



BESSER INTERN VERLINKEN MIT SCREAMING FROG UND KNIME

Michael Hohenleitner

Sie glauben, Daten über APIs abzufragen, ist nur etwas für Programmierer? Weit gefehlt. Mit KNIME lassen sich APIs ganz einfach und ohne Programmierung per POST Request abfragen und in eine Tabelle integrieren, bei voller Flexibilität. Erfahren Sie in diesem Beitrag, wie einfach die Sistrix-API mit KNIME abgefragt werden kann und wie Sie dabei wertvolle Zeit sparen.

Der Sistrix-Sichtbarkeitsindex ist eine in der SEO-Branche weitverbreitete Metrik und er ist für viele die erste Wahl, wenn es darum geht, die Reichweite einer Website in der Google-Suche zu prüfen. Neben dem reinen Sichtbarkeitsindex stellt Sistrix zahlreiche weitere Metriken zur Verfügung, die für einen Audit der eigenen oder einer fremden Website interessant sind. So lässt sich sowohl auf Domain- als auch auf URL-Ebene prüfen, für wie viele Keywords eine Domain in den Top 100 oder den Top 10 rankt.

Die meisten SEOs nutzen für diese Analysen das Webinterface von Sistrix. Dabei stellt der Anbieter auch eine einfach zu bedienende API zur Verfügung. Mit dieser lassen sich Abfragen, die sonst mühevoll einzeln durchgeführt werden

müssten, automatisieren. Mit dem kostenlosen Datenanalysetool KNIME lässt sich in wenigen Schritten und ohne Programmierkenntnisse ein Workflow erstellen, der für eine beliebige Menge URLs Daten per Sistrix-API ermittelt.

URL-Liste auf Top-10-Rankings prüfen

Angenommen, Sie möchten eine Liste mit URLs (beispielsweise aus einem Screaming Frog Crawl) auf die Anzahl der Top-10-Rankings mithilfe der Sistrix-API prüfen. Dazu benötigen Sie eine KNIME-Installation, einen Zugang zu Sistrix mit mindestens einem aktiven Modul und eine CSV-Datei mit den URLs, die geprüft werden sollen.

Foto: ThomasVogel / gettyimages.de

DER AUTOR



Michael Hohenleitner arbeitet gerne mit Rohdaten, um diese in maßgeschneiderten Analysen für seine Kunden aufzubereiten.

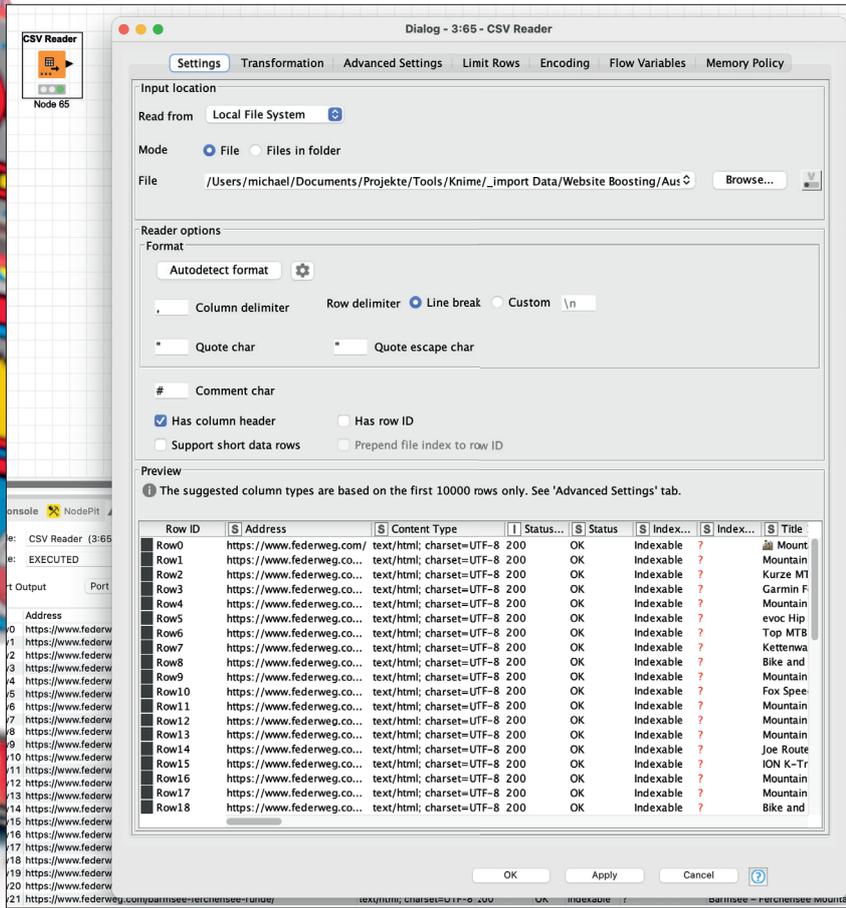


Abb. 1: Einstellungen der Node CSV-Reader in KNIME

SISTRIX CREDITS

Die Sistrix-API funktioniert über ein Credit-System. Für jedes bezahlte Sistrix-Modul werden pro Woche 10.000 Credits für API-Abfragen zur Verfügung gestellt. Wie viele Credits pro Abfrage benötigt werden, hängt von der Anzahl der Daten ab, die erhoben werden. Dies ist in der Beschreibung der einzelnen Abfragemethoden in der Dokumentation unter <https://www.sistrix.de/api/> beschrieben.

Erstellen Sie in KNIME einen neuen Workflow und ziehen Sie die CSV-Datei per Drag-and-drop auf die Arbeitsfläche. KNIME erstellt nun automatisch eine Node vom Typ CSV-Reader auf ihrer Arbeitsfläche und es öffnet das Konfigurationsmenü (Abb. 1).

Dort sehen Sie eine Vorschau, wie die Daten in KNIME importiert werden. Die Software ermittelt dabei selbst, wie die CSV-Datei aufgebaut ist, weshalb meistens keine Anpassungen in den Einstellungen nötig sind. Sehen die Daten im Bereich Preview gut aus, kann das Menü geschlossen und die Node ausgeführt werden. Durch einen Rechtsklick auf die Node und den Befehl Execute 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 Node Monitor. Nun sehen Sie unterhalb der Arbeitsfläche beim Klicken auf eine Node stets eine Vorschau der verarbeiteten Daten.

Im nächsten Schritt wird an die importierte Tabelle mit den URLs eine weitere Spalte angehängt, die die URL enthält, mit der die Sistrix-API abgefragt wird. Ziehen Sie dazu eine Node mit der Bezeichnung Constant Value Column auf die Arbeitsfläche und verbinden Sie diese mit der Node CSV-Reader, indem Sie die schwarzen Drei-

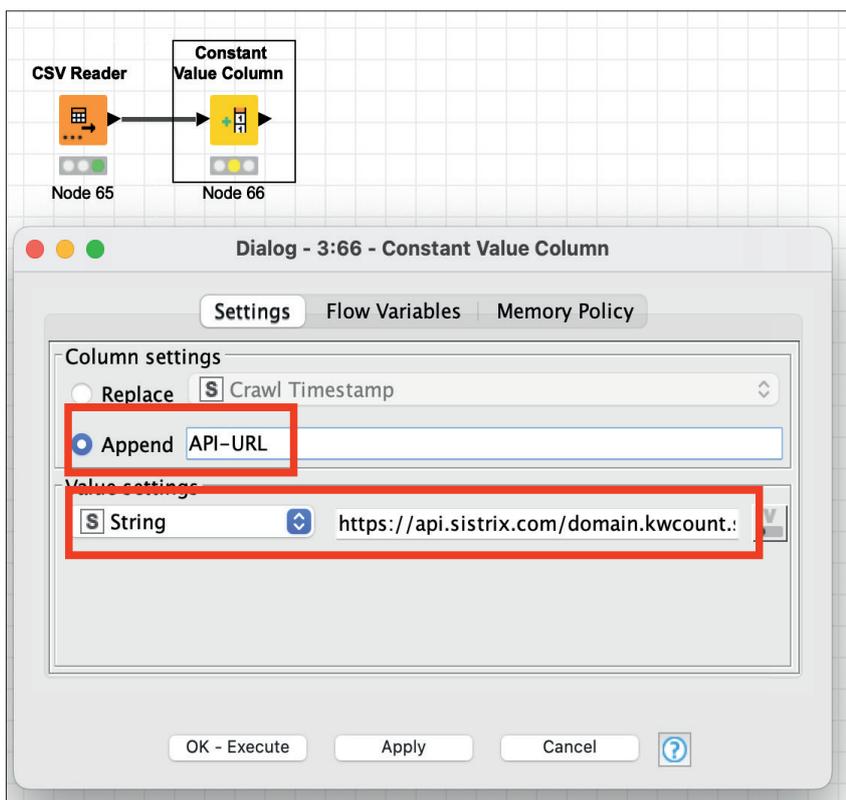


Abb. 2: Einstellungen der Node Constant Value Column

ecke per Drag-and-drop miteinander verbinden. Mit dieser Node können Sie eine neue Spalte erzeugen (oder eine bestehende überschreiben), die einen konstanten, von Ihnen vorgegebenen Wert enthält.

Die Node Constant Value Column wird nun per Doppelklick wie folgt konfiguriert (Abb. 2):

- » Column Settings: Wählen Sie Append und geben Sie in das Feld einen Namen für die neue Spalte, z. B. API-URL, an.
- » Value Settings: Das Auswahlmü bleibt auf „String“ und in das Feld wird nun die URL eingetragen, die die Daten von Sistrix ermittelt. Diese ist wie folgt aufgebaut:

https://api.sistrix.com/abfragemethode?api_key=[IHR_API_KEY]&url.

Die URLs, für die die Daten abgefragt werden sollen, sind bereits mit der CSV-Datei in KNIME importiert worden und sie werden im nächsten Schritt an die API-URL angehängt. Deshalb reicht es zunächst, die Abfragemethode und den API-Key anzugeben. Um mehr Daten zu erhalten und diese einfacher verarbeiten zu können, sollten die Parameter `&extended=true` (mehr Daten) und `&format=json` (gibt die Daten im JSON-Format zurück) in die Abfrage-URL aufgenommen werden. Eine finale Abfrage-URL sollte also wie folgt aussehen:

https://api.sistrix.com/domain.kwcount.seo.top10?api_key=[IHR_API_KEY]&extended=true&format=json&url=.

Um die Daten abzufragen, benötigen Sie ihren persönlichen API-Schlüssel, den Sie unter <https://app.sistrix.com/account/api> kopieren können und den Sie an der Stelle [IHR_API_KEY] einfügen.

Bevor die API-URL und die URL, für die die Rankings ermittelt werden sollen, verbunden werden können, muss

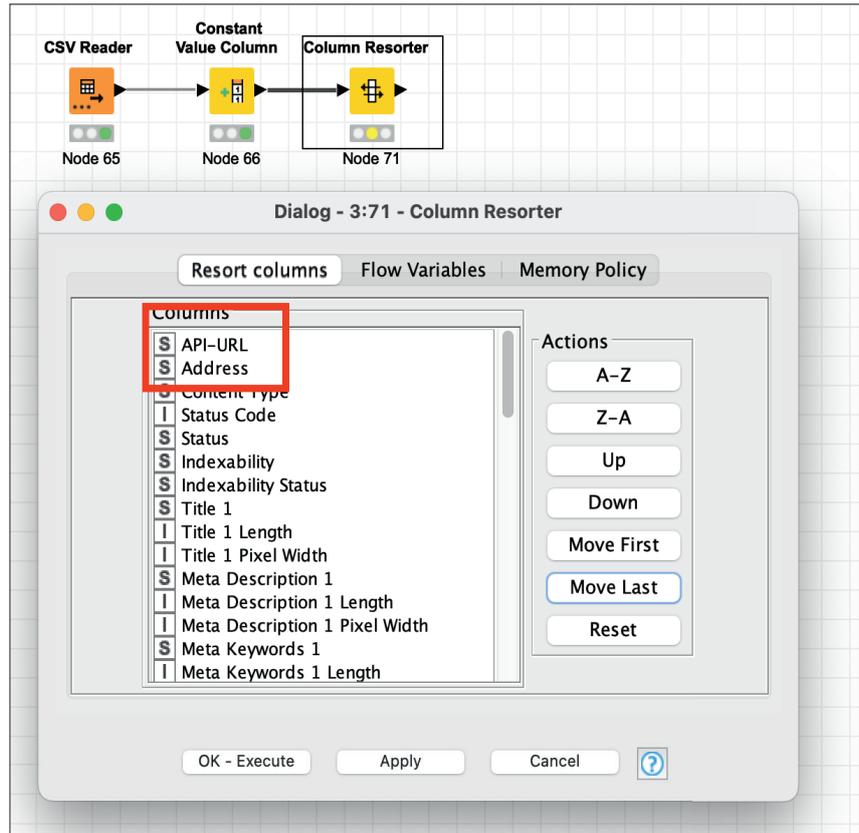


Abb. 3: Sortierung der Spalten anpassen, bevor sie verbunden werden

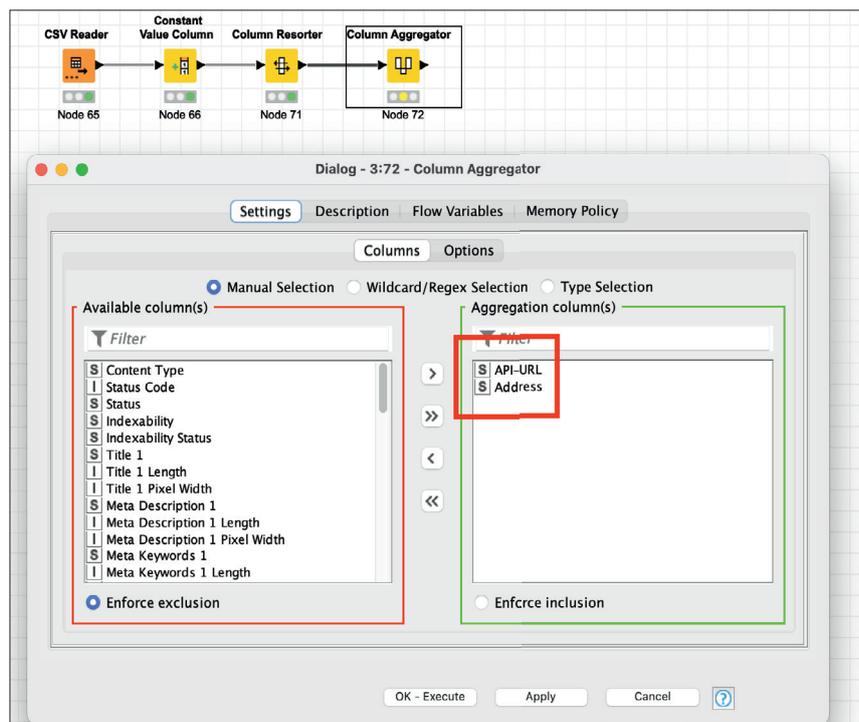


Abb. 4: Aggregation der Spalten API-URL und Address

zunächst mit der Node Column Resorter die Reihenfolge der Spalten angepasst werden. Wichtig hierbei ist, dass die Spalte API-URL vor der Spalte mit der abzufragenden URL steht (Abb. 3).

Abfrage-URLs für die API automatisch erstellen

Nun können beide Spalte mit der Node Column Aggregator aggregiert werden. Verschieben Sie dabei in der

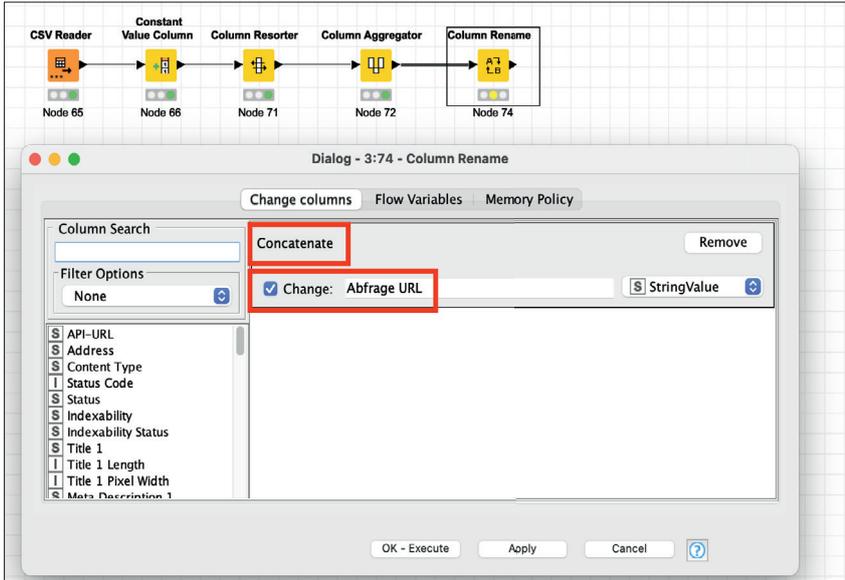


Abb. 5: Umbenennen der Spalte „Concatenate“

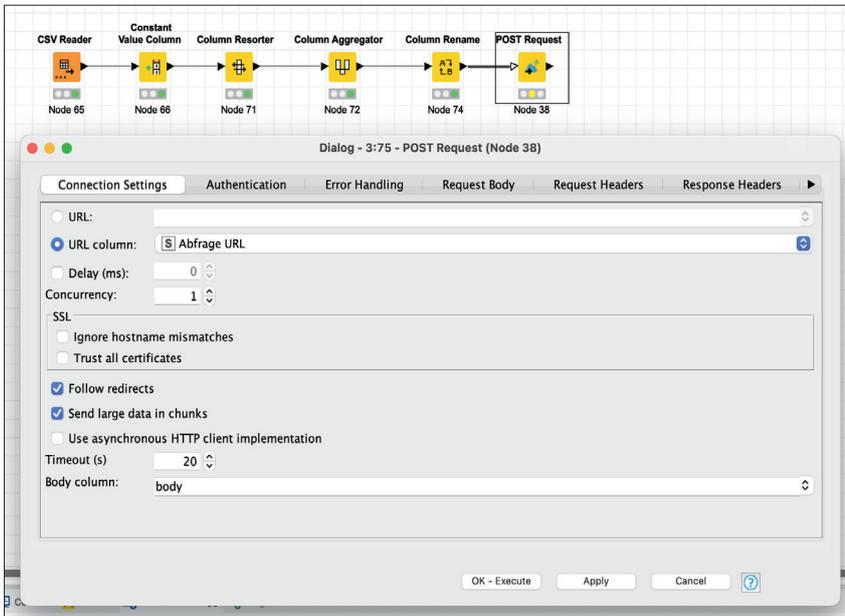


Abb. 6: Konfiguration der Node POST Request

Console NodePit Node Monitor Progress

Node: JSON to Table (3:76)
State: EXECUTED

Port Output Port 0 Load data Rows: 119, Columns: 66

url	date	value	used
o.top10] https://www.federweg.com/	2022-10-31T00:00:00+01:00	15	
o.top10] https://www.federweg.com/mountainbiken-lenggries/	2022-10-31T00:00:00+01:00	16	
o.top10] https://www.federweg.com/brenner-grenzkamm/	2022-10-31T00:00:00+01:00	0	
o.top10] https://www.federweg.com/garmin-fenix-6/	2022-10-31T00:00:00+01:00	5	
o.top10] https://www.federweg.com/albrauf-mtb/	2022-10-31T00:00:00+01:00	19	
o.top10] https://www.federweg.com/evoc-hip-pack-capture-7l-test/	2022-10-31T00:00:00+01:00	8	
o.top10] https://www.federweg.com/touren/bayern/	2022-10-31T00:00:00+01:00	105	
o.top10] https://www.federweg.com/kettenwachs-von-optimize-im-test/	2022-10-31T00:00:00+01:00	3	
o.top10] https://www.federweg.com/bike-and-hike/	2022-10-31T00:00:00+01:00	14	
o.top10] https://www.federweg.com/mountainbiken-tegernsee-schliersee/	2022-10-31T00:00:00+01:00	17	
o.top10] https://www.federweg.com/fox-speedframe-pro-test/	2022-10-31T00:00:00+01:00	29	
o.top10] https://www.federweg.com/mountainbiken-zillertal/	2022-10-31T00:00:00+01:00	11	
o.top10] https://www.federweg.com/mtb-transalp/	2022-10-31T00:00:00+01:00	145	
o.top10] https://www.federweg.com/ausruetzung/	2022-10-31T00:00:00+01:00	0	
o.top10] https://www.federweg.com/joe-route/	2022-10-31T00:00:00+01:00	10	

Abb. 7: JSON-Daten als Tabellenspalten

Konfiguration der Node alle Spalten außer der Spalte API-URL und der Spalte Address aus dem grünen in den roten Bereich, um nur die beiden Spalten bei der Aggregation zu berücksichtigen (Abb. 4). Wechseln Sie nun in den Tab Options und wählen Sie Concatenate und entfernen Sie im Feld Value delimiter das Komma, sodass dieses Feld leer ist.

Nach dem Ausführen der Nodes sehen Sie nun in der Tabelle eine neue Spalte mit dem Namen Concatenate. Sie enthält die zusammengeführten Daten aus API-URL und der URL, für die die Daten ermittelt werden sollen.

Um die Namen der Spalten übersichtlich zu halten, kann die neu erzeugte Spalte mit der Node Column Rename umbenannt werden, beispielsweise in den aussagekräftigeren Namen Abfrage URL (Abb. 5). Öffnen Sie dazu die Einstellungen der Node und klicken Sie doppelt auf die umzubenennende Spalte (in diesem Fall Concatenate). Sie erscheint nun in der rechten Spalte und sie kann mit einem neuen Namen versehen werden.

API-Daten per POST Request abfragen

Nun ist es an der Zeit, endlich die Daten bei Sistrix abzufragen. Dazu ergänzen Sie den Workflow um die Node POST Request. Mit einem POST Request können Daten (z. B. aus einem Kontaktformular) an einen Webserver gesendet und empfangen werden. Die Node sendet also die Abfrage-URL an den Sistrix-Server und dieser antwortet mit den abgefragten Daten.

In der Konfiguration der Node wählen Sie lediglich die Spalte mit der Abfrage-URL aus (Abb. 6). Nach dem Ausführen der Node enthält der Workflow eine weitere Spalte (body), in der die Daten aus der Sistrix-API im JSON-Format enthalten sind.

In einem Schritt: JSON-Daten als Tabellenspalten

Um die JSON-Daten aufzubereiten, dient die Node JSON to Table. Bei dieser Node muss nichts konfiguriert werden. Sie muss lediglich an den Workflow angehängt und ausgeführt werden. Die Node konvertiert den Inhalt der Spalte mit den JSON-Daten in mehrere Spalten. Anschließend gibt es für jeden Wert, den die API ausgibt, eine neue Spalte in der Tabelle (Abb. 7).

Die Anzahl an Keywords, für die jede abgefragte URL in den Top 10 rankt, finden Sie in der Spalte value.

Außerdem werden noch Informationen über die Anzahl der verwendeten Credits (Spalte used), das Abfragedatum, die abgefragte URL, die Abfragemethode und einige weitere als eigene Spalten angezeigt.

Um die Tabelle übersichtlich zu halten, können Sie bei Bedarf mit der Node Column Filter nicht benötigte Spalten herausfiltern oder mit der Node Sorter die Tabelle nach Werten, wie der Anzahl der Rankings sortieren. Anschließend lassen sich die Daten als Balkendiagramm visualisieren (Node: Bar Chart).

Fazit

Dies war nur ein Beispiel, wie Daten aus der Sistris-API in KNIME genutzt werden können. In der Dokumentation der API unter www.sistris.de/api/ sind viele weitere Möglichkeiten, die API abzufragen, aufgeführt. Mit wenigen Handgriffen lässt sich der KNIME-Workflow auch auf andere Abfragemethoden (z. B. den Sichtbarkeitsindex) anwenden. Probieren Sie es aus. Wie immer gibt es diesen Workflow auch als Download unter: <https://einfach.st/knime77>. ¶



WEBSITE BOOSTING #078 erscheint am 14.02.2023

Herausgeber & Chefredakteur (verantwortlich):

Mario Fischer
E-Mail: redaktion@websiteboosting.com

Autoren dieser Ausgabe:

Dr. Martin Bahr, Britta Behrens, Kathleen Bier,
Simon Griesser, Michael Hohenleitner,
Sara Inverso, Mareike Krücken,
Lara Marie Massmann, Rico Melzer,
Maik Metzen, Lena Möslein, Uwe Roll,
Sarah van den Berg, Sarah Weitnauer

Anzeigenleitung:

Markus Lutz
E-Mail: anzeigenleitung@websiteboosting.com

Art Direction, Layout/Produktion:

Kai Neugebauer

Lektorat:

Bärbel Philipp, textperlen.de

Fotos & Illustrationen:

Website Boosting / GettyImages / DALL-E 2

Druck:

Vogel Druck und Medienservice GmbH
Leibnizstr. 5, 97204 Höchberg

Vertrieb:

PressUp GmbH
Postfach 70 13 11
22013 Hamburg
E-Mail: websiteboosting@pressup.de

Abonnement:

Website Boosting Aboservice
PressUp GmbH
Postfach 70 13 11
22013 Hamburg
Tel. 040 / 38 6666 - 342
Fax: 040 / 38 6666 - 299
E-Mail: websiteboosting@pressup.de

Erscheinungsweise: 6 x jährlich

Bezugspreis: Einzelheft: 11,80€

Bezugspreis Inland jährlich 62,00€ inkl. Versand

Bezugspreis Ausland jährlich 70,80€
inkl. Versand

Studenten im Inland erhalten gegen Vorlage einer
Immatrikulationsbescheinigung einen
Preisvorteil – Details finden Sie auf der Website.

Verlagsleitung:

Michael Müßig
Tel: +49 931 / 26 038 04,
verlag@websiteboosting.com

Anschrift des Verlages

Hotspot Verlag GmbH
Obere Landwehr 4a, 97204 Höchberg
Tel: +49 931 / 26 038 04
Fax: +49 931 / 26 038 05
E-Mail: verlag@hotspotverlag.de
www.hotspotverlag.de

Geschäftsführung:

Kai Neugebauer

Die Inhaber- und Beteiligungsverhältnisse
lauten wie folgt: Gesellschafter zu 100%
ist die Webvalue Holding GmbH

ISSN: 2191-6241

Für unverlangt eingereichte Texte und Daten kann keine Haftung übernommen werden. Sämtliche Veröffentlichungen in Website Boosting erfolgen ohne Berücksichtigung eines eventuellen Patentschutzes. Markennamen werden ohne Gewährleistung einer freien Verwendung benutzt. Trotz sorgfältiger Recherche kann für die Richtigkeit des Inhalts keine Haftung übernommen werden. Namentlich gekennzeichnete Artikel geben nicht unbedingt die Meinung der Redaktion wider.