

## Pagerank

Stand: 21.07.2022

### Der PageRank-Algorithmus

Die folgende rekursive Formel dient zur Berechnung des PageRank:

$$PR(A) = (1-d) + d (PR(T_1)/C(T_1) + \dots + PR(T_n)/C(T_n))$$

Hierbei gilt:

- $PR(A)$  ist der PageRank einer Seite A.
- $PR(T_1)$  bis  $PR(T_n)$  ist der PageRank der Seiten  $T_1$  bis  $T_n$ , von denen Links auf Seite A verweisen.
- $C(T_1)$  bis  $C(T_n)$  ist die Gesamtzahl der Links auf den Seiten  $T_1$  bis  $T_n$ .
- $d$  ist ein Dämpfungsfaktor zwischen 0 und 1.

Aus der Formel ergibt sich: **Je mehr ausgehende Links eine Seite T aufweist, umso weniger PageRank vererbt sie an die Zielseite A.** Der Effekt der linkgebenden Seite T „verwässert“ sich gewissermaßen.

Der **Dämpfungsfaktor d** bewirkt, dass die Linkverfolgung nicht unendlich fortgesetzt wird.

Bei der hier genannten Formel handelt es sich um den **ursprünglichen PageRank-Algorithmus**, wie er zuerst von Larry Page und Sergey Brin an der Universität Stanford vorgestellt wurde. In der Zwischenzeit wurde der Algorithmus **mehrfach im Rahmen der Google-Updates modifiziert**. Google veröffentlicht diese neueren Varianten aber nicht, um Manipulationen beim Ranking vorzubeugen.

### Das Random-Surfer-Modell



Dem ursprünglichen PageRank-Algorithmus lag das Random-Surfer-Modell zugrunde. Hierbei wurde ein Benutzer angenommen, der sich von Link zu Link durch das Netz klickt, ohne dabei auf Inhalte zu achten.

Dieser Zufallssurfer befindet sich mit einer bestimmten Wahrscheinlichkeit auf einer Webseite, welche durch den PageRank abgebildet wird. **Abhängig davon, wie viele Links er nun zur Auswahl hat, ergibt sich die Wahrscheinlichkeit, mit der er sich auf die von uns betrachtete Webseite A weiterbewegt.** Das ist also der Grund, warum der Algorithmus den PageRank der verlinkenden Seite durch die Gesamtzahl der ausgehenden Links teilt.

Allerdings wird der Zufalls-Surfer sich nicht unendlich lange durch das Internet klicken – irgendwann bricht er ab. Um das abzubilden, **enthält der Algorithmus den Dämpfungsfaktor  $d$** , der die Wahrscheinlichkeit angibt, mit der Links immer weiterverfolgt werden. Geht der Dämpfungsfaktor gegen 1, werden Links länger verfolgt; geht er gegen 0, bricht der Zufallssurfer frühzeitig ab.

## Das Reasonable-Surfer-Modell

Im Jahr 2004 hat Google eine Weiterentwicklung des Random-Surfer-Modells zum Patent angemeldet, das Reasonable-Surfer-Modell. Hier klickt der User nicht länger beliebig auf einen der vorgefundenen Links, sondern **nimmt eine Einschätzung vor, welche Links für ihn relevanter sind**. Beispielsweise klicken User nur **selten auf Links wie die AGB oder Werbelinks**. Aber auch die Position, das Design, die Länge und der Kontext des Links werden beim Reasonable-Surfer-Modell berücksichtigt. Links sind hier also nicht länger gleichbeschaffen, sondern wirken sich unterschiedlich stark auf den PageRank der Zielseite aus.

## Praktische Verfahren zur Ermittlung des PageRank

Bis April 2016 gab es im Wesentlichen zwei praktische Wege, um den PageRank einer Seite zu ermitteln: PageRank-Kalkulationstabellen wie die von Bob Wakfer und die Ermittlung des PageRank per Add-on.

## Die PageRank-Kalkulationstabelle

PageRank-Kalkulationstabellen klärten, wie viele Links mit PR x benötigt werden, um einen PR y zu erreichen. Das machte es Online-Marketern leichter, Prognosen und Aufwandsabschätzungen vorzunehmen. Da Google den PageRank-Algorithmus nicht bekanntgab, stellte diese Tabelle eine Näherung dar, aber eine in der Praxis sehr brauchbare:

	1	2	3	4	5	6	7	8	9	10
2	101	18	3	1	1	1	1	1	1	1
3	555	101	18	3	1	1	1	1	1	1
4	3055	555	101	18	3	1	1	1	1	1
5	16803	3055	555	101	18	3	1	1	1	1
6	92414	16803	3055	555	101	18	3	1	1	1
7	0,5 Mio.	92414	16803	3055	555	101	18	3	1	1
8	2,8 Mio.	0,5 Mio.	92414	16803	3055	555	101	18	3	1
9	15 Mio.	2,8 Mio.	0,5 Mio.	92414	16803	3055	555	101	18	3
10	84 Mio.	15 Mio.	2,8 Mio.	0,5 Mio.	92414	16803	3055	555	101	18

Ein Beispiel für die Verwendung der Tabelle:

Um einen PR 4 zu erreichen, brauchte man einen [Backlink](#) von einer PR-6-Seite. Alternativ brauchte man 3 Backlinks von PR-5-Seiten oder 18 Backlinks von PR-4-Seiten usw.

Grundsätzlich galt: Um mit einem einzelnen Link einen bestimmte PR zu erreichen, brauchte man einen Backlink von einer Seite, deren PR zwei Punkte höher lag. Um mit niedrigen PR-Seiten einen höheren PR für die eigene Seite zu erreichen, bedurfte es entsprechend vieler Links: Für einen angepeilten PR 4 brauchte man beispielsweise 3055 Backlinks von PR-1-Seiten.

## Add-ons

Browser-Add-ons stellten bis April 2016 die einfachste Möglichkeit dar, den PageRank zu ermitteln. Einmal installiert, zeigte die Extension beim Besuch einer Webseite automatisch oben rechts den PR der jeweiligen Seite an, meist in Verbindung mit einem grünen Balken. Beliebte Add-ons waren und sind z.B. der von Google selbst bereitgestellte **Google-Toolbar**, oder der **MozBar**. Neben dem PR ermittelten solche Add-ons oft auch den **Alexa-Rank** und andere für die [Suchmaschinenoptimierung](#) relevante Kennwerte.

## Geschichte



**Der PageRank-Algorithmus wurde 1996 von Sergey Brin und dem namensgebenden Larry Page an der Universität Stanford entwickelt und 1997 zum Patent angemeldet.** 1998 erschien das erste Paper zum Thema. Kurz darauf gründeten Brin und Page zusammen Google Inc., das Unternehmen hinter der Suchmaschine Google.

Das Patent für den PageRank-Algorithmus **gehört der Universität Stanford**, nicht Brin und Page. Google Inc. verfügt allerdings über die exklusiven Nutzungsrechte. Die Stanford University erhielt von Google 1,8 Millionen Anteile für das Patent, welche sie 2005 für 336 Millionen Dollar verkaufte.

In der empirischen Sozialforschung gab es bereits in den 50er-Jahren Überlegungen zu den Beziehungen zwischen Gruppenmitgliedern, welche die Kerngedanken des PageRank-Algorithmus vorwegnehmen.

## Die Bedeutung des PageRank

### Die Bedeutung des PageRank bis 2013

Bis Dezember 2013 aktualisierte Google den PageRank relativ regelmäßig und stellte die Bewertungen öffentlich zur Verfügung. Bis zu dem Zeitpunkt war der PageRank **eines der wichtigsten Kriterien für die Suchmaschinenoptimierung**. Seiten, die einen hohen PR aufwiesen, rankten besonders gut und verfügten in den Augen von Google über viel „Trust“. Beispiele für solche Seiten waren die des **Weißes Hauses in den USA (PR 10) sowie Seiten von Universitäten, Tageszeitungen und NGOs**. Diese Backlinks waren sehr begehrt, weil sich so der eigene PR und die Sichtbarkeit bei Google steigern ließen.

## Missbrauch des PageRank-Algorithmus

Allerdings bildete sich unter diesen Bedingungen ein florierender Markt für **PageRank-Manipulationen** heraus. Betreiber von [Linkfarmen](#) erzeugten ein künstliches Netzwerk von Seiten, um eine Zielseite A mit möglichst vielen Backlinks zu versorgen. Beim Linkkauf konnten Webseitenbetreiber **Backlinks von Seiten mit hohem PR einkaufen**, um das Ranking ihrer eigenen Seite zu verbessern. Und beim Verkauf von Domains war es möglich, den PageRank durch bestimmte technische Manipulationen kurzfristig höher erscheinen zu lassen, als er tatsächlich war, um damit einen höheren Verkaufswert für die Domain zu erzielen.

Diese weit verbreiteten Missbrauchstaktiken waren ein wesentlicher Grund dafür, warum Google den PageRank zuletzt am 6. Dezember 2013 aktualisierte. Bis April 2016 erfolgte keine Aktualisierung mehr,

der letzte offizielle PageRank war aber weiterhin einsehbar. **Seit April 2016 gibt Google den PageRank nicht mehr bekannt.**

Die Bedeutung des PageRank heute



**Es ist davon auszugehen, dass der PageRank oder ein dem PageRank ähnlicher Algorithmus nach wie vor eine Rolle für das Ranking von Webseiten spielt.** Allerdings ist er mittlerweile ein Faktor unter vielen – **ihm gegenüber gewinnen Nutzerdaten wie die Verweildauer oder andere Nutzeraktionen zunehmend an Bedeutung.** Googles Bestreben geht also dahin, die Wertigkeit von Webseiteninhalten für den menschlichen User zum Kriterium zu machen und den Such-Algorithmus dahingehend anzupassen. Es ist aber nach wie vor möglich, das Ranking der eigenen Webseite durch das Erzeugen von geeigneten Backlinks positiv zu beeinflussen.

## Alternative Verfahren

Das Abschalten des PageRank durch Google hat auch dazu geführt, dass einige Anbieter aus dem Online-Marketing-Bereich mittlerweile mit eigenen Metriken arbeiten; ein Beispiel für einen solchen Ersatz-PageRank ist der sog. **MozRank**, der sich über den bereits erwähnten MozBar ablesen lässt. Ein weiteres Beispiel ist der sog. **Trust Flow** des Anbieters Majestic.