

HALLO. ICH BIN EIN HTML-DOKUMENT.

Karl Kratz

Onsite- und Onpage-Optimierung von Websites gelten aktuell als sichere Investition. Worauf muss man eigentlich achten, wenn man leidenschaftliche und konversionsstarke Online-Inhalte für Menschen erstellen und trotzdem gute Suchmaschinenergebnisse erzielen möchte?
Ein HTML-Dokument berichtet live und exklusiv, wie eine angetrunkene Suchmaschine „auspackt“.

DER AUTOR



Karl Kratz schreibt gerade an der Erweiterung des neuen Ebooks „Welcome to the System“. Googeln Sie einfach mal danach.

Mein Name ist „vibrator.html“. Ich bin in einer großen Website zu Hause. Die Menschen nennen mein Zuhause Online-Erotik-Shop. Eine hübsche Texterin mit tollen Augen hat mich vor ein paar Tagen auf die Welt gebracht. Ein Grafiker hat mir wunderschöne Bilder geschenkt. Zum Schluss hat mich noch irgend so ein dicker „SEO“ hochnäsiger überflogen, an mir herumgefummelt und dann behauptet: „So rankt es besser.“

Ich wurde innerhalb meines Zimmers liebevoll mit anderen Dokumenten verknüpft. Der SEO sagt dazu: „Innerhalb der Kategorie verlinkt.“ Nur mein Footer, also die Fußzeile, gefällt ihm noch nicht. Da will er „noch mal ran“. So ein Ferkel?! Und gestern ist etwas sehr Aufregendes passiert: Ich wurde in kurzer Zeit von mehreren Programmen besucht, die mich von oben bis unten komplett durchgelesen haben ... das hat gekribbelt, meine Herren! Tat aber auch irgendwie gut. Die meisten Besucher kamen, blieben kurz und verschwanden dann wieder!

Gerade öffne ich die Augen und sehe, dass mich ziemlich viele Menschen betrachten: Da stehen diese unwiderstehliche Texterin, der Grafiker mit den fettigen Haaren, der dicke SEO und noch ein paar andere Gestalten. Und alle schauen recht unglücklich drein. „Der Mist wird nie ranken!“, schimpft der dicke SEO. Die Texterin schaut ihn wütend an und verzieht ihr hübsches Gesicht.

Ein Kerl mit zerrissenen Jeans, der sich als „Linkbuilder“ vorstellt, schlägt vor, man solle mir einfach mehr Links geben. Ein vornehm gekleideter Herr mit Brille wiegelt sofort ab: „Zu teuer. Lohnt sich nicht!“ Das macht mich sehr traurig ... bin ich ... wertlos? Eine AS-CII-Träne kullert über meinen Quelltext und versickert im Server.

Der dicke SEO drängelt sich vor den Bildschirm, baut sich imposant auf und tönt in die Runde: „Die vertrauenswür-

digste Datenbasis innerhalb eines Information-Retrieval-Systems sind die darin enthaltenen Dokumente. Bei allen anderen Faktoren ist davon auszugehen, dass sie mit hoher Wahrscheinlichkeit manipuliert werden. Das sind Faktoren, die Suchmaschinenbetreiber mit genauso hoher Wahrscheinlichkeit als Nächstes ins Visier nehmen. Also sollten wir mit unserem mickrigen Budget hauptsächlich die Optimierung der Dokumente angehen.“

Ich schaue wieder zum Monitor hinaus und lasse mir nichts anmerken. Mittlerweile sind alle weg ... bis auf meine bezaubernde Texterin. Sie sitzt vor dem Monitor und streichelt mich über ihre Tastatur ... fügt ein paar Wörter hinzu, entfernt ein paar andere. Das ist so schön! Ich beginne zu dösen und alles um mich herum wird dumpf und schumrig.

Bärbel, die Suchmaschine

Ich stehe vor einer großen, klappri-gen Maschine mit der Aufschrift „Suchmaschine Bärbel II, einfaches Information-Retrieval-System“. Die Maschine hat keine Schalter, nur eine kleine Tür und eine Klingel. Ich klinge und warte gespannt. Die Tür geht auf und aus der Dunkelheit fragt mich eine tiefe (weibliche!?) Stimme: „Wer stört? Was willst du? Und **worum geht es bei dir?**“

„Worum es bei mir geht?“ Ich werde bleich. Keine Ahnung ... hastig schaue ich all meine Wörter an.

„Was ist dein unmissverständlichstes Term-Signal?“ Mir verschlägt es den Atem. Die Suchmaschine riecht wie eine **Destille!** Mir fällt hektisch ein, dass mein Dateiname vibrator.html lautet und dass in meinem Dokumenten-Titel das Wort „Vibrator“ steht. Ich nuschle etwas von: „Bei mir geht es um das Thema ‚Vibrator‘.“ „LAUTER!!!“ Ich spüre, wie sich mein Körper mit Wörtern füllt, und ich schreie laut: „BEI MIR GEHT ES UM DAS THEMA VIBRATOR!“

Ich werde gescannt. Die Suchmaschine grunzt zufrieden: „**Verhaltensauffällig** bist du schon mal nicht. Ich hasse nämlich **Spam**. Ich bin übergigens **Bärbel**.“ Ich werde wieder gescannt. Diesmal ist Bärbel überhaupt nicht zufrieden: „Kleiner ... so wird das nichts. Bei dir geht es zwar um das Thema ‚Vibrator‘. Aber so richtig überzeugt mich das nicht.“ Ich werde ein drittes Mal gescannt. Die Suchmaschine zieht eine Augenbraue hoch, kritzelt eine Zahl auf einen Zettel und drückt ihn mir in die Hand. Darauf steht einfach nur: Position #307.

Ein dumpfer Schlag trifft mich plötzlich auf den <head>-Bereich. Um mich wird es dunkel. Aus der Ferne höre ich, wie der SEO zu meiner Texterin sagt: „Platz #307. Na toll. Wir müssen die Termgewichtung für ‚Vibrator‘ deutlich erhöhen und noch mal die Proof-Keywords checken.“ Ich verstehe gar nichts. Alles dreht sich und verschwimmt ...

Termgewichtung $w = wdf * idf$ in einfachen Information-Retrieval-Systemen

„Wo bin ich?“ Ich komme wieder zu mir und blinzele. Bärbel setzt mich vorsichtig auf eine ausrangierte Festplatte. Nach einem kräftigen Schluck aus einem Glas Wodka meint sie genervt: „Deine Eltern wollen, dass du bei Suchmaschinenabfragen für das Wort ‚Vibrator‘ unter den relevantesten Ergebnissen erscheinst. Deshalb manipulieren sie dich jetzt so lange, bis sie das hinbekommen.“ Ich bin verunsichert und frage: „Meine Texterin? Der dicke SEO? Und was machen die mit mir?“

Die Suchmaschine schaut mich mit fahrigem Blick an und meint dann: „Gerade versuchen sie, dich besser aussehen zu lassen als du eigentlich bist. **Sie verändern die Gewichtungen deiner Wörter.**“ Ich verstehe genau gar nichts und bitte Bärbel um eine Erklärung.

$$w_{i,j} = WDF_{i,j} * IDF_i$$

Abb. 1: Termgewichtung $w = wdf * idf$

$$WDF_{i,j} = \frac{lb(TF_{i,j} + 1)}{lb(L_j)}$$

Abb. 2: Berechnung der Within-Dokument-Frequency

Die Suchmaschine zieht eine Zigarre aus der Tasche und beginnt zu erzählen: „Im Jahr 1992 beschrieb Donna Harmann eine Formel, um den Wörtern eines Dokuments innerhalb einer Sammlung von vielen Dokumenten eine Gewichtung beizumessen. Diese Formel, die man heute noch in Lehrbüchern nachschlagen kann, lautet wie folgt: (siehe Abb. 1)“

„Die Gewichtung eines Terms i (zum Beispiel ‚Vibrator‘) in einem Dokument j (zum Beispiel ‚vibrator.html‘) lässt sich als Produkt aus der dokumentspezifischen Gewichtung des Terms innerhalb des Dokuments und der inversen Dokumenthäufigkeit des Terms i innerhalb eines definierten Dokumentekorpus berechnen.“ Ich starre Bärbel an. Ich frage mich, ob alle Suchmaschi-

nen Wodka trinken und Zigarre rauchen, um dann so ein Zeug vom Stapel zu lassen.

„Ich habe es nicht verstanden. Bitte erkläre es mir noch mal. Und zwar ganz langsam“, sage ich und gieße meiner neuen Freundin noch ein Gläschen des 50-prozentigen Treibstoffs ein.

Bestimmung der dokument-spezifischen Termgewichtung über die Within-Dokument-Frequency

Bärbel kippt den Wodka ex und beginnt: „Lass uns mit dem WDF-Teil beginnen. Grundsätzlich ist diese **Within-Dokument-Frequency (WDF)** nichts anderes als die Suchbegriff-Dichte, welche um einen Logarithmus zur Basis 2 im Zähler und Nenner

versehen wurde. Diese über den Logarithmus erzielte ‚Stauchung‘ ist der Natürlichkeit unserer Sprache geschuldet.“

„Die Within-Dokument-Frequency ist also der Quotient aus dem Logarithmus zur Basis 2 der Termfrequenz i und dem Logarithmus zur Basis 2 der Gesamtanzahl der vorkommenden Terme im Dokument j . Wenn in dir, liebes Dokument, 400 Wörter enthalten sind und der Term ‚Vibrator‘ darin 19-mal vorkommt, dann berechnet sich der Gewichtungswert $WDF(\text{Vibrator}')$ wie folgt: (siehe Abb. 3)“

„Aaaah. Einfach den \log_2 von der Anzahl eines Wortes plus eins, und dann durch den \log_2 der Menge aller Wörter im Dokument dividieren. So einfach ist das?“, frage ich. Bärbel pafft unbeirrt weiter: „Ja und nein. Diese Berechnung ist vom Funktionsprinzip her gültig. Aber sie liefert beispielsweise bei zusammenhängenden Termen fehlerhafte Werte. Genauer wäre es, bei Termkombinationen die Anzahl der Einzeltermine von der Summe der zu betrachtenden Terme im Dokument abzuziehen: (siehe Abb. 4)“

Das ist ja einfach! „Dann muss ich ja nur 1.000-mal das Wort ‚Vibrator‘ in mir tragen, und schon weiß jeder, dass es bei mir um das Thema ‚Vibrator‘ geht!“, frohlocke ich. Bärbel sieht mich an. Ich sehe Bärbel an. „Ist das nicht so?“, frage ich zögerlich. Die Suchmaschine kippt noch ein Gläschen und antwortet genervt: „Nein. Das ist nicht so. Das wäre Keyword-Stuffing. Das wäre Spam. Dann wäre unsere Freundschaft zu Ende.“

Ich beschließe, mich mit meinen blöden Ideen ein bisschen zurückzuhalten, bis ich wirklich alles kapiert habe.

Die Inverse-Dokument-Frequency

„Du bist nicht allein auf dieser Welt“, sagt die Suchmaschine zu mir

und pafft mit ihrer Zigarre einen Krinkel. „Wenn es innerhalb eines Information-Retrieval-Systems **nur ein einziges Dokument gäbe**, ließe sich dein unmissverständlichstes Term-Signal in der Tat über die Suchbegriff-Dichte und eine Liste mit Stopp-Worten ermitteln. **Aber:** Es ist eben so, dass es sehr viele andere Dokumente im World Wide Web beziehungsweise in meiner Datenbank gibt.

Wenn wir also herausfinden wollen, welche Relevanz du für eine Suchanfrage gegenüber all den anderen Dokumenten besitzt, muss ich dich ins Verhältnis zu allen anderen Dokumenten, die ich kenne, setzen. Diese Relevanz kann ich berechnen und nützliche Informationen daraus ableiten.“

Bärbel zieht ein dickes Buch mit dem Titel „Information-Retrieval-Systeme – Grundlagen“ heraus, kippt noch schnell ein Gläschen und beginnt vorzulesen: „(...) die Inverse-Document-Frequency $IDF(i)$ wird aus dem Logarithmus zur Basis 10 des Quotienten aus dem Dokumentenkörper $N(D)$ und der Menge der Dokumente $N(i)$ gebildet, wobei $N(i)$ die Anzahl der Dokumente ist, die den Term i enthalten.“

Mein <head>-Bereich meldet tausend Fragezeichen. Eigentlich wäre es mir jetzt auch nach einem Gläschen, aber ich reiße mich zusammen. „Bärbel, das ist extrem kompliziert! Kannst du mir das nicht einfacher erklären?“ Bärbel guckt mich an. Wie verlegt sie aussieht, denke ich. Aber irgendwie schon sexy. „Schau, Kleiner. Du musst dich nur mit Dingen messen, die dich betreffen:

1. Für jedes Wort in dir gibt es da draußen eine bestimmte Anzahl an Dokumenten, die dieses Wort auch enthalten. Die Summe daraus nennt man ‚Dokumentenkörper‘, welches in der Formel als $N(D)$ bezeichnet wird.
2. Für ein bestimmtes Wort (Term i) deiner Wörter gibt es da draußen

The screenshot shows a website page for 'Vibrato.de' with the headline 'Gestatten: Vibrator, nett Dich kennenzulernen!' and an illustration of a woman. A simplified WDF formula is overlaid on the right side of the page:

$$WDF_{i,j} = \frac{lb(19 + 1)}{lb(400)} = 0,5$$

Abb. 3: Beispiel für eine vereinfachte WDF-Berechnung

The diagram shows the correct WDF formula with labels for its components:

$$W_{i,j} = \frac{lb(TF_{i,j} + 1)}{lb(L_j - (TF_{i,j} * (tc - 1)))}$$

Labels and their corresponding parts in the formula:

- Within-Document-Frequency für den Term i innerhalb des Dokument j.** points to $TF_{i,j}$
- Anzahl aller Terme im Dokument j.** points to L_j
- Häufigkeit von Term i im Dokument j** points to $TF_{i,j}$
- Anzahl der Einzeltermine bei Termkombinationen** points to $(tc - 1)$

Abb. 4: Korrekte WDF-Berechnung für Term-Kombinationen

The diagram shows the IDF formula with labels for its components:

$$IDF_i = \lg\left(1 + \frac{N_D}{N_i}\right)$$

Labels and their corresponding parts in the formula:

- Inverse-Document-Frequency für den Term i innerhalb des definierten Dokument-Korpus.** points to the entire formula
- Anzahl aller Dokumente im zu betrachtenden Dokument-Korpus** points to N_D
- Anzahl aller Dokumente in denen der Term i vorkommt.** points to N_i

Abb. 5: Berechnung der Inverse-Document-Frequency

ebenfalls eine ganz klar messbare Anzahl an Dokumenten, das ist $N(i)$.

Wenn ich das Dokumentenkörper durch die Anzahl der Dokumente teile, in denen das bestimmte Wort (Term i) vorkommt, erhalte ich eine Kennzahl, die es mir ermöglicht, die Wichtigkeit dieses Wortes innerhalb des Gesamtsystems einzustufen.“

Plötzlich komme ich mir sehr klein vor. So viele Dokumente. Viele sehen viel stärker und mächtiger aus. Manche haben ganz viele Verbindungen von anderen Dokumenten bekommen. Und manche haben bunte Fähnchen mit Zahlen drauf.

„Und wie berechne ich das Dokumentenkörper? Das kenne ich doch gar nicht?“, frage ich. „Oh, da gibt es unterschiedliche Verfahren. Ich zeige dir einen ganz, ganz einfachen Weg. Für jedes deiner Wörter gibt es eine gewisse Anzahl an Suchergebnissen in meiner Datenbank. Wenn ich diese Überlegung nun für alle deine Wörter vornehme, erhalte ich eine Anzahl an Dokumenten, in denen mindestens eines deiner Wörter vorkommt. Das ist das Dokumentenkörper.“

„Aber ich warne dich!“, raunt Bärbel und wird plötzlich sehr ernst. „Dieser Weg ist ungenau. Und wenn man versucht, mir Werte auf diesem Weg zu entlocken, liefere ich gerne auch mal ungenaue Werte aus ... um die Menschen zu verwirren ... hihhi ...“

Außerdem ist eine einfache Summenbildung beim Dokumentenkörper $N(D)$ nicht wirklich das Gelbe vom Ei: Es ist schlicht und ergreifend nicht präzise und methodisch eigentlich falsch. Innerhalb der ermittelten Dokumente der Suchmaschinenresultate können schließlich auch andere Terme aus dem Dokument j vorkommen. Das bedeutet, dass hier die Berechnung über die **Vereinigungsmenge** erforderlich ist. Die Vereinigungsmenge ist die Menge aller Wörter, die entweder im Dokument 1

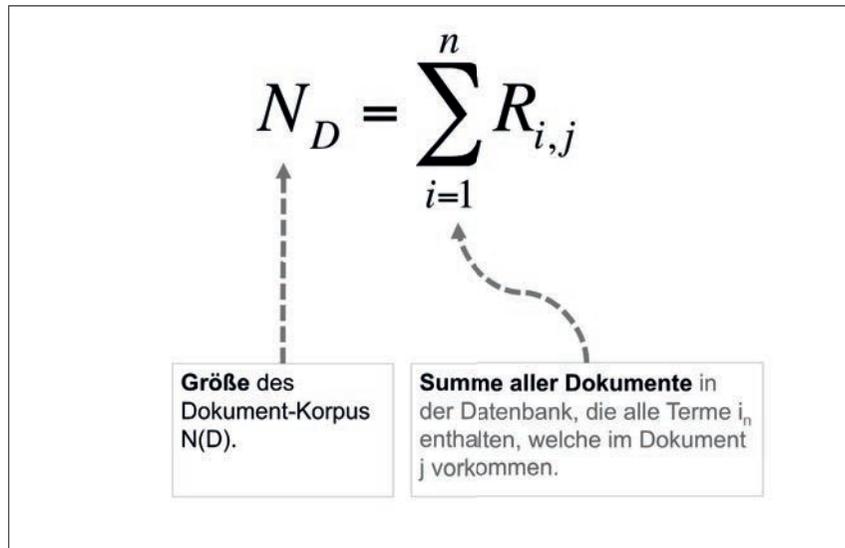


Abb. 6: Vereinfachte Methode zur Ermittlung der Größe des Dokumentenkörpers

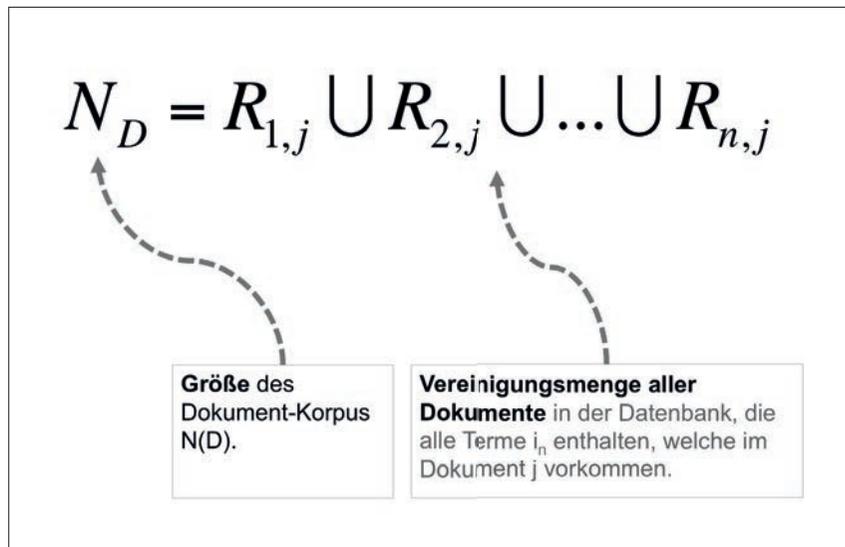


Abb. 7: Genauere Methode zur Ermittlung der Größe des Dokumentenkörpers

oder im Dokument 2 oder im Dokument 3 ... usw. enthalten sind. Diese Berechnung ist aufwendig und setzt voraus, **dass man jedes Dokument im Index kennt.** Spricht's und schenkt sich noch ein Gläschen Wodka nach. „Auf die Betreiber von Analyse-Werkzeugen!“, grölt sie.

„Also gucken wir zuerst nach den Wörtern in mir und bewerten, was dort wichtig ist. Und danach vergleichen wir mich mit allen anderen Dokumenten in deinem Index. Habe ich das richtig verstanden?“, frage ich unsicher. Bärbel strahlt über beide Ohren und legt ihre Hand auf meine Schulter. „Du bist

ein richtig smartes Kerlchen. Und jetzt bauen wir das zusammen: Dafür setzen wir die Within-Document-Frequency (WDF) und die Inverse-Document-Frequency (IDF) einfach in die Formel für die Termgewichtung ein (siehe Abb.8).“

„So, mein liebes Dokument. Jetzt können wir für **jedes deiner Wörter und jede zusammenhängende Wortkombination die Termgewichtung im Verhältnis zu allen in meiner Datenbank gespeicherten Dokumenten** berechnen. Damit erfahre ich sehr viele interessante Dinge über dich.“

Bärbel spart sich mittlerweile den ineffizienten Umweg über das Glas und

$$w_{i,j} = \frac{\text{lb}(TF_{i,j} + 1)}{\text{lb}(L_j - (TF_{i,j} * (tc - 1)))} * \text{ld} \left(1 + \frac{R_{1,j} \cup R_{2,j} \cup \dots \cup R_{n,j}}{N_i} \right)$$

Gewichtung eines Terms i im Dokument j

Anzahl aller Terme im Dokument j.

Häufigkeit des Term i im Dokument j.

Anzahl der Einzelterme der Term-Kombination

Anzahl aller Dokumente in denen der Term i vorkommt.

Vereinigungsmenge aller Resultate in der Datenbank, die alle Terme i_n enthalten, welche im Dokument j vorkommen.

Abb. 8: Vollständige Berechnung für die Termgewichtung in einfachen Information-Retrieval-Systemen

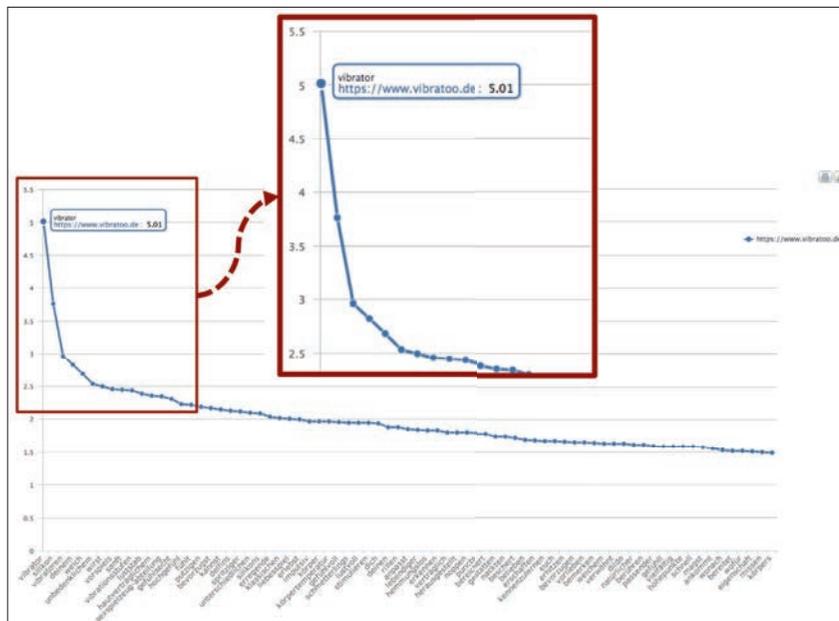


Abb. 9: Praxisbeispiel für eine dokumentenspezifische Termgewichtungs-Kurve

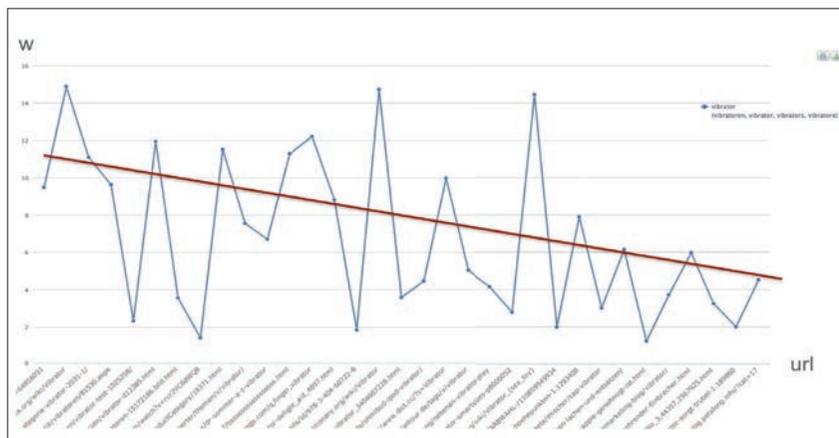


Abb. 10: Auswertung einer wortspezifischen Termgewichtung über die ersten 30 Suchmaschinenergebnisse

setzt die Wodkaflasche direkt an. Noch einen Zug an der Zigarre. Ich frage mich wirklich, ob alle Suchmaschinen so hart drauf sind. Vielleicht erfahre ich das ja eines Tages.

Was ist dein unmissverständlichstes Signal?

„Erinnerst du dich, was ich dich fragte, als wir uns kennenlernten?“, fragt mich die Suchmaschine. Ich nicke: „Du wolltest wissen, **worum es bei mir geht.**“ „Genau. Für mich als Suchmaschine ist diese Frage sehr wichtig. Es ist für mich **einfacher und kostengünstiger**, deinen thematischen Fokus durch die Termgewichtung zu berechnen, anstatt mir diese Informationen aufwendig über Links zusammenzutragen. Schau her, so siehst du in der Praxis aus: Dein unmissverständlichstes Term-Signal ist das Wort ‚Vibrator‘ mit einem Wert von 5,01.“

„5,01?“, frage ich. „Ist das gut? Oder ist das schlecht? Ist das viel?“ Bärbel lässt wortlos einen Bericht über ihre 30 relevantesten Suchmaschinenergebnisse für das Wort ‚Vibrator‘ heraus, starrt eine Weile darauf, um dann von Hand eine rote Linie hineinzuzichnen. „Schätzchen, alles ist relativ. Ob etwas gut oder schlecht, wenig oder viel ist, das bestimmt das Kollektiv aus allen Dokumenten. Als Suchmaschine kann ich nur das verwerten, was mir geliefert wird. Wenn es also sehr, sehr viele Dokumente mit einem Termgewichtungswert für ‚Vibrator‘ von über 10 gibt, dann wird das schon seine Richtigkeit haben. Aus der Grafik kannst du sehen, dass dein unmissverständlichstes Term-Signal von 5,01 gegenüber den relevantesten Dokumenten eher niedrig erscheint (siehe Abb. 10).“

„O. k. Dann sollte meine Texterin also mehr ein Augenmerk darauf legen, wie gut positionierte Dokumente zu diesem Suchbegriff aussehen?“ Bärbel wird plötzlich sehr anhänglich, was ich

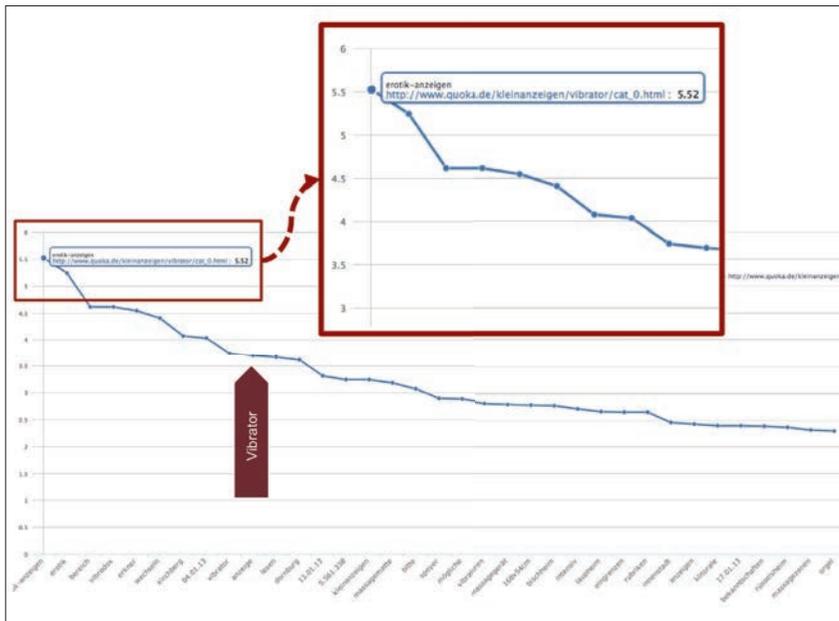


Abb. 11: Praxisbeispiel einer schlecht positionierten Webseite mit wenig aussagekräftiger Termgewichtung

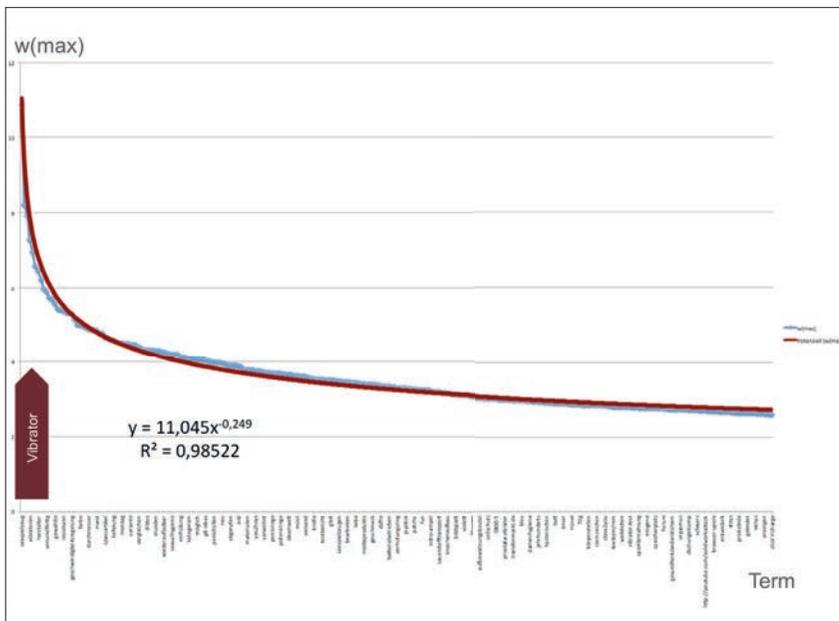


Abb. 12: Hüllkurve, bestehend aus den Termgewichtungswerten aller Terme aus den zehn relevantesten Suchmaschinenergebnissen für den Begriff ‚Vibrator‘

auf die mittlerweile drei leeren Wodkaflaschen schiebe. „Schatzi, das sollte deine Texterin in der Tat machen. Soll ich dir mal ein schlechtes Beispiel zeigen? Warte ... komm, wir gehen bei meinen Suchmaschinenresultaten für das Suchwort ‚Vibrator‘ mal auf das Ergebnis mit der Position #200. Das sieht dann so aus wie in Abb. 11 zu sehen ist“

Bärbel zippelt an einer neuen Zigarettenschachtel herum und erklärt: „In diesem Dokument geht es unter anderem auch um das Wort ‚Vibrator‘ mit einer Termgewichtung von ca. 3,6. Also nicht besonders hoch. Aber viel wichtiger ist, dass das unmissverständliche Term-Signal vom Wort ‚Erotik-Anzeigen‘ kommt. Ich musste erst einmal ein paar Millionen Dokumente durchfors-

ten, um zu lernen, dass Erotik etwas mit Vibrator zu tun hat. Das war anstrengend. Unterm Strich habe ich das Dokument für den Suchbegriff ‚Vibrator‘ hauptsächlich aus zwei Gründen auf Platz #200 verbannt:

1. Das **unmissverständlichste Term-Signal** ist nicht ‚Vibrator‘, sondern Erotik-Anzeigen und
2. ‚Vibrator‘ hat nur einen **niedrigen Termgewichtungs-Wert**.

Dieses Dokument ist für den Suchbegriff ‚Vibrator‘ allein auf der Basis dieser zwei Bewertungsfaktoren einfach nicht relevant genug.“

No Spam, Babe!

Langsam beginne ich, zu verstehen. Bärbel, die gerade die vierte Wodkaflasche öffnet, ist eigentlich sehr berechnend und entscheidet **auf der Basis einzelner Signale**, wie relevant ein Dokument für eine Suchabfrage ist! Doch weiter komme ich mit meiner Überlegung nicht. „Komm her, Schnuckel!“, lallt meine Lieblings-Suchmaschine und nimmt mich in den Schwitzkasten. „Lass uns mal gemeinsam ein bisschen an die Grenze gehen ...!“ Spätestens jetzt wird mir Bärbel unheimlich.

„Ich habe für dich eine **Termgewichtungs-Berechnung der zehn relevantesten Suchmaschinenresultate** für das Suchwort ‚Vibrator‘ erstellt. In dieser Grafik habe ich die alle Terme aus allen Dokumenten mit ihren Gewichtungswerten absteigend sortiert und zeige dabei nur die jeweiligen Maximalwerte pro Term an (siehe Abb. 12).“

„Was du hier siehst, ist eine sehr natürliche Kurve. Wenn ich so etwas sehe, werde ich immer sehr heiß und scharf.“ Bärbel schielt dabei auf meinen <body>. Ich versuche, mir nichts anmerken zu lassen. Aber irgendwie ist sie auch nett, wenn sie betrunken ist. Die nüchterne Art der Suchmaschine

holt mich sofort auf den Boden zurück. „Anhand dieser Hüllkurve stelle ich fest, ob in einem Dokument Wörter unangemessen häufig verwendet werden. Die Kurve bildet sich aus Dokumenten mit der höchsten Relevanz. Sag also deinem Super-SEO, dass er es nicht übertreiben soll ... ansonsten sehen wir uns auf Platz #1.000 wieder.“

Ich konzentriere mich: „Wenn ich das zusammenfassen darf, dann sind dir also drei Dinge sehr wichtig:

1. Ich brauche ein **unmissverständliches und gut abgegrenztes Term-Signal**, damit du auf einen Blick verstehst, worum es bei mir geht.
2. **Bestimmte Wörter** sollten in Relation zu meinem gesamten Text, aber auch mit Blick auf die Gesamtheit aller Dokumente für dieses Wort **angemessen häufig** in mir enthalten sein.
3. Solange die Termgewichtungswerte meiner Wörter **nicht deine rote Spam-Linie** überschreiten, ist alles in Ordnung.“

Bäbel freut sich und gibt mir einen euphorischen Kuss auf meinen <head>-Bereich. Ich laufe rot an. „Eine Sache zeige ich dir noch. Aber dann ist auch Schluss für heute. Weißt du, ich bin ein bisschen angeschickert, seit ich an meinem Gläschen genippt habe.“

Kannst du beweisen, worum es bei dir geht?

Schlagartig kommt mir Bäbel wieder extrem nüchtern vor, als sie mich fragt: „Du behauptest also, dass es bei dir hauptsächlich um das Thema ‚Vibrator‘ geht. Das kann ich bestätigen. Ich kann auch bestätigen, dass du keine unangemessen hohen Wortanhäufungen enthälst. Aber sage mir eines: **Kannst du aus dir selbst heraus beweisen**, dass es bei dir um das Thema ‚Vibrator‘ geht?“

Dazu fällt mir nichts ein. Ich schaue

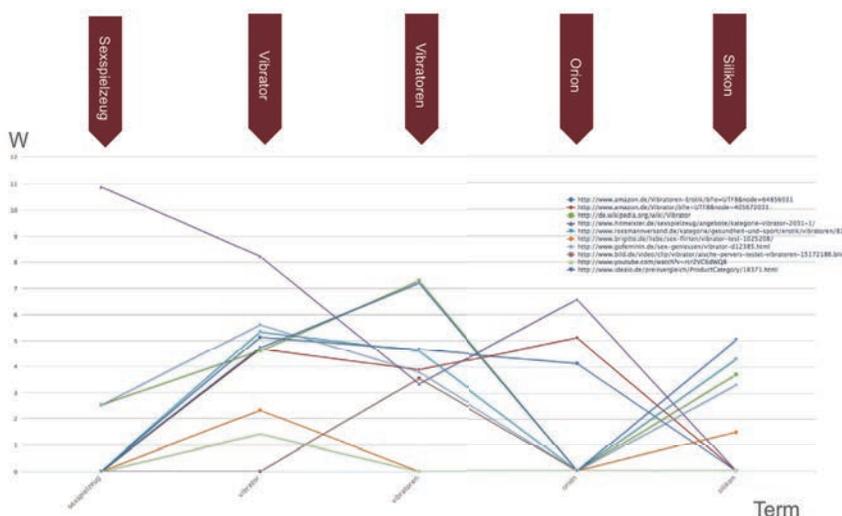


Abb. 13: Beweisführende Wörter für den Suchbegriff ‚Vibrator‘, extrahiert aus den zehn relevantesten Suchmaschinenergebnissen

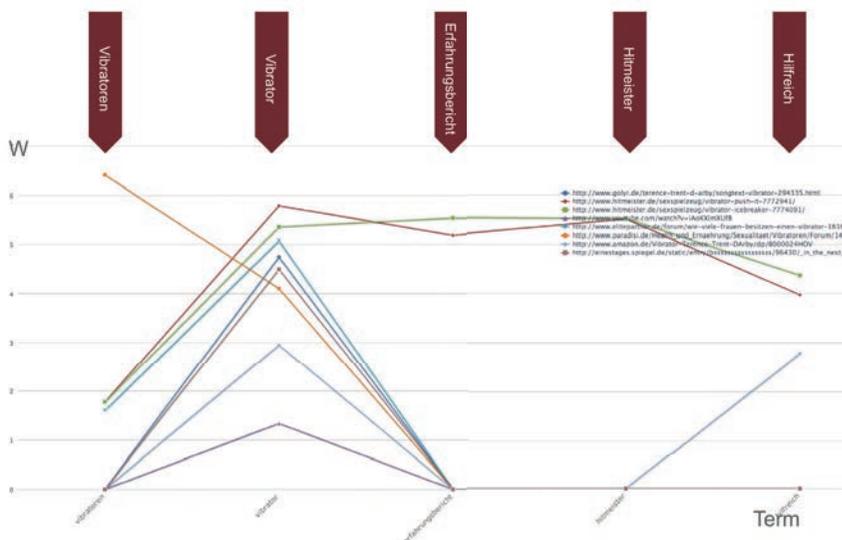


Abb. 14: Beweisführende Wörter der Dokumente der Suchmaschinenergebnisse #100 – #110

auf meine Wörter im <body>. Keine Ahnung, was Bäbel meint. Doch mir schwant, dass das eine Fangfrage war. „Oooh, da muss ich dich wohl noch mal durch-s-c-a-n-n-e-n ...“, flüstert sie mir verrückt und angeheitert ins Ohr. „Lass uns zuerst schauen, **welche Wörter in den zehn relevantesten Suchmaschinenresultaten insgesamt eine hohe Termgewichtung erhalten**: (siehe Abb. 13)“

Bäbel nimmt einen kräftigen Zug an ihrer Zigarre und fährt fort: „Wenn ich die Termgewichtung der zehn relevantesten Suchmaschinenergebnisse

auswerte und mir die Wörter anzeigen lasse, die insgesamt eine sehr unmissverständliche Gewichtung erfahren, dann lauten diese Wörter:

- » Sexspielzeug
- » Vibrator
- » Vibratoren
- » Orion und
- » Silikon

Ich habe mir von Menschen bestätigen lassen, dass das alles miteinander zu tun hat.

Übrigens: Je schlechter und irrelevanter die Suchmaschinenergebnisse



IST ES WIRKLICH SO EINFACH?

Karl Kratz, der Autor des vorangegangenen Dialogs eines HTML-Dokuments mit einer betrunkenen Suchmaschine, wurde Ende letzten Jahres von der Community rund um SEO-United.de zum SEO des Jahres gewählt. Die Branche scheint seine Ausführungen also ernst zu nehmen. Grund genug für uns, entsprechend nachzuhaken ...

Website Boosting: Karl, erst mal herzlichen Glückwunsch zur Wahl „SEO des Jahres 2012“! Hast du denn mit dieser fachlichen Sympathieexplosion gerechnet?

Karl Kratz: Ach Mario, komm – hör mir auf! Der jährliche SEO-Wahlen-Zirkus ist ja höchst amüsant und immer eine große Gaudi. Inhaltlich ist das jedoch keinen Pfifferling wert: Die SEO-Wahlen sind weder manipulationsicher noch finden geeignete Wahlkriterien Anwendung. Ich persönlich finde das Tragen eines Titels wie „SEO des Jahres“ anmaßend und möchte ihn auch nicht haben. Ich „mache mein Ding und gut is.“

„Sympathieexplosionen“ wären auch die falschen Bekundungen. Angebracht wäre kritisches Hinterfragen und vor allem konstruktives, öffentliches Weiterdenken. Nur so etwas bringt die Szene insgesamt doch erst voran!

Deine Bescheidenheit ehrt Dich natürlich. Wobei ich nicht sicher bin, dass es gleich anmaßend ist, sich offen über solche Nominierungen oder viele Stimmen zu freuen. Manipulationssicher ist wohl überhaupt nichts im Web, da hast Du sicher Recht. Marco Janck hat dich in einem

Posting zur SEOKomm als „Klaus Kinski des Online-Marketings“ bezeichnet, der „das Internet ein Stück erklärbarer macht.“ Siehst du dich selbst auch so? Was treibt dich an, so ungewöhnlich und so anders zu schreiben?

Ich lege keinen Wert auf einen Vergleich mit Klaus Kinski – schon gar nicht angesichts der letzten Nachrichten über ihn. Kinski war Kinski, Karl ist Karl :-)

Wenn sich die Möglichkeit ergibt, zu einem Online-Marketing-Thema eine noch nicht ausgeprägte Perspektive hinzuzufügen, welche einen Mehrwert für meine Leser darstellt, dann nutze ich diese. Das ist dann nicht „ungewöhnlich“ oder „so anders“, sondern einfach nur ein „Ergänzen“. Das kann eigentlich auch jeder andere: Feststellen, was fehlt und ergänzen. Dafür muss man weder besonders klug noch besonders innovativ sein. Es genügt meiner Meinung nach die Lust und Leidenschaft, sich mit Online-Themen intensiv, kritisch und bis hin zur selbstzerstörerischen, ekstatischen Manie auseinanderzusetzen ... Mehr ist es wirklich nicht.

Du hast in deinem Beitrag gezeigt, dass man nur anders schreiben müsste, um auf Top-Positionen bei Bärbel oder Google zu gelangen. Ist es nach deiner Erfahrung wirklich so einfach? Man schärft Dokumente hin-

sichtlich der verwendeten Wörter und sticht damit andere Webseiten aus, die viele gute andere Signale aufweisen?

„Nur anders“ genügt nicht und die „Einfachheit“ ist grundsätzlich von der termspezifischen Wettbewerbssituation abhängig.

Anbieter von Online-Inhalten müssen sich ja nicht gegen Google durchsetzen: Anbieter von Online-Inhalten treten gegen jeweils alle anderen Anbieter von Online-Inhalten an, bei denen terminologische Überschneidungen auftreten. In diesem Fall gewinnt, wer Onsite/[Onpage*](#)-, [Offpage*](#)-, Social- und Infrastruktur-Faktoren am besten erfüllt.

Jeder SEO kennt die Feld-Wald-Wiesen-Version davon: Es ist regelmäßig ein Kinderspiel, ohne Links für einen Drei- oder Vierwort-Term eine gute Positionierung in Suchmaschinen zu erhalten. Weshalb? Weil es genauso häufig nicht besonders viele Dokumente im Index mit dieser Termkombination gibt.

Es genügt, ungeachtet der Optimierung weiterer Rankingfaktoren, bereits einfach ein klein wenig besser zu sein als die Mitbewerberdokumente. Dieselbe Logik gilt auch für umkämpfte Einwort-Terme wie Kredit, Risikolebensversicherung, Sterbegeld usw. Mit einer präzisen Analyse und den daraus resultierenden Bemühungen lassen sich in der Tat die erstaunlichsten Ergebnisse erzielen.

*siehe Online-Glossar unter www.websiteboosting.com

Mal angenommen, mit einer „terminorientierten Gewichtungsoptimierung“ würde man tatsächlich deutlich bessere Rankings bekommen. Und nach Eindringen dieser Erkenntnis in die SEO-Köpfe würde dies dann auch großflächig eingesetzt. Hätte man und hätten Bärbel und Google dann nicht ein ähnliches Bewertungsproblem wie vorher, aber jetzt mit zu wenig „Unterschiedlichkeit“? Mit anderen Worten: Wenn alle nach dieser Methode optimieren, kann Google am Ende onpage überhaupt keine Unterscheidungen zwischen Dokumenten zu einem Thema machen, weil sie zu nahe beieinander liegen.

In der Tat lässt sich ein „Outranking“ von Mitbewerber-Websites unter definierten Voraussetzungen einzig und allein auf Basis der Termgewichtung erzielen. Dazu gab es vor allem in den hochkompetitiven Finanz- und Versicherungsbereichen bisher genügend private und öffentliche Demonstrationen.

Natürlich stellt sich die Frage: „Was ist, wenn das alle machen?“ Aber diese Frage muss doch ausnahmslos bei jedem anderen Rankingfaktor ebenfalls gestellt werden. Dazu kommen bei der Termgewichtungsanalyse ein paar „kleine Hürden“:

Grundsätzlich bedingt die Termgewichtungsanalyse und -gestützte Texterstellung den Einsatz komplexer und infrastrukturintensiver Anwendungen. Nicht „alle“ setzen solche Werkzeuge überhaupt ein, im Gegenteil: Ein ausgereiftes Termgewichtungswerkzeug auf der Basis einer Vereinigungsmengenberechnung verursacht hohe Betriebskosten. Allein durch diesen Umstand kann davon ausgegangen werden, dass nicht besonders viele Menschen „richtige“ Werkzeuge einsetzen.

Weiter gilt: „All tools are not created equal“, und nicht alle Werkzeuge sind sowohl von ihren Algorithmen als auch vom Funktionsprinzip her wirklich brauchbar. Wenn Werkzeugbetreiber z. B. Google-Re-

sultate automatisiert abrufen (scrapen), verstoßen sie gegen die Geschäftsbedingungen von Google und nutzen massiv fehlerbehaftete Daten. Werkzeugbetreiber, die zur Analyselaufzeit einen eigenen Index erstellen, sind da fein raus. Diese können und werden ihr Werkzeug wiederum nicht einer breiten Masse von Anwendern zur Verfügung stellen. Wieder fällt ein Großteil der Anwender aus der Gruppe „alle“ weg, die das Funktionsprinzip wirklich korrekt anwenden.

Zu guter Letzt kommt noch ein nicht zu unterschätzender Faktor ins Spiel: Die unendliche Bequemlichkeit des Menschen. Nun gibt es ein exzellentes Werkzeug, aber dieses macht ja den Produktionsprozess von Online-Inhalten erst einmal deutlich komplexer! Plötzlich muss ein Text nicht nur inhaltlich sinnvoll und korrekt, anregend zu lesen und konversionsstark sein, sondern auch noch ganz bestimmten SEO-technischen Anforderungen genügen, die über einen Keyword-Density-Faktor weit, weit hinausgehen.

Unterm Strich bleiben nicht mehr viele Menschen, die sowohl die richtige Technologie einsetzen als auch die Energie und Disziplin haben, perfekte Inhalte zu liefern. Eigentlich bleibt also alles beim Alten ... ;-)

Abgesehen davon sollten wir bedenken, dass wir hier von einem einzigen Signal sprechen.

Ja klar, aber Du misst diesem Signal ja zumindest eine starke Bedeutung zu. Google schläft ja nun aber auch nicht. Wie groß schätzt du die Gefahr ein, dass man dort den Bewertungsmaßstab für die Topickerkennung möglicherweise spürbar verschiebt, wenn sie bewusst und vor allem gerade von Spammern manipuliert wird? Geben wir uns keiner Illusion hin: Bevor der typische Mittelständler davon etwas mitbekommt, dass und wie er oder seine Agentur texten muss, werden die Manipulatoren und Vermittler zuschlagen.

Interessanterweise erkennt man mit etwas Erfahrung (und/oder Mathematik) am Verlauf der Termgewichtungskurve eines Textes sehr einfach, ob ein Text „natürlich“ erstellt und danach optimiert wurde oder ob ganz „gezielt“ getextet wurde. Im ersten Fall beschreibt die Termgewichtungskurve den Idealverlauf einer negativen Exponentialfunktion, im zweiten Fall wird diese insbesondere durch treppenartige Abweichungen und/oder ein auffällig niedriges Bestimmtheitsmaß geprägt.

Auch wenn es nichts direkt mit deiner Frage zu tun hat: Aus diesem Grund möchte ich Nutzern von Termgewichtungs Werkzeugen gerne nahelegen, Online-Inhalte zunächst völlig fernab von jeglichen Termen, Kurven und Werten zu erstellen und erst nach der Fertigstellung einfache Termgewichtungskorrekturen durchzuführen. Wurde ein Text bereits in der Entstehungsphase mit Blick auf „bestimmte Werte“ erstellt, kam in der Praxis bisher ausnahmslos kaum brauchbarer ASCII-Schrott heraus.

Content-Spammer sollten also schnellstmöglichst eine Weiterbildung im Bereich „Computerlinguistik“ und „automatisierte Texterstellung“ machen. Saim Alkan von aexea ist da sicher eine gute Anlaufstelle ...

Auf der SEOkomm wurdest du nach deinem Vortrag gefragt, was du glaubst, wie viele der Anwesenden das verstanden haben, was du vorgetragen hast, bzw. wie viele dir tatsächlich folgen konnten. Es gingen zwar viele Finger hoch, weil sich niemand eine Blöße geben wollte ... Denkst du, ein typischer Agenturmitarbeiter, der Webseiten für beauftragende Unternehmen erstellt, kann das System verstehen und anwenden?

Man könnte glauben, dass wir allesamt mit dem Eintritt ins Berufsleben gestresst, denkfaul und träge werden: Da verstehen die Teilnehmer die EINFACHSTEN Formeln nicht. Hallo? Das ist doch

nur ein Logarithmus!!! Vereinigungsmengenberechnung haben wir in der Grundschule mit bunten, sich überschneidenden Kreisen gemalt!

Spaß beiseite, im Ernst: Wenn mein Thema von den Teilnehmern nicht verstanden wurde, habe ich einen schlechten Job gemacht. Wenn ich durch Teilnehmer-Feedback merke, „dass ich schlecht verstanden wurde“, ist das für mich eine große Motivation, das Thema erneut so weit zu bearbeiten und zu vereinfachen, bis es auch in einem Umfeld aus Alltagsstress und multimedialer Dauerablenkung verstanden werden kann.

Ich arbeite daran und hoffe, dass „Bärbel“ ein wenig aufklären konnte.

Ja, die „Bärbel“-Darstellung hilft sicherlich ein großes Stück, die Thematik besser zu verstehen. Aber wie kann man diese Art der Optimierung denn nun in der Praxis tatsächlich anwenden? Gibt es Tools, die einem hier Hilfestellung geben? Hast du ernst zu nehmendes Feedback von anderen, dass nach der Optimierung Rankings tatsächlich spürbar angestiegen sind? Wurde das deines Wissens schon breiter eingesetzt?

Für den Praxiseinsatz wurden bereits Werkzeuge entwickelt. Mit diesen kann zum Beispiel die Termgewichtungskurve eines Textes bei Änderungen in Echtzeit angezeigt werden. Das liefert dem Texter wertvolle Informationen:

- » Was ist das unmissverständlichste Term-signal meines Dokuments?
- » Sind beweisführende Worte im angemessenen Rahmen vorhanden?
- » Werden keine unangemessen hohen Worthäufungen bzw. verfälschende Worte eingesetzt?

Die Teilnehmer des karlsCORE-Projektes nutzen solche Werkzeuge und melden aktuell ausschließlich Erfolge. Eines der vielen aktuellen Praxisbeispiele sind die Terme „Erotikkatalog“/„Erotik Katalog“ für redstore.de: Hier wurde eine bestehende Seite nur im Rahmen der Term-

gewichtungsanalyse behandelt und erzielte binnen kürzester Zeit ihr heute prominentes Ranking. Insgesamt sind mir unzählige Fälle bekannt, welche auch in hochkompetitiven Bereichen auf diese Weise exzellente Werte in der Praxis erzielen.

Ich hatte auch das Vergnügen, bei den Jungs von OnPage.org ein wenig „unter die Haube“ schauen zu dürfen. Die Ergebnisse des dort enthaltenen Termgewichtungs- und Textwerkzeugs sind aus meiner Sicht durchaus für ordentliche Optimierungen geeignet.

Ein Problem mit den Google-Richtlinien kann man eigentlich nicht bekommen, denn diese Art der Optimierung wird dort nicht erwähnt bzw. ist nicht explizit als schädlich ausgeschlossen. Solange das nur die „Guten“ anwenden, kann das Google ja eigentlich nur recht sein. Oder wie siehst du das?

Definitiv. Sobald sich ein Texter oder SEO ernsthaft mit dem Thema der Termgewichtungsanalyse beschäftigt, werden regelmäßig qualitativ hochwertigere Inhalte erstellt.

Grundsätzlich ist es für Suchmaschinen kostengünstiger, rankingrelevante Signale vom Dokument selbst zu beziehen, anstatt sich dokumentspezifische Informationen aufwendig über verlinkende Quellen zu errechnen.

Gibt es denn auch eine Art Faustformel in wenigen Worten, wie man texten sollte – oder muss – und sollte man wirklich immer die Mathematik und ein entsprechendes Tool zu Hilfe nehmen? Früher hieß es ja, der Hauptterm (also das Keyword bzw. die Keywordkombination) solle in einem Dokument gegenüber allen anderen mit deutlichem Abstand häufiger auftauchen. Praktisch gesehen also z. B. mit einer Dichte von ca. 3-4 %, wobei andere Worte und Kombinationen nicht über ein Prozent kommen sollten. Zusätzlich sollte man ähnliche und wortverwandte Begriffe

aus dem Themenumfeld verwenden. Was hat sich daran jetzt genau geändert bzw. wie würdest du das ergänzen/schärfen?

Mario! Bitte sag doch nicht mehr (Suchbegriff-)Dichte ... :-)

Ich muss dich enttäuschen, es gibt weder eine „Faustformel“, noch lässt sich diese Aufgabe ohne ein entsprechendes Analysewerkzeug durchführen.

In welchem Umfang der Hauptterm sowie beweisführende Terme im Text verwendet werden, ergibt sich aus der Zusammensetzung aller Terme im Dokument, dem daraus resultierenden Dokumentenkörper sowie den relevantesten Dokumenten für eine Suchabfrage. Änderst du auch nur ein einziges Wort in deinem Text, hast du sofort eine andere dokumentspezifische Termgewichtung sowie durch den nun andersartigen Dokumentenkörper komplett andere Voraussetzungen für die Berechnung. Spätestens jetzt sollte jeder einsehen, dass diese Methode mit der eindimensionalen Denkweise beim Einsatz der Suchbegriff-Dichte schon lange nichts mehr zu tun hat.

Dichte, Dichte, Dichte! ;-) Noch eine Frage zum Schluss: Glaubst du persönlich, dass die Gewichtung von onpage bei Google in Zukunft eher noch steigen oder in Summe sinken wird?

Meiner Meinung nach wird das Suchmaschinenoptimierungs-Pendel weiter zwischen Onsite/Onpage, Offpage, Social und Infrastruktur schwingen. In jedem Bereich eben mit unterschiedlicher Amplitude und unterschiedlicher Frequenz.

Onsite- /Onpage- und Infrastrukturverbesserungen sind sicher weiterhin gute Maßnahmen, zumal ein schneller und reibungsloser Zugriff auf Websites für Suchmaschinen grundsätzlich auch eine betriebswirtschaftliche Ersparnis bedeutet.

Vielen Dank für das Gespräch!