

Thomas Mindnich & Stefan Fischerländer

»Die zeitliche Dimension des Linkprofils

Die aktuellen Google-Updates lenken die Aufmerksamkeit der Suchmaschinenoptimierer verstärkt auf die Zusammensetzung des Linkprofils. Die üblichen Backlinkanalysen beinhalten aber oft nur einfache URL-Listen und übersehen so wichtige Aspekte des zeitlichen Verlaufs. Thomas Mindnich und Stefan Fischerländer haben sich hierzu einmal ausführliche Gedanken gemacht.

Aktuelle Google Penalties als Anlass

Ende Februar stellten viele Webmaster erst einen dramatischen Rankingverlust ihrer wohl-optimierten Website fest und fanden wenige Tage später eine Nachricht in den Google-Webmaster-Tools vor. Dort hieß es dann, es seien unnatürliche Backlinks festgestellt worden. Davon aufgeschreckt begannen viele Suchmaschinenoptimierer, sich verstärkt mit den Backlinkprofilen ihrer Sites zu beschäftigen.

Schnell wurde wieder in der Szene diskutiert, wie genau ein möglichst natürliches Linkprofil aussieht und an welchen Durchschnittswerten sich dies festmachen lässt. In den marktüblichen Tools schaut man sich dann gern die Verteilung bestimmter Parameter an und hofft, dass man möglichst nah an den allgemeinen Durchschnittswerten die Nase vorn hat und so auch Google der Webseite ein natürliches Linkprofil attestiert.

So kursieren „normale“ Verhältnisse von Deeplinks zu Startseitenlinks, von Einwort- oder Money-Keywords zu Longtail-Phrasen in Linktexten, Nofollow- zu Follow-Links, PageRank-Verteilungen und einigen weiteren leicht messbaren und auswertbaren Parametern in der internationalen SEO-Szene.

Mit vorhandenen Daten arbeiten

Unbestritten bilden Links die Basis des Rankingalgorithmus von Google. Dass es auch andere Faktoren und Merkmale für Google gibt, die zum Teil direkt im Zusammenhang mit Links stehen (z. B. nehmen echte Nutzer den Link wahr, gibt es also Traffic über den Link) und einen unterschiedlich hohen Einfluss auf das Ranking in den SERPs haben, sei für die folgenden Überlegungen einmal außer Acht gelassen.

Allerdings müssen wir Suchmaschinenoptimierer schon eingestehen, dass wir sehr gern auf einfach zu messende Parameter in unseren Optimierungsbetrachtungen zurückgreifen. War dies früher der PageRank, den man sich mit entsprechenden Links erkaufen konnte, so ist das heute eher die [IP-Popularität*](#), die zur Not über Social-Bookmarks aufgehübscht wird.

Allen diesen Parametern (die ja nicht zu selten auch als Business-[KPI*](#) herhalten müssen) ist vor allem eines gemein: Sie sind leicht zugänglich! Damit ist gemeint, dass man auch ohne die pure Größe und überragenden Möglichkeiten eines Suchmaschinenbetreibers diese Werte selbst ermitteln kann oder von Toolanbietern zugeliefert bekommt.

Listen statt Analysen

Nachteilig für unsere Analysen wirkt sich nun aber aus, dass es durchaus noch weitere relevante Metriken aus dem Linkkontext gibt, die deutlich schwerer zu erheben sind. So liefern viele Toolanbieter entweder nur unüberschaubare Linklisten oder fassen die Daten nach den einfach zu ermittelnden Parametern wie Linktext oder Nofollow-Attribut zusammen. Damit unterliegen wir bei der Analyse leicht einer Fehleinschätzung, die wir anhand einiger Diagramme zu veranschaulichen versuchen möchten. Als Ausgangssituation betrachten wir eine Website, die bislang gut und „natürlich“ verlinkt wurde. (Wir ersparen uns hier die Definition einer „natürlichen“ Verlinkung, denn von Details abgesehen, sollte jeder SEO eine Vorstellung davon haben, wie die Backlinkstruktur einer Website ohne manipulativen Linkaufbau aussieht.) Damit sind viele Links mit „natürlichen“ Linktexten vorhanden, für gewöhnlich

DER AUTOR



Thomas Mindnich ist seit 1996 in der Onlinebranche mit Fokus auf SEO tätig und geschäftsführender Gesellschafter der ITam GmbH.

DER AUTOR



Stefan Fischerländer arbeitet seit zehn Jahren als SEO-Consultant und ist Mitinhaber der SEO-Agentur Gipfelstolz in Passau.

* siehe Glossar Seite 110-112

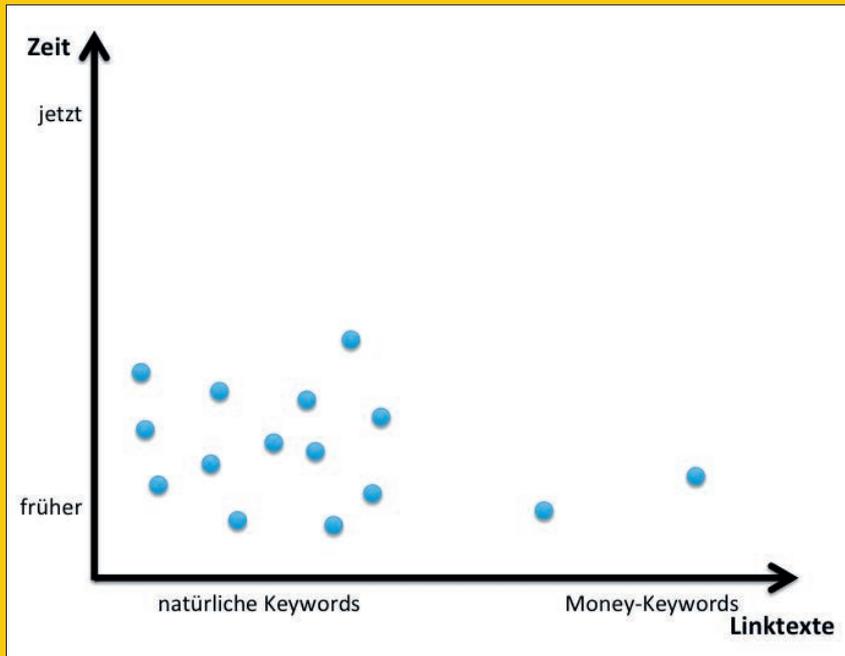


Diagramm 1: Jeder blaue Punkt steht für einen Link, der ohne Zutun eines SEOs gesetzt wurde. Nach oben ist der Zeitpunkt der Linksetzung aufgetragen, nach rechts die „Natürlichkeit“ des Linktextes.



Diagramm 2: Die gleiche Verlinkung wie in Diagramm 1, nur haben wir dieses Mal die zeitliche Komponente ausgeblendet. Diesen aggregierten Blick gewähren die üblichen Backlink-Tools.



Diagramm 3: Die roten Punkte stehen für die neu hinzugekommenen SEO-Links.

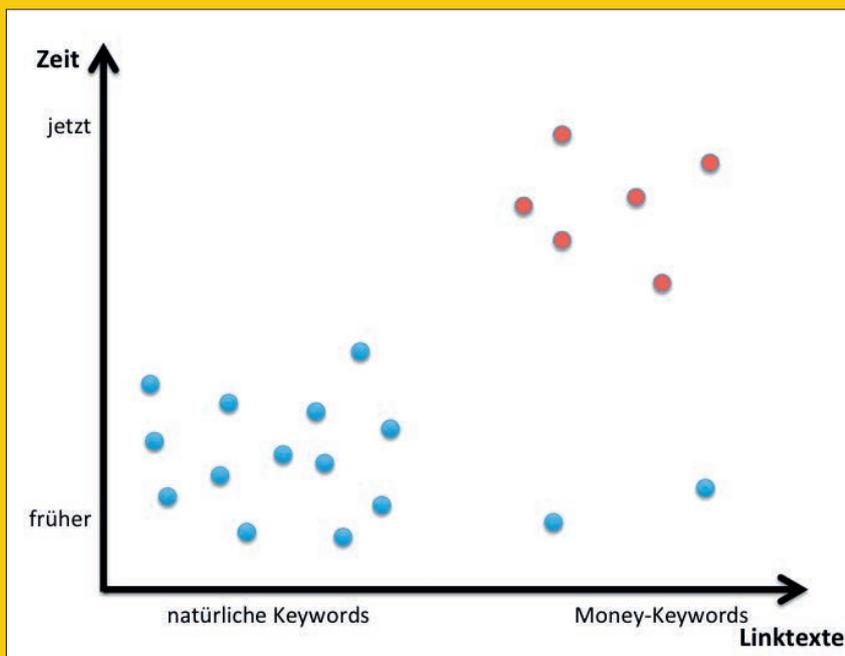


Diagramm 4: Nehmen wir die Zeitdimension mit den neuen Links wieder hinzu, so erhalten wir ein Bild, dass die Problematik deutlich zeigt.

gibt es aber kaum Links mit Money-Keywords. Diese Situation ist in Diagramm 1 dargestellt, wobei jeder blaue Punkt für einen „natürlichen“ Link steht.

Betrachten wir diese Verlinkung durch die Brille eines der üblichen Backlink-Tools, müssen wir die Zeitdimension ausblenden und erhalten eine Darstellung, wie sie Diagramm 2 zeigt. Diese aggregierte Darstellung suggeriert, dass die Website noch viele Links mit Money-Keywords im Linktext vertragen kann. Also legen wir los und bauen nun munter solche SEO-Links auf.

In der aggregierten Ansicht zeigt sich zwar, dass nun das Verhältnis von Money-Keywords zu „natürlichen“ Linktexten nicht mehr so ausgewogen ist wie zuvor, aber wir dürften in der Gesamtschau immer noch im akzeptablen Bereich liegen (Diagramm 3).

Nehmen wir aber nun im Diagramm 4 die Zeitdimension wieder zu unserer Analyse hinzu, fällt sofort ins Auge, dass alle neu gesetzten Links mit Money-Keywords versehen sind. Unsere zuvor getroffene Einschätzung, dass die Website SEO-Links vertragen könne, ist damit hinfällig, denn Google erkennt sofort, dass alle zuletzt aufgebauten Links nur der Rankingmanipulation dienen, ein klarer Verstoß gegen die Google-Richtlinien. Wir lernen somit, dass wir keine „bösen“ Links zwischen den „guten“ Links verstecken können, wenn wir nur auf die aggregierten Daten blicken und die Zeitkomponente außer Acht lassen.

Gesprächige Blogsysteme

Nachdem wir nun überzeugt sind, dass die Zeitkomponente eine gewichtige Rolle in der Linkanalyse spielt, stellt sich sofort die Frage, wie wir an diese Zeitdaten kommen. Für rein statische Seiten könnten wir den Last-Modified-Header der Serverantwort auswerten – doch wie viele Websites gibt es,

LINKQUELLE	ANZAHL	QUELL-IP	REIF	IMG	LINKTEXT	LINKZIEL	DATUM
http://www.seosweet.de/blog	1	88.198.217.131			Penguin + Panda Update von Google	website-boosting.de	17.05.12 13:52
http://www.sitefactor.de/blog	1	178.77.91.8			Zeiten des Penguin-Updates nicht pauschal schlechter bewerte ...	website-boosting.de	16.05.12 12:33
http://www.selfmarketing-online.com/blog	19	82.165.114.184			Website Boosting 2.0	website-boosting.de	16.05.12 08:55
http://www.selbstaendig-im-netz.de	95	109.239.61.143			Website Boosting	website-boosting.de	14.05.12 11:59
http://www.seokratie.de	2	188.40.85.150			Das Penguin Update von Google, der "SEO-Filter"	website-boosting.de	11.05.12 14:17
http://www.luna-park.de	1	188.93.8.214			Mario Fischer, Website-Boosting	website-boosting.de	10.05.12 13:43
http://blog.firstmedia.de	11	213.240.172.50			Website Boosting (Prof. M. Fischer)	website-boosting.de	10.05.12 08:00
http://felixbeilharz.de	2	46.4.82.204			Mario Fischer empfiehlt in seinem Beitrag	website-boosting.de	09.05.12 11:25
http://www.seoblog.net	1	84.200.180.207			Mario Fischer	website-boosting.de	09.05.12 10:30
http://www.carina-meyer-broicher.de	2	82.165.80.106			Mario Fischer in seinem erhellend	website-boosting.de	08.05.12 14:10

Abb. 1: Aktuelle Backlinks auf website-boosting.de, in der Zeitdimension dargestellt.

die noch mit statischen HTML-Seiten arbeiten? Der Brute-Force-Ansatz wäre, in sehr kurzen Abständen (zum Beispiel täglich) das ganze Web abzucrawlen. Doch dies erforderte einen gewaltigen Hardware- und Trafficaufwand und wäre keineswegs praktikabel.

Wir benötigen also eine Funktion, die uns mitteilt, wenn irgendwo im Web eine neue Seite erscheint. Und in der Tat gibt es eine solche Funktion bereits seit mehr als zehn Jahren: Blog-Systeme setzen automatisiert im Hintergrund einen sogenannten Update-Ping an einen (einstellbaren) zentralen Server ab, sobald ein Artikel veröffentlicht oder geändert wird. In der Standardinstallation von Wordpress ist dies Pingomatic (<http://pingomatic.com>), ein Service, der die Update-Pings weiterverteilt und anderen Diensten zur Verfügung stellt. Zwar erhalten wir so nur die Aktualisierungen von Sites, die auf Blogsystemen laufen, aber das ist mehr, als wir zunächst vermuteten. Neben „echten“ Blogs laufen viele Unternehmenssites, aber auch klassische Linkbuilding-Projekte wie Artikelverzeichnisse oder Bookmark-Dienste auf Blogsoftware. Zudem setzen auch eine ganze Reihe von Foren oder Zeitungswebsites Update-Pings ab.

Ein solcher Update-Ping enthält lediglich die URL einer Website, die von sich behauptet, neue Inhalte zu haben. Um diese neuen Inhalte zu erhalten, muss man die angegebene URL besuchen und die neuen Inhalte selbst finden. Da fast immer ein RSS-Feed auf pingenden Websites bereitgestellt wird, lassen sich aber die neuen Inhalte schnell entdecken. Auf unserem Testsystem itam.net (www.itam.net) brauchen wir selten länger als fünfzehn Minuten, bis wir den Inhalt und die Links einer neu erstellten oder aktualisierten Seite in der Datenbank gespeichert haben. Schneller sind lediglich die diversen Googlebots, die meist schon nach wenigen Sekunden auftauchen. Einen Blick in eine derartige Datenbank sehen Sie in Abbildung 1.

Natürlich greift auch Google längst auf diese Ping*-Daten zu und kann damit viele Analysen durchführen. Da eine neue Seite sofort nach Veröffentlichung gecrawlt wird, lassen sich beispielsweise später hinzugekommene Links leicht identifizieren. Die gängige Praxis, thematisch passende Blogbeiträge zu suchen und dort vom Blogbetreiber einen Link unterbringen zu lassen, kann Google also problemlos erkennen. Umgekehrt wird Google fest-

stellen, ob etwa als Blogkommentare gesetzte Links vom Betreiber wieder entfernt wurden. Das ist bedeutend, schließlich sind gelöschte Links ein besonders drastisches Signal, dass dieser Link unerwünscht war. Zudem lassen sich site-wide gesetzte Links leicht erkennen, ohne dass auch nur eine Zeile HTML-Code geparkt (d. h. maschinell analysiert) werden müsste: Diese Links fallen schlicht und ergreifend durch ihren identischen Linktext und das stets gleiche Linkziel auf.

Diese nicht immer gewollte Transparenz im Linkbuilding wird noch dadurch vergrößert, dass viele professionelle Linknetze auf Blogsystemen, meist auf Wordpress, laufen. In den letzten Jahren gab es kaum eine SEO-Konferenz, auf der nicht wenigstens ein Vortrag zeigte, wie sich mit Wordpress eine ganze Armada an Linktausch-Blogs aufsetzen lässt. Das bedeutet aber nicht weniger als die mehr oder weniger komplette Transparenz der klassischen Linkkauf-Ökonomie!

Real-Time-Konkurrenzanalyse

Doch nicht nur Linkkauf wird transparent, auch weitere Linkbuilding-Strategien sind leicht zu überwachen. So zeigt ein Blick auf die aktuell gesetzten

* siehe Glossar Seite 110-112

LINKQUELLE	ANZAHL	QUELL-IP	REF	IMG	LINKTEXT	LINKZIEL	DATUM
http://amg-personal.blog-hr.de	46	82.165.109.36	nf		Thorsten		18.05.12 09:00
http://www.e-five.de	1	78.46.122.207			mit dem Anbieter der Mitarbeiterausweise Kontakt aufnehmen		14.05.12 22:37
http://www.blog-n-biz.de	12	109.237.138.22			Kartendrucker Shop		01.05.12 00:56
http://artikelhannes.de	1	46.59.90.233			Ausweisdrucker		18.04.12 10:26
http://artikelgeier.de	1	46.59.95.245			Kartendrucker		18.04.12 09:57
http://www.reputare.de	1	93.158.64.134			Kartendrucker		11.04.12 15:13
http://artikel-checker.de	1	91.211.114.124			Ausweisdrucker		04.04.12 15:17
http://www.shop-blog24.de	2	89.144.4.41			Kartendrucker		03.04.12 10:49
http://www.meinungsplanet.de	1	85.13.146.101			Plastikkarte		28.03.12 18:51
http://www.technik-portal24.de	1	89.144.7.25			Kartendrucker		28.03.12 00:02

Abb. 2: Die Linktexte von offensichtlich gezielt aufgebauten Backlinks gewähren deutliche Einblicke in die Keywords, die diese Website besonders wichtig findet – perfekt für die zeitnahe Konkurrenzanalyse.

LINKQUELLE	ANZAHL	QUELL-IP	REF	IMG	LINKTEXT	LINKZIEL	DATUM
http://www.presse-artikel.at	5	212.41.237.18			Die Hauptfrage ist , wo die Hochzeit zelebriert werden soll		16.05.12 08:05
http://www.blogverzeichnis24.de	1	62.112.144.181			Die Auswahl für den Ort der Hochzeit sollte rechtzeitig getr ...		14.05.12 09:22
http://www.presstext-redaktion.at	4	212.41.237.20			Möglicherweise eine klassische Hochzeit samt kirchlicher Tra ...		14.05.12 08:32
http://online-presstext-publisher.at	2	212.41.237.22			eine traditionelle Hochzeit mit Trauung in der Kirche – ankn ...		10.05.12 08:25
http://www.artikelmaschine.de	1	62.241.40.136			klassische Hochzeit		30.03.12 14:35
http://www.artikel-blogger.de	1	62.241.43.11			Hochzeit		30.03.12 14:32
http://artikel-platz.de	1	91.211.115.124			Hochzeit		30.03.12 14:32
http://artikel-portal24.de	1	91.211.112.124			Hochzeit		25.03.12 19:51
http://artikel-checker.de	1	91.211.114.124			Hochzeit		25.03.12 17:21

Abb. 3: Während bei dieser Site im März noch Keyword-Links dominierten, weisen die zuletzt gesetzten Links längere und damit vielleicht „natürlichere“ Linktexte auf.

Backlinks der Konkurrenzsites, welche Keywords diese Konkurrenz derzeit besonders pushen möchte (Abbildung 2). Auch Linkbaits werden so schnell sichtbar und lassen sich eventuell für den eigenen Linkaufbau nutzen. Manchmal ist es sogar möglich, die von der Konkurrenz entdeckte Backlinkquelle ebenfalls anzupapfen, etwa bei Links, die in einer Forumdiskussion gesetzt wurden.

Googles trickreiche Zeitangaben

Hier stellt sich die Frage, wie denn Google eigentlich an seine Zeitangaben kommt, die manchmal zu einzelnen Treffern angezeigt werden. Den ersten

Verdacht, es handele sich dabei nur um den Crawlzeitpunkt, konnten wir nicht bestätigen. Vielmehr bedient sich Google offenbar fleißig an Datumsangaben im Inhalt der Seite. Deutlich macht dies eine Suche, die auf den 1. Januar 1970 eingeschränkt ist. Dieses Datum entspricht dem Zeitpunkt null in Unix-Systemen und wird bei unsauberer Programmierung oft als Standard-Datum gesetzt, so lange keine anderweitigen Angaben gemacht werden.

Entsprechend viele Treffer zeigt uns Google für diesen Tag an, wie in Abbildung 4 zu sehen ist. Dass es sich dabei nicht um das Crawldatum handeln kann, ist offensichtlich – wurde

„Google bedient sich offenbar fleißig an Datumsangaben im Inhalt einer Seite“

Suche Ungefähr 1.120.000 Ergebnisse (0,24 Sekunden)

Alles 1. Jan 1970

Bilder [Universität Konstanz | Intranet](#)
[printservice.uni-konstanz.de/startseite/](#)

Maps 1. Jan. 1970 - **Universität Konstanz** · Login | · Deutsch; | English; | Français; | Español; | Russian; | Chinese; | Polski; | Türkçe. Home · Impressum · Intranet · Planungs- und ...

Videos

News [Zugangsberechtigungen zu DV-Anlagen - Universität Regensburg](#)

Shopping [www.uni-regensburg.de/Einrichtungen/Rechenzentrum/.../Accounts/](#)

Mehr 1. Jan. 1970 - An der **Universität** gibt es zentral und dezentral viele verschiedene Rechner und Rechnersysteme. Will man eines dieser Systeme nutzen, so benötigt man dafür ...

Abb. 4: Ergebnis für die Google-Suche nach [uni] und der Einschränkung, nur Treffer für den 1. Januar 1970 anzuzeigen.

SQL-Befehl:

```
SELECT to_site, GROUP_CONCAT(DISTINCT linktext), COUNT(DISTINCT linktext) AS anz
FROM links, sites, posts
WHERE from_post_id = posts.id AND sites.id = links.from_site_id
AND linktext in ('Ben', 'Leon', 'Lukas', 'Jonas', 'Max', 'Maximilian', 'Luis', 'Paul', 'Felix', 'Moritz', 'Alexander', 'Thomas', 'Stefan', 'Jan', 'Jakob', 'Tom', 'Niklas', 'Alex', 'Axel', 'Eric', 'Fabian', 'David', 'Tim', 'Timm', 'Daniel', 'Tobias', 'Boris', 'Theo', 'Christian', 'Tobi', 'Jörn', 'Achim', 'Markus', 'Marcus', 'Phillip', 'Julian', 'Sasha', 'Sascha', 'Sven', 'Klaus', 'Josef', 'Joseph', 'Horst', 'Dominik', 'Maik', 'Michael', 'Uwe', 'Sebastian', 'Dennis', 'Florian', 'Marcel', 'Robert', 'Johannes', 'Chris', 'Anna', 'Lea', 'Sarah', 'Hannah', 'Laura', 'Lisa', 'Lena', 'Lara', 'Anne', 'Anni', 'Maria', 'Julia', 'Johanna', 'Sabine', 'Kerstin', 'Marie', 'Katharina', 'Sophie', 'Antonia', 'Alina', 'Jasmin', 'Nina', 'Emma', 'Luisa', 'Alexandra', 'Sophia', [...])
```

Zeige: 30 Datensätze, beginnend ab 30

untereinander angeordnet und wiederhole die Kopfzeilen nach 100 Datensätzen.

to_site	GROUP_CONCAT(DISTINCT linktext)	anz
www.f...n.de	Maik, Sandra, Uwe, Markus, Sasha, Lisa, Klaus, Julian, Ach...	26
twitter.com	Sebastian, Christian, Uwe, Michael, Jan, alexander, Hann...	19
www.facebook.com	Tim, Joseph, Sven, Lara, Sebastian, Christian, markus, J...	18

Abb. 5: Wenn eine Finanzwebsite innerhalb weniger Tage mit 26 verschiedenen Vornamen verlinkt wird, erscheint das recht ungewöhnlich.

Mahnungen im Grunde für alle Parameter, mit denen sich ein Link charakterisieren lässt. So haben wir beispielsweise untersucht, welche Websites besonders häufig mit unterschiedlichen Vornamen verlinkt wurden. (Abbildung 5) Dass soziale Netzwerke hier besonders oft vertreten sind, war zu erwarten und ist Ergebnis natürlicher Verlinkung. Aber in unserer Stichprobe, die nur Daten weniger Wochen enthielt, bekam eine Finanzwebsite Links mit 26 verschiedenen Vornamen, deutlich mehr als Twitter oder Facebook. Ein klares Zeichen, dass die dortigen SEOs intensiv auf Blog-Kommentare setzen.

doch das Web erst mehr als zwei Jahrzehnte später entwickelt. Und es zeigt, wie schlampig im Web häufig programmiert wird...

Gegenmaßnahmen

Wir haben gezeigt, wie die Update-Pings der Blogs es ermöglichen, extrem zielgenau und sparsam zu crawlen und trotzdem viele Links zeitnah zu finden. Wer nun die dadurch ermöglichte große Transparenz nicht zu schätzen weiß, sollte in seinen Linknetz-WordPress-Blogs unbedingt den Update-Service ausschalten. Zudem ist ein Blick in die Liste der eingesetzten Plug-ins sinnvoll, denn es gibt noch weitere Me-

chanismen, die ähnlich wie die Update-Pings funktionieren. So stellt etwas Googles PubSubHubbub (der Name stimmt tatsächlich, zu finden unter <http://einfach.st/pubsub>) eine Erweiterung der klassischen RSS-Feeds mit Push-Funktionalität dar. Alle auf wordpress.com gehosteten Blogs etwa informieren bei Änderungen seit 2010 über dieses Protokoll.

Die Zeitdimension ist nur ein Beispiel!

Der Schwerpunkt unserer Betrachtung lag auf der zeitlichen Dimension, die bislang in der Linkanalyse eher vernachlässigt wurde. Doch gelten unsere

Expertenblick aufs Linkprofil

Bislang haben wir uns mit den relativ einfachen Linkanalysen der SEO-Tools nicht aus Bequemlichkeit zufriedengegeben, sondern weil tiefere Einblicke zum Verständnis der Google-Rankings kaum nötig waren. Die Update-Orgie der zurückliegenden zwölf Monate macht aber deutlich, dass Google die schier unendlichen Daten seines Linkgraphen immer detaillierter auszuwerten scheint. Das wiederum zwingt Suchmaschinenoptimierer dazu, den Linkaufbau noch stärker zu professionalisieren und damit einhergehend die Anforderungen an die Linkanalyse-Tools nach oben zu schrauben.¶