Sebastian Erlhofer, Mario Fischer

BERECHNEN SIE IHREN PAGERANK DOCH SELBST!

Mit der Berücksichtigung eingehender Links für das Ranking eines Dokuments lieferte Google damals deutlich bessere Ergebnisse als andere Suchmaschinen. Seither ist viel passiert. Sehr viele weitere Rankingfaktoren kamen hinzu und auch die PageRank-Berechnung wurde deutlich differenzierter. Entgegen vielen Vermutungen berechnet Google nach eigenen Angaben diesen Wert auch heute noch fortlaufend neu. In die Berechnung fließen sowohl interne als auch externe Links (Backlinks) ein. Eine iterative Berechnung über das gesamte Web ist eine der kompliziertesten Angelegenheiten, weil die schiere Datenmenge nur noch über Annäherungsalgorithmen bewältigt werden kann und unvorstellbar viel Rechenpower benötigt. Experten weisen immer wieder darauf hin, dass die interne Verlinkung innerhalb einer Domain großen Einfluss auf die Wahrnehmung und damit auch auf das Ranking durch und in Google hat. Das ist Grund genug, sich mit diesem Faktor etwas intensiver zu beschäftigen und im nächsten Schritt den eigenen internen PageRank zu berechnen, um nutzbares Optimierungspotenzial offenzulegen. Sebastian Erlhofer und Mario Fischer zeigen Ihnen die Möglichkeiten dazu auf und führen Sie durch die nötigen Schritte. Das Beste dabei ist – Sie brauchen keine



Mario Fischer ist Professor für E-Commerce an der Hochschule Würzburg-Schweinfurt und Herausgeber der Website Boosting. Er beschäftigt sich u. a. seit 1996 intensiv mit dem Thema Suchmaschinen.

DER AUTOR



Sebastian Erlhofer ist Geschäftsführer von mindshape in Köln und Buchautor des Bestsellers "Suchmaschinen-Optimierung". Er begeistert sich für datengetriebene Suchmaschinen-Optimierung – insbesondere Graphen und die Google Search Console haben es ihm derzeit angetan. Programmierkenntnisse. Zwischen Ihnen und einer transparenten Berechnung steht nur ein klein wenig Experimentierfreude, die Sie dafür investieren müssten.

"PageRank ist unwichtig." Seit es diesen Wert gibt, werden Suchmaschinenoptimierer nicht müde, ihn kleinzureden. Die Gründe sind dabei vielfältig. Gering geschätzt wurde er zum einen dann, wenn man selbst einen nur niedrigen Wert auf der Domain hatte. Zu den Zeiten, als der PageRank (PR) noch in der Toolbar und Tools ausgewiesen wurde, war das oft tatsächlich ein kleines Glaubwürdigkeitsproblem. Zum anderen wurde der Wert häufig auch falsch interpretiert. Er wurde nur in ganzen Zahlen von 1 bis 10 ausgewiesen, die Abstände zwischen den Zahlenwerten waren aber nicht linear. Um die nächste ganzzahlige Stufe zu erreichen, waren die dazu nötigen Backlink(-Werte) jeweils um den Faktor fünf höher. Zudem bedeutete ein hoher PR nicht automatisch auch ein gutes Ranking. Wenn die anderen Signale schlecht waren und/oder der hohe Wert nur durch wenige PR-starke – oft gezielt gekaufte – Links zustande kam, war das Ranking PR-schwächerer Seiten oft sehr viel besser. Schließlich werden durch die PR-Berechnung nur die Linkbeziehungen von Dokumenten ausgewiesen, völlig unabhängig vom Inhalt. Wer nur eindimensional dachte, kam daher zwangsläufig zu dem Schluss, PR könne gar nicht so wichtig sein – schließlich gab es keinen direkten und immer zu beobachten Zusammenhang: hoher PR -> gutes Ranking.

Wie funktioniert die PR-Vererbung?

Vereinfach gesagt, setzt sich der PR einer Seite A aus den anteiligen PR-Summen der Seiten zusammen, die auf die Seite A verlinken. Anteilig deshalb, weil sich der vererbbare Wert einer Seite durch die Anzahl der abgehenden Links von dieser Seite aufteilt.

Ein Beispiel: Gehen von einer Seite Z genau 100 Links auf andere Seiten, bekommt jeder Link nur ein Hundertstel des originären PR-Wertes, den die Seite Z hat, mit auf den Weg. Würden nur 50 Link abgehen, hätten alle Links entsprechend den doppelten Vererbungswert. Gäbe es nur zehn abgehende Links, würde der vererbbare Wert nur durch zehn geteilt. Je mehr

 $PR_i = rac{1-d}{n} + d$

Abb. 1: Die Zusammensetzung des PageRank-Gewichts für eine Seite (n = Gesamtzahl der Seiten; d = Dämpfungsfaktor; c = Anzahl abgehender Links)

Links also abgehen, desto weniger Wert hat der einzelne Link für die "linkempfangende" Seite und umgekehrt. Die Formel in Abbildung 1 zeigt dies nochmals formal auf.

Wichtig ist zu unterscheiden, dass es eben nicht nur um die Anzahl an Links geht, sondern um die Aufteilung der "Wichtigkeit", also des PR-Wertes über die Links. Insofern reicht es nicht, ieweils nur die Anzahl an Links zu zählen, die auf eine Seite zeigen. Ein Link von einer PR-starken Seite, von der zudem noch wenige Links abgehen, erhöht den PR-Wert der linkempfangenden Seite überproportional mehr als ein Link von einer PR-schwachen Seite, die zudem noch sehr viele abgehende Links hat. Mit anderen Worten kann ein einziger "starker Link" einige Hundert oder Tausend Mal mehr wert sein als Tausende von "schwachen Links".

Das Originaldokument "The Anatomy of a Large-Scale Hypertextual Web Search Engine" ist noch heute auf dem Server der Stanford-Universität verfügbar unter *http://infolab.stanford. edu/~backrub/google.html.* Das Patent mit dem Titel "Method for node ranking in a linked database" ist u. a. über *www. google.de/patents/US6285999* einsehbar. Es wurde am 9. Januar 1998 eingetragen und am 4. September 2001 veröffentlicht.

Wie führt man die komplexen Berechnungen durch?

Keine Sorge. Wie eingangs erwähnt, brauchen Sie keine Programmierkenntnisse. Es gibt Software, welche die Berechnung des PageRanks für jede einzelne Webseite durchführt. "Die Anzahl eingehender Links ist aus PageRank-Sicht nicht der entscheidende Faktor."

Für Einsteiger eignet sich wahrscheinlich die freie Visualsierungssoftware Gephi am besten für eine einfache Kalkulation. In einem weiteren Schritt kann man die Daten auch optisch aufbereiten, was durchaus weitere Erkenntnisse bringen kann. Vereinfacht kann man sich die Software wie eine Art Adobe Photoshop für Graphen vorstellen. Doch dazu später mehr.

Der Download von Gephi geht ganz einfach direkt von gephi.org. Die Software ist für Windows, Mac OS X und auch für Linux verfügbar. Dort finden sich auch bei Bedarf entsprechende Tutorials für einen tieferen Einstieg. Wer einmal das Funktionsprinzip von Gephi verstanden hat, wird schnell

TIPP

PageRank alternativ mit "R" berechnen Wer mathematisch angehaucht ist, kennt wahrscheinlich die freie Software "R", mit der sich praktisch jede Art von Berechnungen durchführen lässt. Für eine Kalkulation des PageRanks muss nur zusätzlich ein Package installiert werden, an das man durch die Eingabe des Befehls install.packages("igraph") gelangt. Die weiteren Programmierbefehle findet man als Quelltext unter http://einfach.st/ prgithub.

Digital-Marketing-Seminare

zu Online-Marketing, SEO, AdWords, Social Media, Analyse & Optimierung

» Die Seminare von 121WATT vermitteln wirklich gelungen praktisches Wissen mit nützlichen Tipps. Ich hatte gleich das Gefühl, hier sind Profis am Werk. Denn neben den allgemeinen Themen gehen die Seminarleiter auch auf individuelle Fragen ein. Durch die Mischung aus Theorie im Vortrag und Praxis durch Aufgaben wurde es nie langweilig. «



Julia Denkert, Online-Marketing-Manager, Edeka Handelsgesellschaft Rhein-Ruhr mbH

» Ich habe insgesamt 9 Seminare der 121WATT besucht und war jedes Mal begeistert. ... Neben den vermittelten Inhalten hat mir auch das Drumherum sehr gut gefallen. Es war zu jedem Zeitpunkt zu spüren, dass die Zufriedenheit der Kunden insbesondere bei Service und Qualität für alle Mitarbeiter an erster Stelle steht. Eine klare Empfehlung. «



Angelika Simon, Leiterin Werbung und Kommunikation, S-Bahn Berlin GmbH

Alle Themen und Termine unter: www.121watt.de/seminare



"Knoten sind die einzelnen Webseiten, Kanten die verbindenden Links."

vielfältige Einsatzmöglichkeiten finden – auch abseits von Linkanalysen. Wer sich z. B. auch mit Social-Media-Analysen beschäftigt, findet dort echte Powerwerkzeuge.

Nach der Installation und dem Öffnen des Programms erscheint eine kurze Abfrage, in der man einfach ein neues Projekt erstellt (Abbildung 3).

Ein einfaches Beispiel zum Verständnis

Für ein erstes Warmwerden legt man sich am besten manuell einige Daten selbst an. Das erhöht die Übersicht und das Verständnis, was genau passiert. Dazu wechselt man nach dem Anlegen eines neuen leeren Projekts zunächst unter den Pull-down-Menüs oben von der aktivierten Schaltfläche "Übersicht" in "Datenlabor" (Abbildung 4, Markierung 1).

Hier kann man sogenannte Knoten und Kanten anlegen. Knoten kann man sich als Objekte – hier die Webseiten – und Kanten als die Verbindungen – hier Links – zwischen Knoten vorstellen. Klicken Sie auf "Knoten hinzufügen" (Markierung 2) und legen Sie die folgenden Seiten an (Markierung 3):

Seite A Seite B

Seite C

Jetzt haben Sie drei Knoten (Seite A, B und C) angelegt (Markierung 4). Anschließend legen Sie noch vier Links bzw. Kanten an.

Klicken Sie dazu im Reiter links über der Datentabelle zunächst auf "Kanten" (Abbildung 5, Markierung 1).

Das Prinzip funktioniert hier gleich. Nach einem Klick auf "Kante



Abb. 2: Mit Gephi lassen sich wunderbar alle Arten von Beziehungen berechnen und visualisieren (*Quelle: gephi.org*)







Abb. 4: Erstes Beispiel – das Anlegen von Knoten bzw. Seiten



Abb. 4: Erstes Beispiel - das Anlegen von Knoten bzw. Seiten



Abb. 6: Das Testprojekt – drei Seiten und vier Links

hinzufügen" (Markierung 2) öffnet sich ein Pop-up, in dem Sie den Startknoten und den Zielknoten auswählen können. Wählen Sie zunächst als Typ "Gerichtet" (Markierung 3) und dann nach dem folgenden Schema jeweils Start- und Zielseite (Markierung 4):

- Seite A zu Seite B
- Seite A zu Seite C
- Seite B zu Seite C
- Seite C zu Seite A
- Die angelegten Links sehen Sie

fortlaufend in der Datentabelle (Markierung 5). Sie haben nun drei Seiten und vier Links zwischen den Seiten fixiert (Abbildung 6).

Das war es schon an manueller Eingabe. Jetzt können Sie bereits zur Berechnung des PageRanks weitergehen. Dazu wechseln Sie wieder von "Datenlabor" zurück auf "Übersicht". Jetzt erscheint auf der rechten Seite eine Auswahlbox (Abbildung 7). Wählen Sie dort "Statistiken" (Markierung 1) und klicken in der Zeile "PageRank" auf den Button "Starten" (Markierung 2). In der nun erscheinenden Box wählen Sie "Gerichtet" (Markierung 3), als Wahrscheinlichkeit "O.5" (Markierung 4 – Achtung, statt Kommata erwartet Gephi immer den Punkt als Dezimaltrenner!), aktivieren "Kantengewichte verwenden" (Markierung 5) und bestätigen mit OK.

Als Wert für die Wahrscheinlichkeit verwendet man normalerweise 0,85 – so steht es im Originaldokument des PageRank-Patents. Mit dem Wert 0,5 sieht man die Unterscheide in diesem kleinen Miniweb allerdings etwas deutlicher, weswegen wir abweichend diesen verwendet haben. Bei Ihren eigenen Berechnungen sollten Sie die Voreinstellung mit 0,85 übernehmen.

Es erscheint ein Diagramm, das sie ignorieren und schließen können. Nun hat Gephi die PR-Werte unseres kleinen Miniwebs bereits berechnet. Um die Werte zu sehen, wechseln Sie wieder ins Datenlabor (Abbildung 8, Markierung 1), klicken auf "Knoten" (Markierung 2). Der PR wird direkt in einer neuen Spalte ausgewiesen (Markierung 3).

Wie man erkennt, ist der PR-Wert von Seite C der höchste, gefolgt von Seite A. Den niedrigsten Wert weist Seite B auf. Vergleicht man die Werte grob mit dem Schaubild in Abbildung 6, erkennt man, warum.



Abb. 7: Eingaben zur Berechnung des PageRanks

121WATT-Creative-Hub: Spannende Vorträge Intensives Networking Wertvolle Insights



11. Juli 2017 Design Offices Berlin

22. August 2017 Design Offices Düsseldorf

29. September 2017

Design Offices Hamburg

10. Oktober 2017 Design Offices München

Design offices monichen

29. November 2017

Design Offices Düsseldorf

Tickets sichern und dabei sein: hub.121watt.de



Seite A gibt zwei Links ab, an Seite B und C. Der vererbbare Wert teilt sich zwischen beiden Links auf. Das entspricht "halber Power" pro Link. Seite B hat nur einen ausgehenden Link (auf C) und auch Seite C hat nur einen einzigen Link, der auf A zeigt. Seite A empfängt also einen vollen Link, Seite B einmal "halbe" Linkpower und Seite C eineinhalbmal. Diese wirklich einfach dargestellte Vererbung findet man exakt berechnet mit einem Dämpfungsfaktor von 0.5 in der Spalte PageRank in Abbildung 8.

Um später mit den berechneten Werten etwas anfangen zu können (z. B. nach Größe sortieren, filtern etc.), exportieren Sie diese am besten in eine Tabellenkalkulation wie z. B. MS Excel. Dazu klicken Sie einfach auf "Tabelle exportieren" (Abbildung 9, Markierung 1) und wählen die Spalten aus, die Sie später haben möchten, in diesem Fall machen "Label" und natürlich "Page-Rank" Sinn (Markierung 2). Diese kleine Tabelle zu exportieren, bringt natürlich nichts. Aber wenn Sie später für Hunderte oder mehrere Tausend URLs Werte erzeugen, empfiehlt sich eine "externe" Weiterverarbeitung. Auch dazu später mehr.

Experimentieren Sie ein wenig selbst weiter

Legen Sie noch einige Seiten mehr an und setzen auch noch weitere Links. Nach einer erneuten PR-Berechnung werden Sie sehen, wie sich die Werte entsprechend verändert haben. Die Dateneingabe geht übrigens auch einfacher! Dazu legen Sie sich die Beziehungen einfach über eine Tabelle in einer Tabellenkalkulation oder sogar nur in einem einfachen Texteditor an. Abbildung 10 zeigt das Datenschema. In der ersten Zelle muss zwingend "Source, Target" stehen. Das benötigt Gephi zum korrekten Einlesen. Danach folgen beliebige Zelleinträge untereinander im dem Schema "URL Linkquelle", dann ein



Abb. 8: Nach wenigen Schritten ist der PageRank fertig berechnet

Datentabelle ×					
Knoten Kanten 😔	Konfiguration 🔂 Kno	iten hinzufügen Hante hinzufügen	🔛 Suchen/Ersetzen	Tabelle importiere	n 📳 Tabelle exportieren
Id	Label	Interval	PageRank		
0	Seite A		0.387643		
1	Seite B		0.214862		
2	Seite C		0.397495		
		Exportiere Tabelle in Trennzeichen: Komm Zeichensatz: UTF-8 Spalten: Id Label <u>Interva</u> PageRank OK	n CSV-Datei	×	

Abb. 9: Die PageRank-Werte lassen sich einfach in eine CSV-Datei exportieren

Komma und danach ohne Leerzeichen "URL Linkziel". Die Einträge können dabei Fantasienamen wie hier "Seite A" sein oder natürlich auch echte URLs. Wichtig ist nur, dass beide Einträge ohne Leerzeichen durch ein Komma getrennt sind und die Datei im CSV-Format abgespeichert wird. Dieses File laden Sie in Gephi ("Tabelle importieren", direkt links neben "Tabelle exportieren"). Im Importdialog (siehe Abbildung 14 weiter hinten) wählen Sie als Trennzeichen "Komma", als Tabelle "Kanten-Tabelle" und den passenden Zeichensatz. Nach dem Klick auf "Weiter" aktivieren Sie noch das Kästchen bei "Erstelle fehlende Knoten". Nach dem Import sehen Sie Ihre Daten im Datenlabor. Denken Sie aber daran, ggf. vorherige Testdaten zu löschen oder besser ein neues Projekt anzulegen,

	А	В
1	Source, Target	
2	Seite A,Seite B	
3	Seite A,Seite C	
4	Seite B,Seite C	
5	Seite C,Seite A	
6		
7		
8		

Abb. 10: Dateneingabe über eine einfache CSV-Tabelle

damit die neuen Daten nicht additiv zu den alten hinzugefügt werden.

Nach wie vor gilt: Die Seite(n) mit dem höchsten bzw. den höheren Werten sind in dem künstlichen kleinen Web die "wichtigen". So ist es auch in der Realität bei Ihrer Domain. Jeder Link, den Sie setzen, verändert die Wichtig-



Abb. 11: Screaming Frog – die optimalen Crawling-Einstellungen zur PageRank-Berechnung

keit Ihrer Seiten und damit auch häufig das Ranking.

Netzwerkdaten mit Screaming Frog generieren

Damit man mit realen Daten arbeiten kann, muss man natürlich die Links von und zu Webseiten vorliegen haben. Dies kann man mit Tools wie z. B. Screaming Frog *(www.screamingfrog. co.uk)* leicht bewerkstelligen. In Ausgabe 36 haben wir dieses SEO-Tool schon vorgestellt. Viele SEOs setzen diese Software bereits u. a. zum Crawlen von Websites ein. Die Jahreslizenz kostet je nach Kurs etwas über 170.- €, mit der Free-Version lassen sich bis zu 500 URLs crawlen.

Rufen Sie zunächst unter "Configuration" den Menüpunkt "Spider" auf und wählen Sie im Reiter "Basic" die Optionen wie in Abbildung 11 gezeigt aus. Anschließend geben Sie die Adresse Ihrer Website in das zentrale Eingabefeld oben ("Enter url to spider") ein und starten den Crawlvorgang.

Wenn das Tool fertig ist (Status-

meldung rechts unten), exportieren Sie über den Menüpunkt "Bulk Export" und die Auswahl "All Inlinks" alle Daten einfach in ein XLSX-File. Das war auch schon alles, was wir mit Screaming Frog tun müssen.

Schritt A

Öffnen Sie die Linkdaten in Excel (zur Übersicht siehe Abbildung 13). Am besten kontrollieren Sie zur Sicherheit nochmal, ob die Daten wirklich nur Dokument-URLs (z. B. HTML) enthalten und keine Bilderadressen, CSS, Java-Script oder weitergeleitete URLs. Sollten solche vorhanden sein, entfernen Sie die Zeilen einfach. Wichtig ist dabei, dass Sie keine leeren Zeilen im Datensatz haben. Also löschen Sie nicht nur die Inhalte, sondern tatsächlich auch alle leeren Zeilen!

Löschen Sie zunächst die Spalten A und D bis I. Anschließend löschen Sie noch die erste Zeile A. Übrig bleiben nun nur noch zwei Spalten mit den Einträgen "Source" und "Destination" (Ziffer 1). Ändern Sie nun die Spaltenüberschrift "Destination" in Spalte B in "Target" um. Das braucht später Gephi, um die Daten richtig zuordnen zu können.

Schritt B

Da Gephi CSV-Dateien für den Import erwartet, müssen Sie nun noch aus den beiden Spalten A und B die Inhalte zusammenführen. Das ist ebenfalls ganz einfach. In Zelle C1 schreiben Sie die Formel (siehe Ziffer 2)

=A1&","&B1

Diese sorgt dafür, dass dem Wert der Spalte A am Ende ein Komma angefügt wird und danach sofort der Inhalt der Spalte B kommt. Die beiden Zellinhalte werden also im Ergebnis verbunden und durch ein Komma getrennt. Sie stehen dann als ein Zelleintrag in Spalte C1. Kopieren Sie die Formel anschließend bis nach unten ans Ende der Daten. Am einfachsten geht das per Doppelklick auf den kleinen Punkt recht unten in der aktiven Zelle, wo die Formel steht (Ziffer 3 in Abbildung 13).



Abb. 12: Nach dem Crawl einfach "All Inlinks" exportieren



Abb. 13: Zur Datenaufbereitung sind einige einfache Änderungen in Excel notwendig

Markieren Sie nun die gesamte Spalte C mit der eben eingetragenen Formel und kopieren Sie diese. Anschließend fügen Sie mit "Werte einfügen" an der gleichen Stelle wieder ein. Das sorgt dafür, dass die Formel mit den berechneten Inhalten überschrieben wird.

Schritt C

Die Spalten A und B brauchen Sie nun nicht mehr. Löschen Sie daher beide Spalten. Jetzt ist nur noch eine Spalte vorhanden, die je zwei URLs enthält, die mit einem Komma getrennt sind. Speichern Sie die Tabelle jetzt im CSV-Format ab. Das war's auch schon.

Jetzt starten Sie Gephi, legen ein neues Projekt an und importieren unter "Datenlabor" mit "Tabelle importieren" die eben erzeugte CSV-Datei. Im Importdialog (Abbildung 14) wählen Sie wie vorhin schon beschrieben als Trennzeichen "Komma" (Ziffer 1), als Tabelle "Kanten-Tabelle" (Ziffer 2) und den passenden Zeichensatz. Wie unter Ziffer 3 zu sehen ist, bekommen Sie bereits eine aktive Vorschau auf die Daten. Nach



Abb. 14: Der Gephi-Importdialog

dem Klick auf "Weiter" aktivieren Sie im nächsten Dialog noch das Kästchen bei "Erstelle fehlende Knoten".

Butter bei die Fische – ein erstes Optimierungsprojekt

Nachdem nun das grundlegende Funktionsprinzip der Berechnung mit Gephi klar geworden ist, können Sie jetzt tiefer in die Analyse von Netzwerkgraphen einsteigen. Damit Sie dabei nicht gleich den Überblick bei einer großen Website verlieren, nehmen wir als Beispiel einen kleinen Online-Shop, der Socken und Hosen anbietet. Die Navigations- und Footer-Links wurden nicht berücksichtigt. Bei der Beispiel-Analyse soll die interne Verlinkung im Bereich des Main-Contents optimiert werden. Die Netzwerkdaten für den Import (zum Selbstausprobieren) sowie das schon fertige Gephi-File zum einfachen Import finden Sie unter *www.websiteboosting. com/downloads/gephi-1.zip.* Worum geht es? Der Shopbetreiber möchte gerne für den Keyword-Cluster "Socken" besser gefunden werden. Die Hosen sind eigentlich nur ein Nebenangebot. Die Produktdetailseiten der Socken ranken im Longtail zwar schon sehr gut. Das Ziel ist es nun also, den generischen Keyword-Cluster um "Socken" zu stärken und damit die direkt von der Startseite verlinkte Socken-Kategorieseite zu pushen. Hier wurde bereits durch Onpage-Content-Maßnahmen und Offpage-Maßnahmen gezielt optimiert. Nun ist die interne Verlinkungsoptimierung an der Reihe.

Bei der Optimierung der internen Verlinkung startet man üblicherweise nach dem Crawlen mit einer ersten groben PageRank-Analyse wie oben beschrieben. Um nicht nur die Page-Rank-Werte in der Datenansicht des Datenlabors nutzen, sondern auch über die Netzwerkvisualisierung Erkenntnisse ziehen zu können, bedienen Sie sich zweier Einstellungen.

Zunächst sollen für unser Beispiel alle Knoten, also URLs, mit hohen PageRank-Werten größer erscheinen als die anderen. Dazu können Sie, nachdem Sie im Reiter "Übersicht" und dann rechts im Reiter "Statistiken" die PageRank-Berechnung einmal gestartet haben, die Knoten entsprechend ausgestalten. Wählen Sie im linken Bereich "Ausgestaltung" den Reiter "Knoten" aus (Abbildung 16, Ziffer 1) und klicken daneben auf das zweite Icon "Größe", welches durch unterschiedlich große Kreise illustriert ist (Ziffer 2). Klicken Sie anschließend noch auf den Unterreiter "Ranking" (Ziffer 3) und wählen als Attribut "PageRank" (Ziffer 4) aus. Beachten Sie bitte, dass diese Option nur zur Verfügung steht, wenn Sie, wie gerade eben beschrieben, die Berechnung des PageRanks auch wirklich vorher durchgeführt haben.

Abschließend können Sie noch eine Minimal- und Maximalgröße der



Abb. 15: Seitenstruktur des Beispiel-Online-Shops



Abb. 16: Menü zur Ausgestaltung von Knoten

Knoten bestimmen (Ziffer 5). Je nach Netzwerkgröße sind die Werte hier unterschiedlich. In dem Shop-Beispiel ist eine passende Minimalgröße 20 und eine Maximalgröße 80. Sobald Sie auf "Anwenden" (Ziffer 6) geklickt haben, erscheinen die URLs mit hohem PageRank größer als die anderen.

Möchten Sie zusätzlich zur besseren Visualisierung etwas Farbe einbringen? Kein Problem. Klicken Sie auf das linke Icon, die Farbpalette (Abbildung 17, Ziffer 1). Dort wählen Sie erneut den Unterreiter "Ranking" und im Drop-down "PageRank" aus. Gephi schlägt standardmäßig einen weißgrünen Farbverlauf vor. Knoten mit hohen PageRank-Werten erscheinen dann in Dunkelgrün. Je weniger Page-Rank ein Knoten hat, desto weniger intensiv wird das Grün. Wenn Ihnen die Farbwahl nicht gefällt, können Sie auch auf das unscheinbare Icon rechts unter dem Drop-down-Feld klicken (Ziffer 2) und eine andere Farbpalette

Ausgestaltung \times		-
Knoten Kanten		루 🕐 🌑 नт
Eindeutig Partition	anking	
PageRank		~
Farbe:	-	~ 2 ⁼
Spline		
	62	Anwenden

Abb. 17: Knoten können Sie je nach PageRank intensiver einfärben

wählen. Auch mehrfarbige Zuordnungen sind hier möglich – in der Praxis hat sich allerdings ein einfarbiger Farbverlauf bewährt.

Doch damit ist es noch nicht genug. Gephi kann noch mehr Farben – und die helfen Ihnen vor allem bei der visuellen Erkundung! Um die Verlinkungsstruktur im Netzwerk besser analysieren zu können, bietet das Tool die Möglichkeit, eingehende und ausgehende Kanten unterschiedlich einzufärben. Dazu klicken Sie im Fußmenü auf das kleine Symbol ganz rechts, damit das Fußmenü aufklappt (Abbildung 18, Ziffer 1).

Den Bereich "Farbe der Selektion" haken Sie an (Ziffer 2) und wählen die gewünschten Farben für eingehend, ausgehend und reziproke Verlinkungen. Im Beispiel wurden Blau, Rot und Grün verwendet.

Wenn Sie nun mit der Maus über einzelne Knoten fahren, dann werden eingehende Verlinkungen als blauer Pfeil angezeigt und ausgehende Verlinkungen als roter Pfeil. In dieser Kombination kann man dann recht gut visuell einen Netzwerkgraphen analysieren. Fahren Sie doch einmal mit der



Abb. 18: Das Fußmenü von Gephi ermöglicht zahlreiche Feineinstellungen



Abb. 19: Gephi unterstützt das visuelle Erkunden von Netzwerken optimal (Originaldarstellung in Gephi)

Maus über den Knoten "Hosen" (Abbildung 19). Die Startseite verlinkt mit einem blauen Link und von der Hosen-Kategorieseite gehen zwei rote Links auf die jeweiligen Produktdetailseiten "hose-marco" und "hose-hanns".

Sollten Sie auf Klick mehr Daten zu einem Knoten benötigen, dann wählen Sie in der linken Toolbar das letzte Icon aus (Abbildung 20). Bei Klick auf ein Knoten erscheinen dann links alle Daten – unter anderem auch das Label, in den meisten Fällen die konkrete URL sowie der berechnete PageRankWert und andere Werte, falls Sie diese zuvor berechnet haben über die rechte Seite.

Wenn Sie sich nun die Visualisierung der internen Verlinkung anschauen, sollten Sie sich eine Frage stellen: Sollen die Knoten, die einen hohen PageRank-Wert haben, auch wirklich einen hohen Wert haben? Oder stimmt die Linkjuice-Verteilung derzeit nicht? Im Beispiel besitzen die untersten Produktdetailseiten der Hosen die höchsten PageRank-Werte. Die Socken-Produktdetailseiten haben weniger PageRank, weil der Linkjuice des Socken-Kategorieknotens darüber durch drei geteilt wird. Bei den Hosen wird er nur durch zwei geteilt – es wird also anteilig mehr Linkjuice vererbt.

Der Shopbetreiber wünscht sich eine Konzentration des Linkjuices auf die Socken-Kategorieseite. Was würden Sie nun tun? Welche Content-Links von wo nach wo würden Sie setzen, um den PageRank der Socken-Seite zu erhöhen? Bei einer solchen Überlegung sollten Sie sich immer die Knoten anschauen, die den höchsten Page-Rank haben. Von dort sollten Sie Links setzen, damit der Linkjuice zu den für Sie wichtigen Seiten fließen kann. In der Praxis reduziert man gleichzeitig von diesen linkgebenden URLs möglichst noch alle ausgehenden Links, damit viel Linkjuice auf das Ziel geht, welches man verlinken möchte. Im Shop-Beispiel fällt die Entscheidung leicht: Von den Produktdetailseiten mit dem hohen PageRank muss jeweils ein Link zurück auf die darüber liegende Socken-Kategorieseite. Als Verstärkung geben wir dann noch einen Link von der Hosen-Kategorieseite auf



die Socken-Kategorieseite – ganz nach dem Motto: "Sie suchen Hosen? Wir haben die passenden Socken dazu!"

Im optimierten Soll-Zustand erhält die Socken-Kategorieseite mit Abstand am meisten PageRank (Abbildung 21). Damit hat Google eindeutig ein Signal, dass es sich hierbei aus Sicht der internen Verlinkung um die wichtigste Seite handelt.

Im Beispiel können Sie wie oben gelernt die Links direkt in Gephi von unten nach oben und von der Hosen-Seite aus zur Socken-Seite manuell einziehen. In der Praxis bei großen Websites ist das allerdings nur mit äußerst hohem Aufwand und praktisch nicht ohne Fehler möglich. Versuchen Sie mal in einem Netzwerk mit Tausenden von Kanten manuell Links zu ziehen. Das ist nahezu unmöglich. In der Praxis werden daher die Optimierungen meist direkt in den Templates bzw. im HTML durchgeführt. Das findet dann natürlich nicht im Livesystem statt, sondern in dem Development- oder Testsystem. Sobald die IT die Änderungen auf Anweisung des SEOs eingespielt hat, crawlt man dann erneut die Domain und erstellt eine zweite Version des Netzwerks. Hier wird dann aufs Neue die PageRank-Berechnung durchgeführt, die Knotengröße wie beschrieben angepasst und schon kann man sich – hoffentlich – über eine optimierte PageRank-Verteilung freuen. Wenn dem so ist, dann steht einem Ausrollen auf das Livesystem nichts mehr im Wege.

Größere Websites visualisieren

In der Realität werden Sie es wohl kaum mit solchen kleinen und übersichtlichen Websites zu tun haben. Schon wenige Hundert URLs mit vielen internen Verlinkungen auf einer Domain sorgen dafür, dass das Netzwerk recht undurchsichtig wird. Das abgebildete Netzwerk besitzt 2.285

Ausgestaltung	Bearbeiten ×	- Graph
Isocken/socke-ma	rio/ - Eigenschaften	Go 🛚
Größe	60.0	
Position (x)	-589.7753	
Position (y)	55.435345	
Position (z)	0.0	
Farbe	[67,165,124]	
Beschriftung Größe	1.0	
Beschriftung Farbe	[61,0,9]	🌚
Beschriftung sichtba	ır 🗹	
Isocken/socke-ma	rio/ - Attribute	
ld	4	
Label	/socken/socke-mario/	
Interval	<nullwert></nullwert>	
PageRank	0.1040618953482472	
Layout ×		

Abb. 20: Für jeden Knoten können die Detaildaten angezeigt werden



Abb. 21: Ziel erfüllt – die Socken-Kategorie besitzt am meisten PageRank (Originaldarstellung in Gephi)

interne Verlinkungen. Wie Sie sehen, erkennen Sie nicht viel. Dabei sind es lediglich 86 URLs! Wenn Sie sich das selbst einmal anschauen möchten, dann finden Sie die Import-Datei unter www.websiteboosting.com/downloads/gephi-2.zip.

Natürlich kann man auch ohne weitere visuelle Unterstützung die PageRank-Berechnung in Gephi durchführen und daraus Optimierungsmaßnahmen ableiten. Aber eine gute Visualisierung hilft hier und da durchaus sehr, um Probleme und Chancen bei der internen Verlinkung zu identifizieren. Holen Sie einfach ein wenig mehr aus den Gephi-Funktionen heraus!

Wenn Gephi ein Netzwerk wie das in Abbildung 22 darstellt, werden die Knoten zunächst zufällig im Raum verteilt. Im ersten Schritt sollten Sie die bereits beschriebene PageRank-Berechnung durchführen, die Knoten einfärben und größentechnisch optimal einstellen. Beachten Sie bei der PageRank-Berechnung, dass Sie auch die Checkbox bei "Kantengewicht verwenden" aktivieren. Das Netzwerk hat teilweise mehrere Links von einer URL auf eine andere und damit ist eine Kante "gewichtiger", also dicker, als andere. Hier fließt dann entsprechend mehr Linkjuice anteilig in die Berechnung ein.

Mit sogenannten Layout-Algorithmen können Sie Ordnung in das Chaos bringen und die einzelnen Knoten nach einer gewissen Logik anordnen lassen. In Gephi stehen verschiedene Layout-Algorithmen zur Verfügung. Leider gibt es keinen, der einer umge-



Abb. 22: So sieht ein komplexeres Netzwerk ohne Bearbeitung in Gephi aus

drehten Baumstruktur nahekommt, wie sie häufig für die Visualisierung von Website-Strukturen genutzt wird. In der Praxis haben sich vor allem zwei Layouts bewährt. Der eine nennt sich "Fruchterman Reingold" (Abbildung 23). Dieser Algorithmus sieht Knoten als Massepartikel an und die Links zwischen ihnen als Verbindung. Der Algorithmus versucht nun, eine ideale Position für jeden Knoten zu finden, wo er als Masseelement im Gleichgewicht mit all seinen Verbindungen ist und perfekt zu anderen Masseartikeln steht. Wie jeder Layout-Algorithmus läuft er iterativ, das heißt, er optimiert die Position der Knoten Runde für Runde. In Gephi können Sie links unten im Bereich "Layout" den "Fruchterman Reingold"-Algorithmus auswählen. Darunter erscheinen dann die möglichen Einstellungen.

Mit "Fläche" stellen Sie die maximale Ausbreitungsfläche ein. Je umfangreicher ein Netzwerk, desto größer sollte der Platz sein. Die

Layout ×	
Fruchterman Reingold	
Fruchterman Reingold	
Noverlap	2
OpenOrd	
Rotate	
Yifan Hu	
Yifan Hu Proportional	
Zufälliges Layout	
Zusammenziehen	
0	ſ
E Fruchterman Reingold	
Fläche	10000.0
Anziehungskraft	10.0
Geschwindigkeit	1.0

Abb. 23: Einstellungen für das Fruchterman-Reingold-Layout

Anziehungskraft ist meist mit dem Wert 10 recht gut brauchbar. Ist Ihnen das Netzwerk visuell zu weit auseinander, erhöhen Sie diese Zahl. Stehen die Knoten zu eng zusammen, senken Sie die Anziehungskraft. Erhöhen Sie die "Geschwindigkeit", wenn Ihnen die Berechnung zu langsam ist. Hier verzichten Sie dann allerdings auf die Genauigkeit bei der visuellen Ausrichtung. Starten Sie das Layout mit Klick auf den "Starten"-Button. Sobald sich visuell nicht mehr viel tut, können Sie die Layout-Iteration mit einem Klick auf "Stop" anhalten. Ein anderes praxisbewährtes Layout ist "Force Atlas 2". Dieses Layout betont noch mehr zusammengehörige Strukturen. Mehr Informationen dazu und zu allen anderen Layouts finden Sie unter *http://einfach. st/gephi4*.

Mit der Einfärbung, der Größendarstellung sowie der Anwendung des Layout-Algorithmus kann man nun schon deutlich mehr erkennen. Links oben befindet sich beispielsweise die Startseite. Sie ist recht isoliert. Im Zentrum sitzt die URL mit dem höchsten Page-Rank, die /blog/-Startseite.

Möchten Sie visuell noch etwas mehr aufräumen? Kein Problem. Das erleichtert in der Praxis nämlich ungemein die Analyse der PageRank-starken Verbindungen. Nutzen Sie dazu rechts den Tab "Filter" (Abbildung 25, Ziffer 1). Wenn Sie nur Knoten anzeigen möchten, die über einen gewissen PageRank verfügen, dann wählen Sie via Doppelklick "Attribute > Bereich" und dort "Page-Rank" (Ziffer 2). Der Eintrag erscheint dann unten im Bereich "Abfragen" (Ziffer 3) und Sie können den Slider entsprechend bedienen (Ziffer 4). Drücken Sie ruhig vorher auf den Button "Filter", damit Sie den Slider quasi live bedienen können.

Die Filtermöglichkeiten lassen sich auch beliebig mit logischen Verknüpfungen verbinden. Probieren Sie ruhig ein wenig aus.

Gephi wäre nicht der Photoshop der Graphendarstellung, wenn Sie nicht auch Cluster aufgrund der internen Verlinkung automatisch erstellen lassen könnten. So können Sie mit Gephi auf Knopfdruck gewissermaßen Silos erkennen, die sich dank der internen Verlinkung identifizieren lassen. Dazu klicken Sie im Tab "Statistiken" auf "Durchschn. Clustering Koeffizient". Sie erhalten damit dann eine neue Knoten-Eigenschaft, den "Clustering Coefficient". Dies ist eine graphentheoretische Kenngröße. Sie ist ein Maß für die Cliquenbildung von Knoten. Eine Clique in der Graphentheorie ist ein Subnetzwerk, in dem alle Nachbarn eines Knotens paarweise verbunden sind. Wenn also jeder Knoten mit jedem verbunden ist, dann spricht man von einer Clique. Für die SEO-Arbeit hilft diese Cliquen-Identifikation, um stark reziprok miteinander verbundene URLs zu erkennen. Ebenfalls spannend ist die "Modularität", die man auch entsprechend granular einstellen kann. Damit lassen sich zusammengehörige Gruppen anhand ihrer Verlinkungsdichte algorithmisch feststellen.

Visualisiert man das Netzwerk von oben im "Force Altas 2"-Layout (Abbildung 26) und kombiniert die Module als farbige Knoten, dann kann man bestimmte Segmente der Website gut erkennen: Die lila Knoten halblinks oben decken den Hauptbereich der Website ab, inklusive der Startseite. Ebenfalls in Lila, allerdings links darüber sowie



Abb. 24: Nach der Anwendung des Layout-Algorithmus wird das Netzwerk spürbar besser visualisiert



Abb. 25: Über Filter lassen sich kleinere Knoten einfach ausblenden



Abb. 26: Wenn man das Netzwerk etwas in Form gebracht hat, lassen sich Cluster sehr gut erkennen



Abb. 27: Mit der Gephi-Vorschau-Funktion kann man solche Graphen als PDF oder SVG exportieren

rechts darunter sind jeweils einzelne Seiten gruppiert nach ihren Verzeichnissen. Orange oben abseits stehen PDF-Publikationen der Website. Den Rest, kugelförmig um die dicke grüne Blog-Startseite angeordnet, sind einzelne Blogbeiträge und Archiv-Seiten des Blogs. Hier sieht man also ein typisches Beispiel einer Domain, deren interne Verlinkung stark von dem Blog geprägt wird. Dennoch besitzen die eigentlichen Website-Inhalte in Lila die stärkeren PageRank-Werte.

Verlinkungsnetzwerk – mach dich schön!

Neben all den nützlichen Funktionen in Gephi muss im Leben eines SEOs aber auch mal hier oder da eine schöne Grafik her, sei es für den Pitch oder die interne Präsentation im Unternehmen. Dazu bietet Gephi den Bereich "Vorschau". Mit diversen Voreinstellungen und ein bisschen Feintuning lassen sich hier meist ganz ansehnliche Graphen erstellen und als PNG, PDF oder sogar als SVG exportieren. Klicken Sie dazu einfach oben auf den dritten Reiter "Vorschau" (rechts neben "Graph" und "Datentabelle") und wählen Sie beispielsweise "Standard – bogenförmig" aus. Hier lassen sich nun diverse optische Einstellungen tätigen, die nahezu keine Wünsche offenlassen: die Rahmenbreiten und -farben von Knoten, Beschriftungen, Kantengrößen und -formen und vieles mehr. Vor allem die Eigenschaften "Dicke" und "Deckkraft" im Bereich "Kanten" sind hilfreiche Stellschrauben bei großen Netzwerken mit vielen Kanten. Hier gilt die Devise: Ausprobieren, bis es schön aussieht.

Fazit

Eigentlich waren das hier nur die Grundlagen, wie Sie schnell und effektiv Ihren PageRank nicht nur selbst berechnen, sondern auch entsprechend visualisieren können. Wen nur die berechneten Werte interessieren, kann die von Gephi erzeugte Datei einfach in Excel oder einer anderen Software weiterverarbeiten und das filtern und sortieren, was ihn interessiert bzw. was optimiert werden müsste. Das geht schnell und ist leichter zu bewerkstelligen, als es auf

BEISPIELE AUS DER PRAXIS – SO KANN MAN ÜBER VISUALISIERUNG SCHNELL EINEN EINDRUCK EINER WEBSITE-STRUKTUR ERHALTEN

Links sehen Sie eine Website eines mittelständischen Herstellers für Türen. In der Mitte stehen zahlreiche URLs, die jede Unternehmenswebsite bietet (über uns, Impressum usw.). Die Startseite sticht nicht einmal hervor – auch alle anderen URLs im Zentrum sind gleichwertig miteinander verbunden, sodass keine in Sachen PageRank herausragt. Auffällig sind die sauber getrennten einzelnen Silos. Hier handelt es sich jeweils um verschiedene Warengruppen (Türen, Schiebetüren, Zargen, Garagentore usw.). Lediglich die Ver-

teilerseiten in das jeweilige Silo bilden den Zugang zu den Unterseiten. Optimierungswürdig wäre hier vor allem die Steuerung des Linkjuices auf zentrale und für das Unternehmen wichtige Seiten. Das Siloing für eine klare thematische Einordnung seitens Google sollte so belassen werden.

Solche extremen Cluster kann man bei vielen Websites sehen. **Im rechten Beispiel** handelt es sich um die Website eines Messebauers. Links oben befindet sich in Orange das gesamte Blog. Hier sticht keine URL so richtig heraus. Rechts unten in Lila ist die eigentliche Website – ähnlich wild miteinander verbunden. In der Mitte in Grün stehen die jeweilige Blog-Startseite und die eigentliche Startseite, welche die beiden Hauptwebsite-Bereiche miteinander verbinden. Hier könnte man durchaus eine stärkere Integration des Blogs in die normale Website als Optimierung ansetzen. Durch gezielte Verlinkung würden die beiden Graphenteile zusammenwachsen und mehr Linkjuice und Traffic miteinander austauschen.

den ersten Blick aussieht – man muss es nur einfach einmal selbst machen.

Aber gerade in der Visualisierung liegt ein wesentlicher Vorteil. Optische geclusterte Darstellungen sind für den Menschen einfacher und intuitiver zu verstehen. Je nach Analysezweck können Sie so für Dritte Entscheidungsvorlagen erstellen. Das beeindruckt in der Regel sehr viel mehr als Zahlensalate. Gephi wirkt im ersten Moment vielleicht etwas abschreckend, weil man erst die Philosophie "dahinter" verstanden haben muss: Daten importieren, nach einem Algorithmus Beziehungen berechnen lassen, nach diesem Algorithmus die erzeugten Werte optisch hervorheben (Größe und Farben), über einen Visualsierungsalgorithmus alles optisch gut aufbereiten lassen und am Ende noch über Filter hervorheben bzw. unterdrücken, was wichtig oder unwichtig ist. Und keine Sorge, wenn Sie sich einmal eingearbeitet haben, werden Sie noch viel mehr nützliche Funktionen für Datendarstellungen der Beziehungen von Objekten aller Art finden – sicher auch außerhalb des Themas SEO. Um Ihnen ein wenig Appetit zu machen: Auch die Fans Ihrer Facebookseite kann man mit Gephi bestens auswerten und z. B. entsprechende Influencer unter ihnen finden. ¶

