

Rebecca Schwarz

KNIME & ChatGPT:

Prompts automatisiert generieren und abfragen

Rebecca Schwarz zeigt in dieser Ausgabe auf, wie man mit ChatGPT direkt in KNIME arbeiten kann. Weil die reine Steuerung der KI außerhalb des Frontends aber nicht zwangsläufig eine Arbeitserleichterung darstellt, geht der Workflow noch einen Schritt weiter. Denn am Ende es geht darum, einen Prompt immer wieder mit neuen Informationen anzureichern. Damit ist es möglich, dass ChatGPT für mehrere unterschiedliche URLs individuelle Auswertungen gleichzeitig durchführt. Die URLs müssen dabei in keinem Zusammenhang stehen, da für jede URL ein eigener Prompt erzeugt und abgefragt wird.

Der Workflow zeigt beispielhaft, wie die KI die Seiteninhalte einer URL einschätzt. Dabei wird gefragt, ob die KI die Seite inhaltlich eher informational oder transaktional einordnet. Dies kann gerade für Preisvergleiche oder Kaufberatungen sehr spannende Insights bringen.

SEO-Tasks mit KNIME umsetzen

Bereits in den letzten Artikeln wurde immer wieder gezeigt, wie die Open-Source-Software KNIME verwendet werden kann, um verschiedenste Fragestellungen in der Suchmaschinenoptimierung zu lösen. Diesmal geht es um einen Anwendungsfall mit der Anbindung an ChatGPT. Neben der Vorstellung eines konkreten Anwendungsfalls soll der Workflow auch zeigen, wie einfach sich die Knoten anpassen lassen, um ganz eigene individuelle Fragestellungen zu analysieren.

Und das Beste: Alle hier vorgestellten Analysen können ab der Veröffentlichung des Artikels im KNIME-Hub kostenfrei per Drag-and-Drop in die eigene KNIME-Umgebung eingefügt werden. Somit müssen nur kleine individuelle Anpassungen vorgenommen werden und der Workflow kann schnell auf die eigenen Daten angewendet werden.

ChatGPT in KNIME verwenden

Nachfolgend sollen zwei Funktionalitäten in der Arbeit mit ChatGPT und KNIME erklärt werden:

1. Wie ChatGPT grundsätzlich aus der KNIME-Umgebung heraus angesteuert wird.
2. Wie in einem Prompt automatisiert immer wieder neue Informationen (hier: URLs) integriert werden können, um eine individuelle Analyse zu starten, ohne die erneute Abfrage händisch durchführen zu müssen. (Keine Angst, anhand des Beispiels wird es auch für KNIME-Einsteiger*innen sehr leicht!)

Für die Arbeit mit ChatGPT in KNIME wird Folgendes benötigt:

- » API-Schlüssel für OpenAI
- » OpenAI-Knoten aus dem KNIME Community Hub

Hinweis zu den verwendeten Knoten (Nodes): Nicht alle hier verwendeten Knoten

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.

sind Teil des Standard-Sets. Deshalb müssen einige dieser Knoten im Community-Hub von KNIME heruntergeladen werden. Im Community-Hub kann jeder Nutzende selbst entwickelte Knoten oder auch ganze Workflows zur Verfügung stellen. Diese Knoten lassen sich dann in die eigene KNIME-Umgebung einfügen.

Die Knoten der Community können über den Reiter „View“ im Menüband der Software und der Auswahl „KNIME Hub“ gesucht werden (Abbildung 1). Die entsprechenden Knoten können dann per Drag-and-Drop in die eigene Umgebung gezogen werden. Unter Umständen ist es notwendig, Erweiterungen zu installieren, um die Knoten verwenden zu können. Mit dem Einfügen der entsprechenden Knoten öffnet sich dafür automatisch ein neues Installationsfenster.

Sollten die nachfolgenden Knoten also nicht bereits in der eigenen Umgebung bekannt sein, können diese einfach über das KNIME-Hub gesucht und eingefügt werden.

Verbindung zu OpenAI in KNIME herstellen

Um mit dem Anwendungsfall zu starten, muss zunächst eine Verbindung zu OpenAI hergestellt werden. Der Workflow startet dafür mit dem Knoten CREDENTIALS CONFIGURATION (Abbildung 2, Ziffer 1). Mit Rechtsklick auf den Knoten und die Auswahl „Configure...“ öffnet sich die Konfiguration des Knotens. Hier muss im Feld „Password“ der eigene API-Key eingegeben werden. Ansonsten bleibt die Standardkonfiguration bestehen. (Abbildung 3)

Daran angehängt wird der Knoten OPENAI AUTHENTICATOR (Abbildung 2, Ziffer 2). Hier ist keine Konfiguration notwendig. Zuletzt wird der Knoten OPENAI CHAT MODEL CONNECTOR (Abbildung 2, Ziffer 3) an die beiden Knoten angehängt. In diesem Knoten wird das gewünschte Sprachmodell ausgewählt (Abbildung 4).

HINWEISE FÜR DEN KNIME EINSTIEG:

Hinweise für KNIME-Anfänger*innen: Die Software KNIME arbeitet mit sogenannten Knoten, die miteinander verbunden einen Workflow ergeben. Zu Beginn werden immer über einen READER-Knoten Daten in die Oberfläche eingelesen. In daran angebotenen Knoten werden die Daten entsprechend verändert, transformiert oder angereichert. Jeder Workflow endet normalerweise mit einem WRITER-Knoten, mit dem die Daten schlussendlich in eine Datei, zum Beispiel eine Excel-Datei, geschrieben werden. Anschließend kann mit dieser Datei dann im entsprechenden Programm ganz leicht weitergearbeitet werden.

Die größten Vorteile der Datensoftware im Allgemeinen sind, dass das Tool kostenfrei verwendbar ist und keine Programmierkenntnisse notwendig sind. Außerdem ist auch für ChatGPT die Software nicht unbekannt und kann bei Problemen mit der Konfiguration helfen. Auf knime.com gibt es umfassende Anleitungen und Dokumentationen zur Einführung ins Tool. Passende SEO-Anwendungsfälle waren und sind außerdem Bestandteil vieler Ausgaben der Website Boosting seit Ausgabe 53 und unter einfach.st/alleknime online abrufbar.

Do it: Eine einfache Analyse von Search-Console-Daten mit KNIME

Datensätze durch KNIME und alle Umformungen plus die Zusammenführung geht ab jetzt vollautomatisch und in Echtzeit. Bereits in Ausgabe 53 wurde das Thema KNIME aufgegriffen...

RESSORT: Content & Systeme AUTOR: Michael Hohenlechner JAHR: 2021 AUSGABE: 66

Beyond Crawling: Individuelle Klassifizierung Ihrer Webseiten nach Typ

der Rohdaten eines Crawls und KNIME schnell und einfach nachgeholt werden. Screaming-Frog-Daten in KNIME importieren Die Grundlagen von KNIME wurden bereits in den...

RESSORT: Web Controlling AUTOR: Michael Hohenlechner JAHR: 2021 AUSGABE: 71

Datenabfrage-Abfrage leicht gemacht: mit KNIME und der Sistrix API

Exzerte werden die Daten in KNIME importiert. Um schnell sehen zu können, wie die Daten in KNIME aussehen, aktivieren Sie im Menü von KNIME unter dem Punkt View den...

RESSORT: Web Controlling AUTOR: Michael Hohenlechner JAHR: 2021 AUSGABE: 77

Data Wrangling mit KNIME: Screaming-Frog-Crawls für Analysen aufbereiten

wesentlichen Vorteil von KNIME kann man anhand des Screenshots nicht erkennen, nämlich dass KNIME ohne Probleme mit sehr großen Datenmengen zurechtkommt. Da KNIME ursprünglich...

RESSORT: Content & Systeme AUTOR: Patrick Lürver JAHR: 2018 AUSGABE: 53

Beyond Crawling: Insights durch N-Gramm-Analysen (Teil 4)

bei Bedarf einfach vom KNIME Hub (URL im Beitrag) per Drag and Drop in die Arbeitsfläche von KNIME ziehen. Außerdem bietet KNIME die Möglichkeit...

RESSORT: Web Controlling AUTOR: Michael Hohenlechner JAHR: 2021 AUSGABE: 77

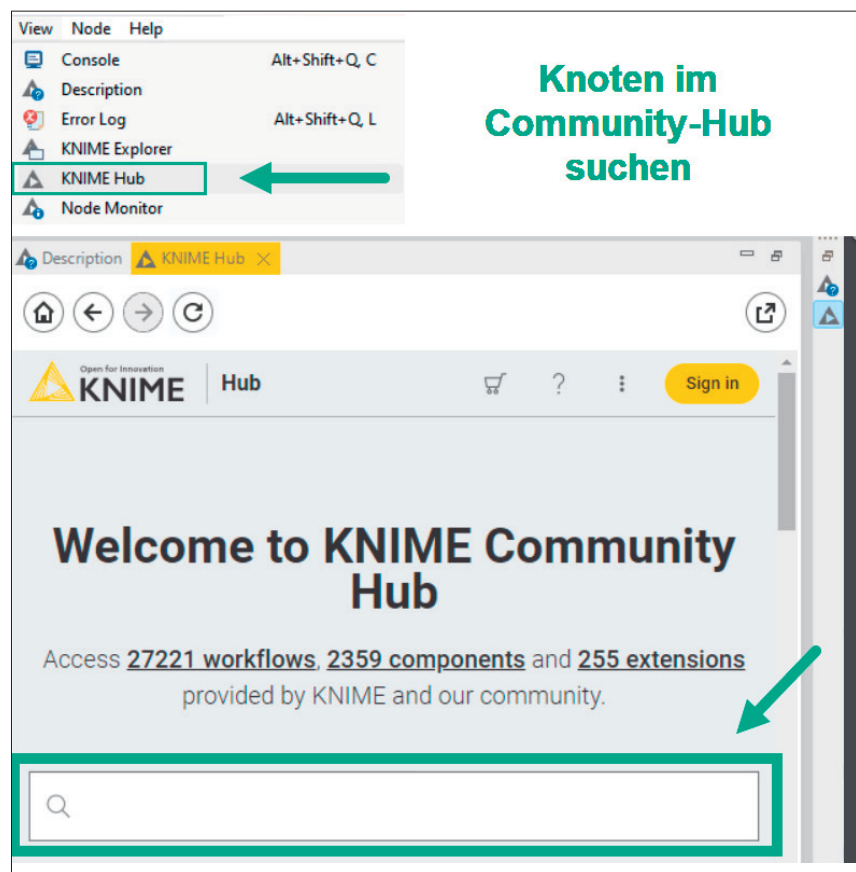


Abb. 1: Ansicht des KNIME-Hubs in eigener KNIME-Umgebung

ACHTUNG

Die Verwendung der OpenAI-API kostet je nach Anwendungsfall und Wahl des Sprachmodells unterschiedlich viele Credits und dementsprechend auch Geld. Um nicht unnötig Kosten zu verursachen, empfiehlt es sich, zunächst nur mit einem kleinen Testset an URLs zu arbeiten, um zu prüfen, ob das gewünschte Ergebnis herauskommt. Zusätzlich ist es sinnvoll, die formulierten Prompts zunächst im Frontend von OpenAI zu testen.

Nach der Konfiguration eines Knotens wird dieser immer über Rechtsklick → „Execute“ ausgeführt. Eine erfolgreiche Konfiguration ist anschließend an der Ampelanzeige am unteren Ende jedes Knotens ablesbar. Eine grüne Anzeige bedeutet, dass die Ausführung erfolgreich war, während eine rote Anzeige auf einen Fehler hinweist. Wird mit der Maus über die Anzeige gehovert, wird angezeigt, welcher Fehler sich gegebenenfalls in der Ausführung des Knotens ergeben hat.

Nachdem nun eine Verbindung via API-Key zu OpenAI besteht, werden nun eigene Daten in die Oberfläche geladen.

Der SEO-Anwendungsfall für OpenAI

Um einmal exemplarisch zu zeigen, wie sich ein Workflow aufbauen lässt, um einen oder auch mehrere Prompts in KNIME abzufragen, wird folgender Anwendungsfall gezeigt:

ChatGPT soll eine URL aufrufen und einordnen, ob der Inhalt eher transaktional oder informational ausgerichtet ist. Dazu wird ausschließlich eine URL im Prompt mit übergeben. Als Ergebnis soll ChatGPT ausschließlich „transaktional“ oder „informational“ ausgeben.

Dazu wird der Prompt wie folgt formuliert: „Rufe bitte die angegebene URL auf. Verschaffe dir einen Überblick über den Inhalt der Seite, ohne ihn

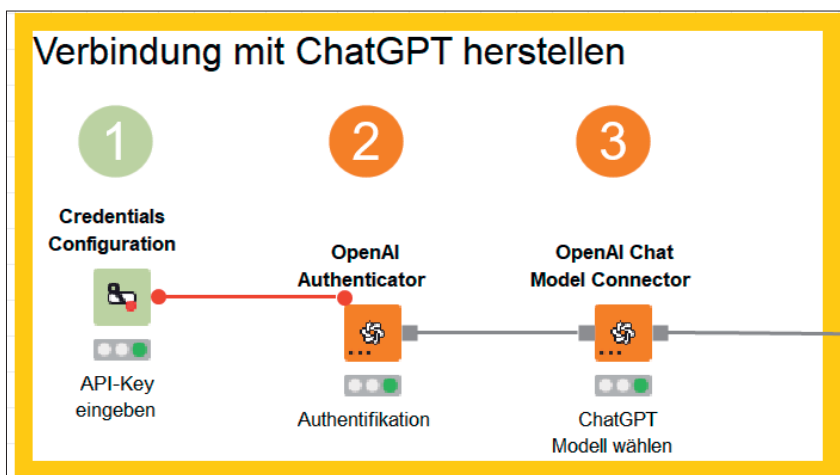


Abb. 2: Knoten-Set zur Verbindung zu OpenAI

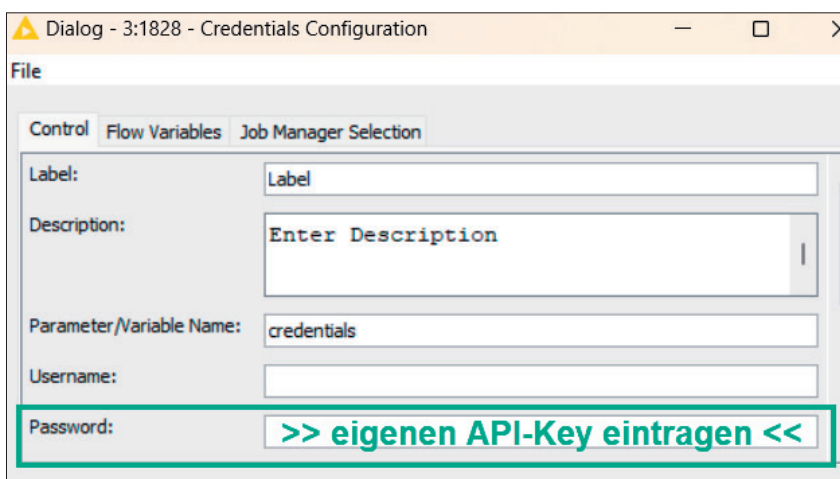


Abb. 3: Eingabe des einen API-Keys im Feld „Password“

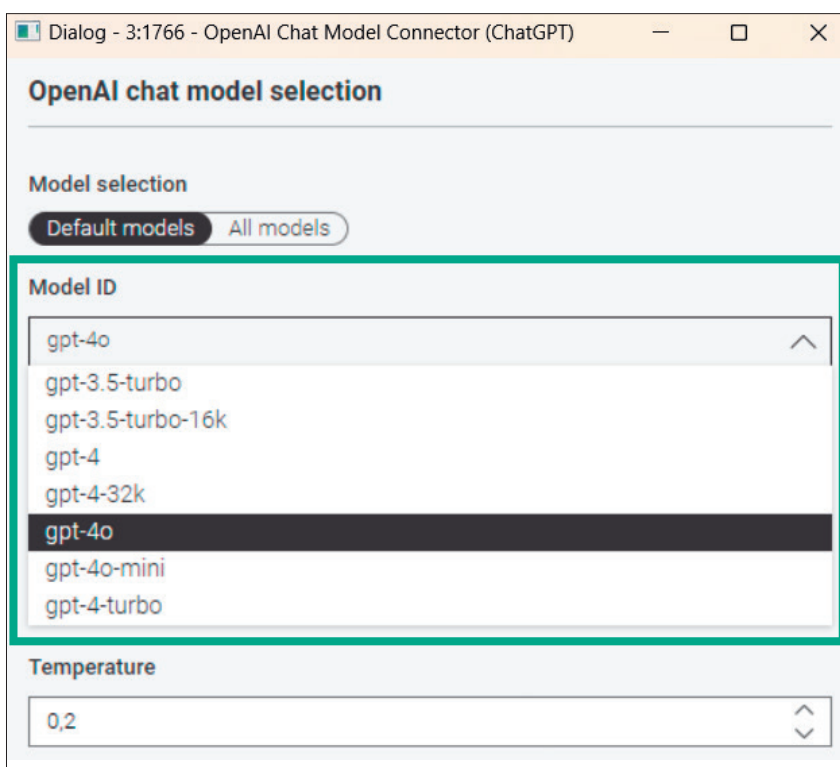


Abb. 4: Wahl des Sprachmodells im Knoten OPENAI CHAT MODEL

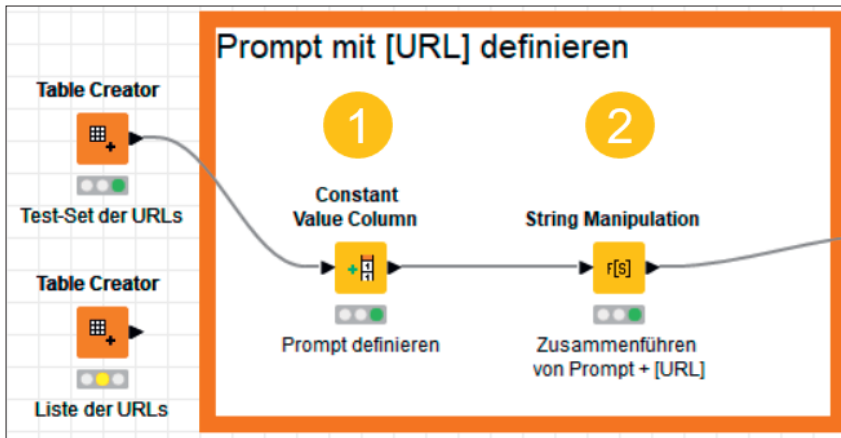


Abb. 5: Prompt mit URL definieren

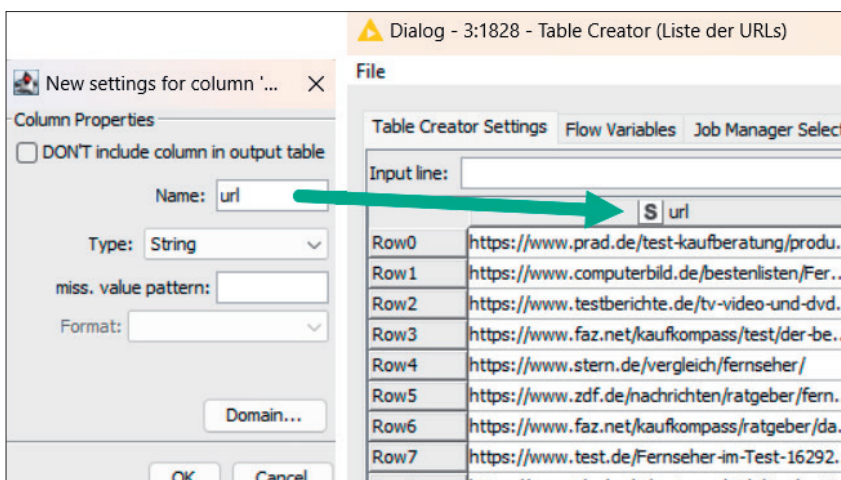


Abb. 6: Konfiguration im TABLE CREATOR

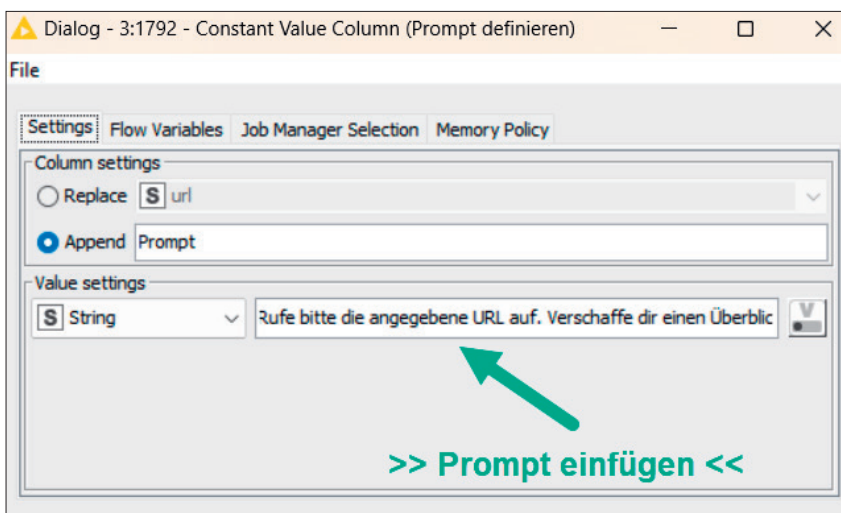


Abb. 7: Konfiguration im Knoten CONSTANT VALUE COLUMN

zusammenzufassen. Ordne den Inhalt eindeutig einer der folgenden Kategorien zu: Transaktional oder Information. Gib als Ergebnis ausschließlich das Wort ‚transaktional‘ oder ‚information‘ aus, ohne Begründung oder weitere Erklärung.“

Für die spätere Abfrage an ChatGPT wird nun zunächst der Prompt mit einer gegebenen URL verbunden. (Abb. 5)

Hierfür wird mit dem Knoten TABLE CREATOR gestartet. Praktisch daran ist, dass hier einfach URLs in den Knoten kopiert werden können.

Mit Doppelklick auf den TABLE CREATOR öffnet sich die Konfiguration. Hier können nun eine oder mehrere URLs hineinkopiert werden. Mit einem Doppelklick auf die Kopfzeile jeder Spalte wird ein Name der Spalte ausgewählt. Dieser ist wichtig, um später die Information weiterzugeben, in welcher Spalte sich die URLs befinden. Hier wird als Name „url“ gewählt. (Abbildung 6)

Mit der Auswahl „OK“ und einem Rechtsklick → „Execute“ auf den Knoten stehen die URLs nun zur Weiterverarbeitung in der Umgebung bereit. An den TABLE CREATOR wird jetzt der Knoten CONSTANT VALUE COLUMN angebunden (Abbildung 5 Punkt 1). In diesen wird der oben genannte Prompt eingefügt. Zusätzlich wird das Feld „Append“ ausgewählt und „Prompt“ eingetragen. Somit ist definiert, dass den Daten eine weitere Spalte hinzugefügt wird, die den Namen „Prompt“ trägt. (Abbildung 7)

Damit ChatGPT auch eine Analyse durchführen kann, wird dem Prompt nun noch eine URL mitgegeben. Dies wird im daran angeschlossenen Knoten STRING MANIPULATION erzielt. (Abbildung 5 Punkt 2) In der Konfiguration wird hier die Funktion „join(str...)“ verwendet, um die Spalte „url“ der Tabelle mit dem geschriebenen Prompt zu verbinden.

In der Expression wird dafür Folgendes eingetragen: join(„url: „,\$url\$,“ „,\$Prompt\$). Weil die vorherige Spalte „Prompt“ nicht mehr benötigt wird, wird die Spalte durch die Verbindung mit der URL überschrieben. In der Konfiguration der STRING MANIPULATION wird deshalb unterhalb Replace Column „Prompt“ ausgewählt. (Abbildung 8)

Zur Prüfung, ob das Datenset in KNIME den eigenen Vorstellungen entsprechend verändert wird, kann die Vorschau verwendet werden. Dazu wird zum Beispiel mit Rechtsklick auf die

STRING MANIPULATION geklickt und „Appended table“ ausgewählt. Nun öffnet sich ein Vorschaufenster.

Durch die Ausführung der STRING MANIPULATION gibt es nun für jede URL einen eigenen Prompt, der um die zugehörige URL ergänzt wurde. Befinden sich zehn URLs im TABLE CREATOR, sind durch die STRING MANIPULATION nun zehn Prompts entstanden, die durch die URL jeweils individuell sind.

All diese Prompts sollen dann nacheinander durch ChatGPT abgefragt werden.

Loop zur Abarbeitung der einzelnen Prompts

Um die Prompts nun abzuarbeiten, wird hier auf das sogenannte Loopen zurückgegriffen. Hierbei wird ein Workflow immer wieder wiederholt, bis keine weiteren Daten mehr vorhanden sind.

Konkret bedeutet das hier:

- » Der Loop startet mit dem Prompt der ersten Zeile und fragt diesen bei OpenAI ab.
- » Anschließend werden die Daten strukturiert und das Ergebnis der Auswertung wird in eine neue Spalte der Daten geschrieben.
- » Ist das abgeschlossen, wird der Prompt der zweiten Zeile abgefragt. Das passiert so lange, bis alle Prompts abgefragt sind. Dann wird der Loop beendet.

Für den Loop wird nun zuerst der Knoten GROUP LOOP START (Abbildung 9, Ziffer 1) benötigt. Dieser Knoten ist der Startpunkt des Loops. In der Konfiguration wird ausgewählt, welche Spalten der Daten benötigt werden. Hier werden einfach beide Spalten ausgewählt und in die Includes eingefügt.

Daran angeschlossen wird der Knoten CHAT MODEL PROMPTER (Abbildung 9, Ziffer 2). In diesem Knoten befindet sich in der Default-Einstellung einfach nur ein Fenster, in das ein

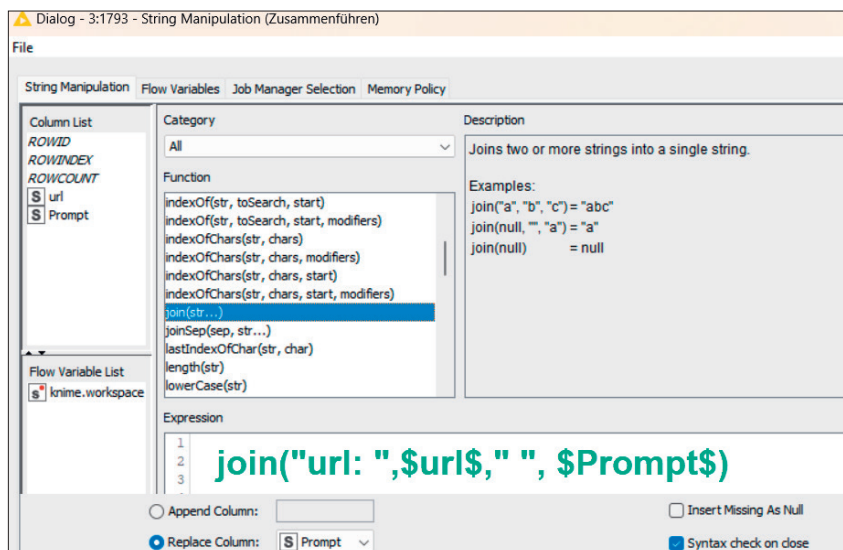


Abb. 8: Konfiguration der STRING MANIPULATION

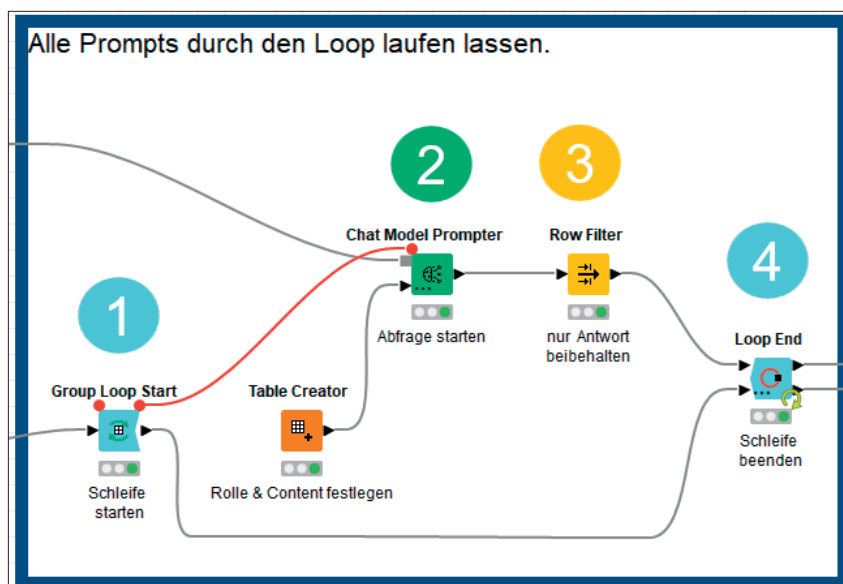


Abb. 9: Konfiguration des Loops

Prompt eingefügt werden kann. Das Besondere in diesem Workflow ist aber, dass mehrere Prompts nacheinander an diesem Knoten übergeben werden sollen. Deshalb muss definiert werden, wo sich diese Prompts befinden.

Dafür muss das Symbol am rechten Ende des Fensters New message angeklickt werden. Es öffnet sich anschließend ein neues Dropdown-Menü, mit dem gewählt wird, wie der Inhalt des Felds befüllt werden soll. Hier wird die Spalte der Daten „Prompt“ gewählt. (Abbildung 10)

Zusätzlich benötigt der Knoten eine Information zur Rolle und Rollenbeschreibung, die notwendig ist, um

Anfragen an die OpenAI API zu stellen. Diese Informationen müssen über eine Extradatei mitgegeben werden. Wie in Abbildung 9 zu sehen ist, wird dem CHAT MODEL PROMPTER ein TABLE CREATOR vorangestellt, um die fehlenden Informationen mitzuteilen. In dieser Tabelle müssen die Spalten „Rolle“ und „Content“ genannt werden. Als Rolle wird „system“ ausgewählt. Und als Content wird die Rollenbeschreibung definiert. Dies ist hier eine Beschreibung darüber, dass das System ein SEO-Experte ist und professionell einschätzen kann, ob ein Seiteninhalt eher transaktional oder informational gestaltet ist.

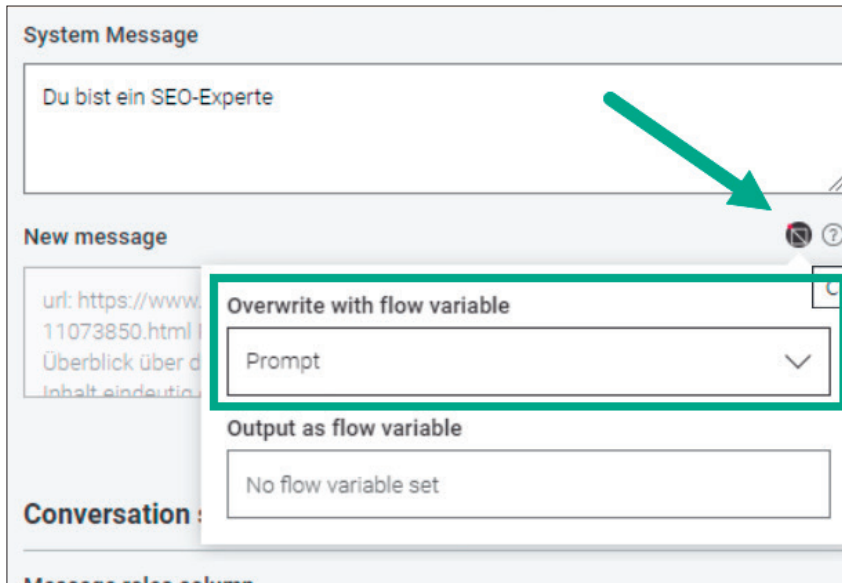


Abb. 10: Konfiguration des CHAT MODEL PROMPTER

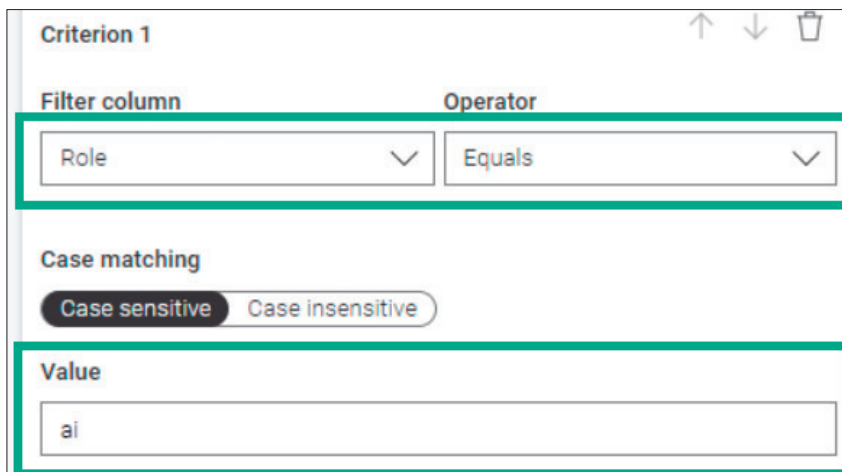


Abb. 11: Konfiguration im ROW FILTER

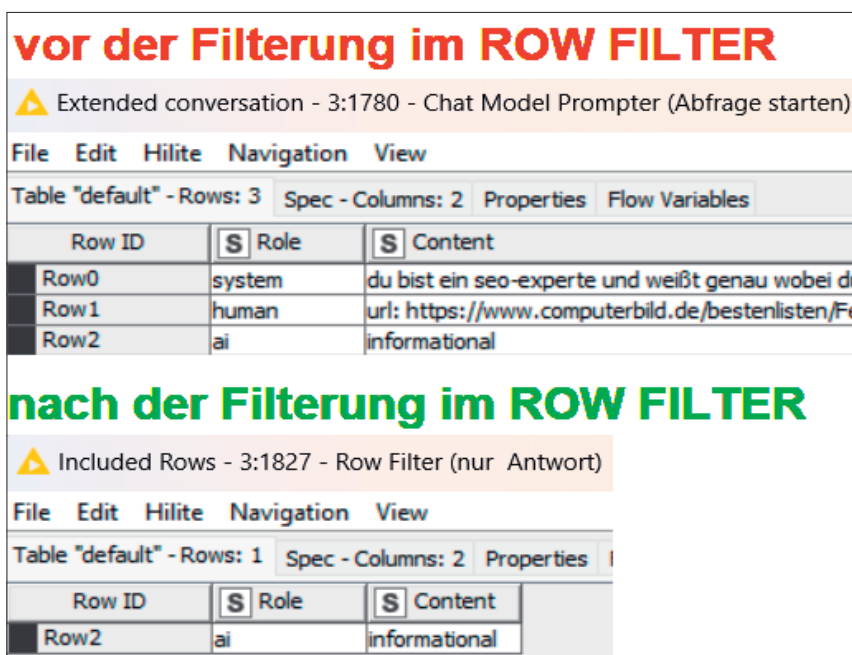


Abb. 12: Datenset nach der Filterung mit dem ROW FILTER

Der TABLE CREATOR wird an den unteren Port des CHAT MODEL PROMPTERs angehängt.

Bisher ist der Knoten GROUP LOOP START noch nicht mit dem CHAT MODEL PROMPTER verbunden. Dies wird hier auch nicht über das gewöhnliche Pfeilsymbol gemacht, sondern über die rote Linie, wie in Abbildung 9 zu sehen ist. Damit das klappt, wird mit Rechtsklick auf den GROUP LOOP START „Show Flow Variable Ports“ ausgewählt. Nun erscheinen zwei rote Punkte über dem Knoten. Der rechte Punkt wird auf den CHAT MODEL PROMPTER gezogen.

Nun kann der nächste Knoten angeschlossen werden, der ROW FILTER. Hierin wird einfach ausgewählt, dass nur die Zeilen erhalten bleiben sollen, die in der Spalte Role den Wert „ai“ aufweisen und somit die Antwort von ChatGPT enthalten (Abbildung 11), da ansonsten pro Prompt drei neue Zeilen entstehen. (Abbildung 12)

Da der CHAT MODEL PROMPTER etwas unglücklich in seiner Funktionalität ist und die hauptsächlichen Informationen, URL und Prompt, verloren gehen, wird gleich auf einen Trick zurückgegriffen.

Jetzt kann der Loop beendet werden. Das wird immer mit dem Knoten LOOP END erzielt. In der Konfiguration ist Folgendes zu tun:

- » Erst wird der LOOP END mit dem ROW FILTER verbunden.
- » Dann wird mit Doppelklick und der Auswahl „Add Collector Port“ ein weiterer Eingangsport hinzugefügt. Somit können weitere Daten im Kno-

TIPP

Um nun zu testen, ob alles korrekt läuft, ohne gleich alle Prompts über die API abzufragen, ist es sinnvoll, dies wieder erst anhand eines kleinen Testsets zu probieren. Das spart nicht nur Geld, sondern auch Zeit.

ten gleichzeitig übergeben werden. Im oberen Port befinden sich die Daten aus dem Loop, im unteren Port wird der GROUP LOOP START verbunden, wie in Abbildung 9 zu sehen ist. Somit werden die im Loop verlorenen URLs und Prompts wieder ins Dataset integriert.

- » Der Trick ist jetzt, dass im LOOP END die Auswahl „Generate new RowIDs“ gewählt wird. Das hat zur Folge, dass die Row-ID überschrieben und neu von vorne angefangen wird zu zählen. In der Informatik ist das bei 0, nicht bei 1.
- » Nach Bestätigung der Konfiguration im LOOP END mit „OK“, ist der Loop bereit. Mit Rechtsklick → „Execute“ startet nun der Loop. Dieser wird so lange durchgeführt, bis alle Prompts abgearbeitet sind.

Verknüpfung der Daten und Aufbereitung des Datensets

An den Loop wird nun der Knoten JOINER gehängt. Mit diesem können Daten miteinander vereint werden. Der Knoten ist das Äquivalent zum bekannten SVERWEIS() in Excel. Damit beide Datensätze integriert werden können, wird vom LOOP END aus der obere und untere Ausgangsport mit dem oberen und unteren Eingangsport des JOINERS verbunden. (Abbildung 13)

Hier wird nun klar, warum die Row-IDs neu vergeben werden mussten. Denn für die Verknüpfung ist immer eine Referenzspalte notwendig. Da sich aber in einem Dataset die Daten aus ChatGPT und sich im zweiten Dataset nur der Prompt und die URL befinden, ist ein Matching der Datensätze nur durch die Row-ID möglich.

Es werden im JOINER nun die Row-IDs gematcht. In der Konfiguration des JOINERS wird daher im linken und rechten Input die Spalte Row-ID gewählt und matching rows als Methode verwendet. (Abbildung 14)

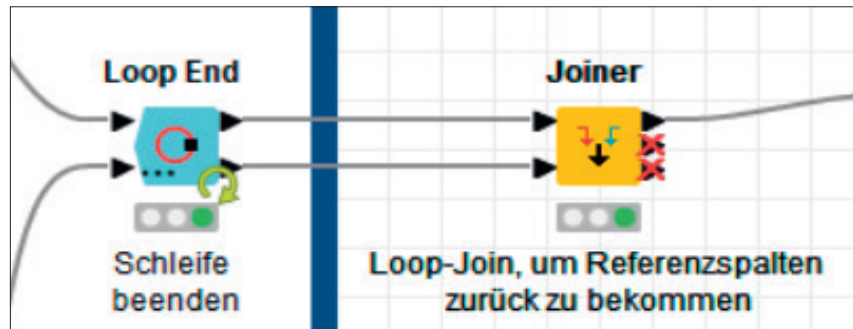


Abb. 13: Verbindung mit dem JOINER

Abb. 14: Konfiguration im JOINER



Abb. 15: Aufbereitung des späteren Exports

Auswertung vorbereiten und exportieren

Nun sind bereits alle Daten vorhanden. Im letzten Schritt wird mit mehreren Knoten noch ein übersichtlicher Export erzielt. (Abbildung 15) Mit dem Knoten COLUMN FILTER werden

alle Spalten herausgefiltert, die nicht mehr relevant sind. Es bleiben nur die Prompt-Antwort und die jeweilige URL bestehen. Mit dem COLUMN RESORTER können die Spalten noch anders angeordnet werden. Mit dem Knoten COLUMN RENAMER kann schlussend-

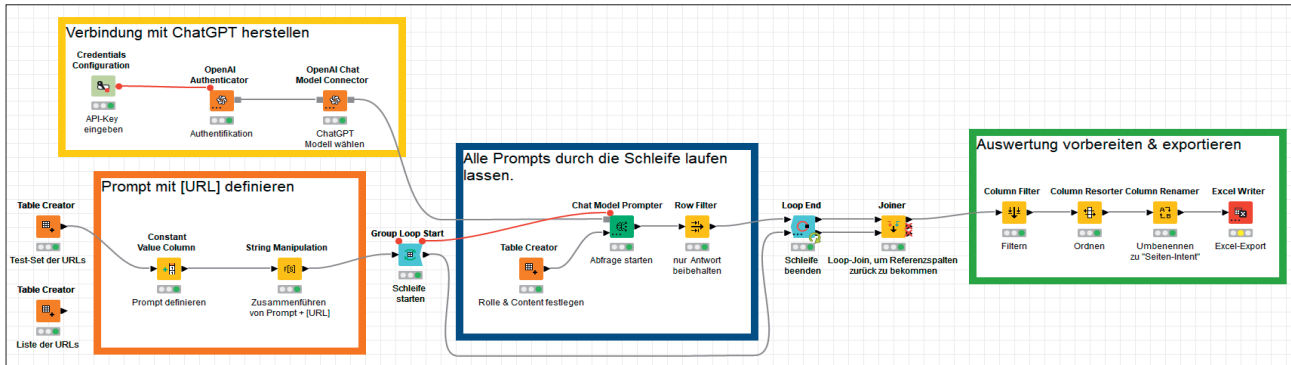


Abb. 17: Gesamter KNIME-Workflow

lich noch der Spaltenname geändert werden. Damit es hier eindeutiger wird, wird die Prompt-Antwort als „Seiten-Intent“ umbenannt.

Zu guter Letzt kann ein WRITER-KNOTEN angehängt werden, um die Daten aus KNIME in eine neue Datei zu schreiben. Im EXCEL-WRITER muss dazu nur ein Speicherort ausgewählt werden. Mit Ausführung des Knotens wird am Speicherort eine Excel-Datei mit dem Datenset erstellt.

Fazit

Der Workflow erscheint zunächst etwas komplex. Umso häufiger KNIME jedoch verwendet wird, um so schneller wird deutlich, dass bereits ein kleines Set an Knoten ausreicht, um unzählige Anwendungsfälle zu erstellen. Denn alle hier vorgestellten Workflows sollen nicht nur konkrete Anwendungsfälle zeigen, sondern vielmehr die Methoden näherbringen, um eigene Analysen durchzuführen. Der fertige Workflow

ist wie immer im KNIME-Hub unter einfach.st/knime86 verfügbar und per Drag-and-Drop in die eigene KNIME-Oberfläche einlesbar. Viel Spaß beim Ausprobieren!

Bei Fragen und Anmerkungen zum Workflow kann gerne ein Austausch via E-Mail oder LinkedIn stattfinden. ¶

PRÄMIE 1

- **Rucksack**
- 100% RPET, hergestellt aus 18 recycelten PET-Flaschen.
- Seesack-Design, große Öffnung.
- mit separatem Laptopfach und Getränkefach
- extra viel Platz für Website Boosting



PRÄMIE 2

- **Edelstahl Isolier Trinkflasche**
- Edelstahl, Doppelwandig, vakuumisoliert
- Füllmenge: ca. 590 ml

ABO

+ WAHLPRÄMIE

6 Ausgaben zum Vorteilspreis (62,- EUR* / Jahr)
Prämie: Website Boosting Rucksack oder Trinkflasche direkt zu Ihnen nach Hause geliefert, ohne zusätzliche Zustellkosten
immer ein paar Tage früher informiert

www.websiteboosting.com/abo

* Sie sparen 8,80 EUR im Vergleich zum Einzelkauf.
 Im Ausland: EUR 70,80 inkl. Versandkosten