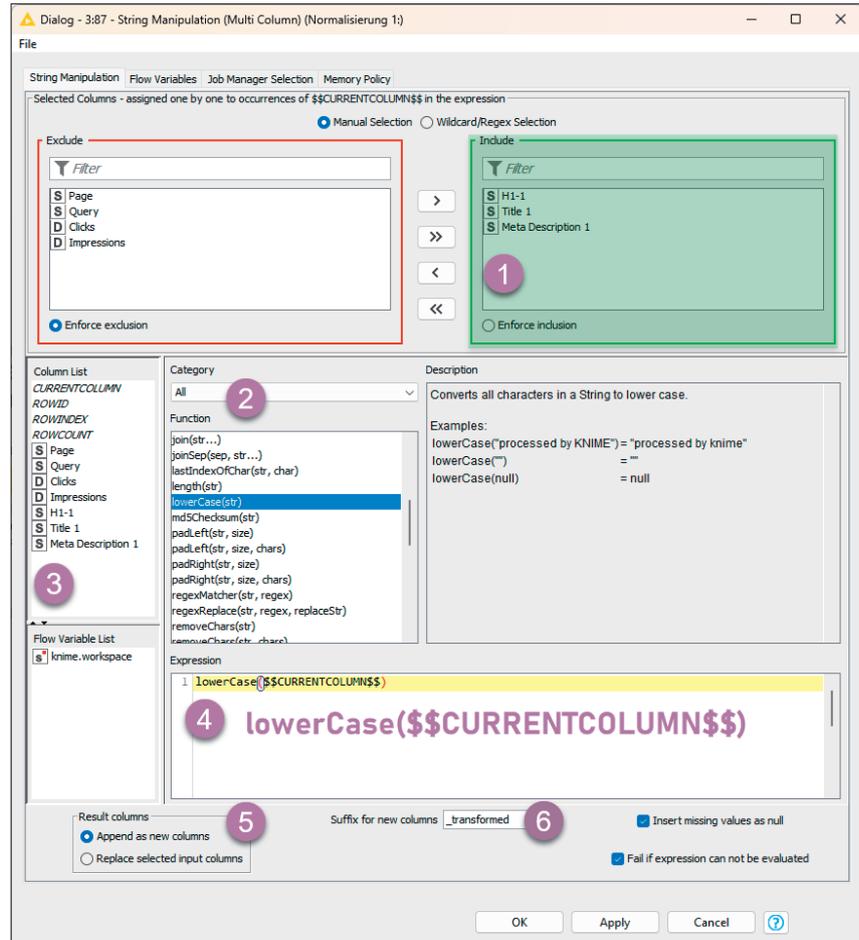


Abb.7: Konfiguration im Knoten STRING MANIPULATION (MULTI COLUMN)

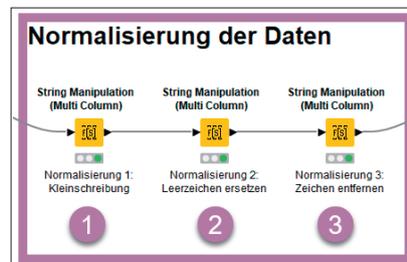


Abb. 8: Normalisierung der Daten in drei Schritten

dreimal der Knoten STRING MANIPULATION (MULTI COLUMN) benötigt.

Expressions für konkrete Normalisierung

1. Normalisierung: Kleinschreibung. Zur Kleinschreibung der Angaben der Spalten Title, Meta-Description und H1-Überschrift wird die Funktion „lowerCase()“ verwendet. Dazu genügt die Expression: lowerCase(\$\$CURRENTCOLUMN\$\$). Als Result column wird „append as new column“ gewählt und als Suffix für

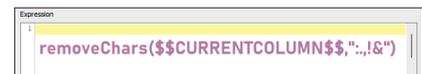


Abb. 9: Formulierung der Expression für Satz- und Sonderzeichen

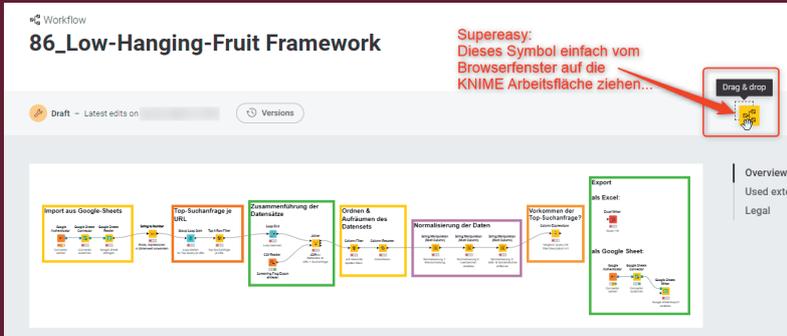
- die neuen Spalten „_transformed“. So kann später noch nachvollzogen werden, ob sich die Angaben der Spalten auch wie gewünscht verändert haben.
2. Normalisierung: Hierbei werden beim Auftreten mehrerer Leerzeichen alle bis auf ein Leerzeichen entfernt. Das ist wichtig, um später auch Such-

HINWEIS

Nach der Ausführung aller drei Knoten sollte mit Rechtsklick → „Output data“ überprüft werden, ob in der Vorschau noch andere Zeichen sichtbar sind. Fallen nun noch weitere Zeichen auf? Wenn ja, können diese einfach noch in den dritten Knoten aufgenommen werden, der dann erneut ausgeführt wird.

INFO

Der gesamte Workflow mit allen Konfigurationen ist kostenlos unter einfach.st/knime86 im KNIME-Hub abrufbar und kann ganz einfach per Drag-and-drop in die eigene KNIME-Umgebung gezogen werden.



anfragen, die aus mehreren Wörtern bestehen, tatsächlich vergleichen zu können. Die Expression lautet: `removeDuplicates($$CURRENTCOLUMN$$)`. Die Einstellung Result columns wird nun auf „Replace selected input columns“ gesetzt, damit alle Normalisierungsschritte in

nur einer neuen Spalte pro Angabe angehängt werden. Ansonsten werden alle Einstellung wie im ersten Knoten durchgeführt.

3. Normalisierung: Entfernung von Satz- und Sonderzeichen. Auch diese Zeichen behindern eher den Abgleich mit der Suchanfrage und

werden deshalb ersatzlos entfernt. Die Expression lautet hierfür: `removeChars($$CURRENTCOLUMN$$,“.:!&“)`. Der erste Teil in den Klammern beschreibt wieder die Auswahl der drei Spalten. Nach dem Komma werden dann alle Zeichen eingetragen, die im Datensatz vorkommen und für den Vergleich hinderlich sind. Wichtig: Die Zeichen müssen mit Anführungszeichen umrahmt werden. Ansonsten bleibt die Konfiguration wie im zweiten Knoten bestehen (Abbildung 9).

Ist die Suchanfrage enthalten?

Erst an diesem Punkt sind alle Daten für den Abgleich ready. Deshalb muss nun der Knoten COLUMN EXPRESSIONS konfiguriert werden. Der Knoten erlaubt es, mehrere Abfragezeilen innerhalb des Knotens zu erstellen. Die Konfiguration funktioniert so, wie in

WE LOVE BOOSTING

STUDENTEN-ABO*

51,- EUR

6 Ausgaben / Jahr
(Ausland: 63,- EUR)



www.websiteboosting.com/studentenabo

Bei Fragen: abo@websiteboosting.com

* auch für Schüler/Innen und Auszubildende (entsprechende Bescheinigung mitschicken!)

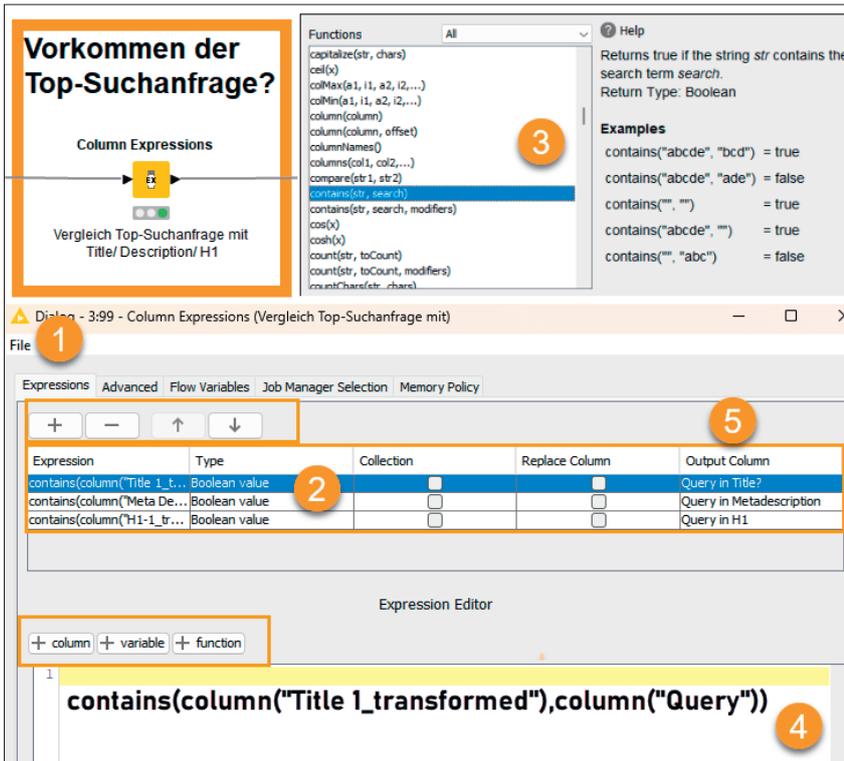


Abb. 10: Konfiguration im Knoten COLUMN EXPRESSIONS

Abbildung 10 zu sehen ist

- Über das Plus-Symbol wird eine neue Zeile ausgewählt (Abbildung 10, Ziffer 1).
- Jetzt muss ein Type gewählt werden (Abbildung 10, Ziffer 2). Dieser gibt an, wie die Auswertung bei Ausführung des Knotens erstellt wird. In diesem Fall wird „Boolean value“ gewählt. Das Ergebnis der Abfrage ist somit entweder „TRUE“ (= stimmt überein) oder „FALSE“ (= stimmt nicht überein).
- Anschließend wird die Expression über die Felder column, variable und function zusammengebaut. Für den gewünschten Vergleich wird als Funktion „contains“ ausgewählt. In der Auswahl (Abbildung 10, Ziffer 3) sind alle verfügbaren Funktionen gelistet und mit Beispielen erklärt.
- Die Anfrage, die in den Editor eingetragen werden muss, lautet: Ist in Spalte „Title 1_transformed“ die entsprechende Suchanfrage aus der Spalte „Query“ enthalten? Als Funktion ergibt das: contains(column(„Title 1_transformed“),column(„Query“)).

- Als Letztes wird über Output Column festgelegt, wie die Spalte heißen soll, in der die Auswertung nach TRUE oder FALSE enthalten ist. Hier: „Query in Title?“
- Der gleiche Ablauf wird nun noch für die Spalten Meta-Description und H1-Überschrift wiederholt. Dazu kann einfach die entsprechende Expression kopiert und für die anderen Zeilen eingefügt werden. Nur der Spaltenname muss für den Vergleich mit der Query angepasst werden. Zusätzlich wird in Output Column der Name der neuen Spalte angepasst. Mit der Ausführung des Knotens ergeben sich in den Daten drei neue Spalten, die jeweils auflisten, ob die Top-Suchanfrage in den vergleichenden Spalten enthalten ist (= TRUE) oder nicht enthalten ist (= FALSE).

Zu guter Letzt – der Export

Zuletzt genügt ein einfacher WRITER-Knoten, um die Daten zu exportieren. Mit dem EXCEL WRITER werden die Daten in eine XLSX-Datei exportiert. Aber auch ein Export in ein Google-Sheet ist möglich. Hierzu sind, wie

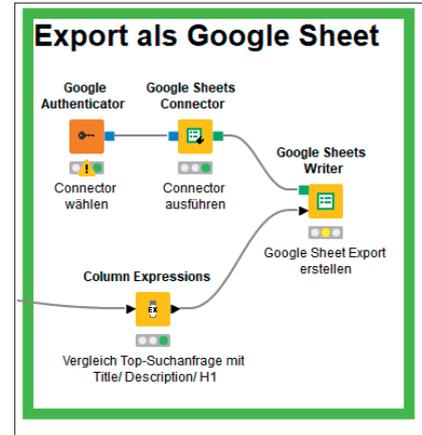


Abb. 11: Export der Daten als Google-Sheet

am Anfang, drei Google-Sheet-Knoten notwendig. Zuerst wird im GOOGLE AUTHENTICATOR wieder der Log-in im Chrome-Browser bestätigt und anschließend mit dem GOOGLE SHEETS CONNECTOR verknüpft. Daran wird der GOOGLE SHEETS WRITER angebunden, in dem nur ein Name für das zu erstellende Sheet unter Spreadsheet name vergeben wird. Der Knoten hat an der linken Seite zwei Ports. An den oberen Port wird, wie in Abbildung 11 zu sehen, der Knoten GOOGLE SHEETS CONNECTOR verbunden. An den unteren Port wird der Knoten COLUMN EXPRESSION angebunden. Mit dem Ausführen wird im gewählten Browser automatisch ein Sheet mit den Daten erstellt. FERTIG!

Fazit

Der vorgestellte Workflow zeigt, wie Daten aus verschiedenen Quellen vereinigt und aufbereitet werden können, um damit direkt weiterzuarbeiten. Durch die Automatisierung der immer gleichen Datenaufbereitung wird Zeit gespart, die dann zum Schreiben von noch besseren Titles, Descriptions und Überschriften zur Verfügung steht. In diesem Sinne viel Spaß beim Ausprobieren in KNIME! ¶