



SEO SATT – DIE SEO CAMPIXX 2016 IN BERLIN

In der zweiten Märzwoche trafen sich in Berlin wie jedes Jahr Suchmaschinenoptimierer und solche, die es werden möchten. Während der sog. Campixx:Week gab es tagesbezogen weitere Online-Marketing-Themen um die Ohren und am Wochenende startete dann die sicherlich von der Art her einzigartige und mittlerweile als legendär zu bezeichnende SEO Campixx. 550 Teilnehmer und Teilnehmerinnen tauschten sich in über 110 Fachvorträgen in zwölf parallelen Tracks aus. Der Veranstalter hatte traditionsgemäß genügend Luft zwischen den Slots gelassen, sodass jeweils genügend Zeit für viele interessante Gespräche und Networking zur Verfügung stand. Wie immer waren alle Tickets kurz nach dem Opening der Buchungseingabe weg und wer je auf der SEO Campixx war, weiß auch, warum das so ist. Website Boosting hatte sich vorgenommen, wichtige Essenzen für alle diejenigen zu filtern, die keine Zeit hatten oder kein Ticket mehr bekamen. Bei diesmal über 40 Seiten handschriftlicher Notizen kein leichtes Unterfangen - schließlich ist der Platz hier vergleichsweise arg begrenzt. Ein Versuch.

Das Programm der allgemein gern auch als „SEO-Klassentreffen“ bezeichneten Unkonferenz war erneut so breit gefächert, dass für jeden genügend dabei war, um etwas Neues zu lernen, vorhandenes Wissen aufzufrischen oder sich inspirieren zu lassen, welche Methoden, Vorgehensweisen oder Tools man im eigenen Unternehmen anwenden bzw. ausprobieren könnte.

Wisdom of the Nerds

Angenehm erschreckend ist mittlerweile, wie viele noch sehr junge Menschen sich im Bereich der professionellen Datenanalyse durch eigene Leidenschaft so extrem gut auskennen, dass sich die Experten mancher größerer Unternehmen davon sicher mehrere Scheiben abschneiden könnten. Da werden just for fun Analysen im Umfang mehrerer Terabyte nur um der Erkenntnis willen in wenigen Sekunden oder Minuten durchgerechnet oder linguistische Modelle in Algorithmen umgesetzt. Schnell mal einige Hundert Server oder Proxies mieten, Speicherplatz und die Power von Hunderten Prozessoren und Hauptspeicher im hohen zweistelligen Gigabytebereich nutzen? Kein Problem für einige der versierten Daten-Nerds. Klick, klick, klick – und schon laufen Programme auf Cloudsystemen in einem Umfang, die bei einem traditionellen Rechenzentrumsleiter sicherlich eine hochgezogene Augenbraue verursachen würden und ggf. längeres Nachdenken. Zehn Terabyte Bewegungsdaten einer Website, alle Daten aus AdWords, der Google-Webmasterkonsole und dem Tag-Manager erst mit OnPage-SEO-Analysedaten und dann allen Werten aus den internen Shop- und Fakturierungsdaten anreichern? So ein Projekt würde bei Unternehmen vom Schlage Siemens wahrscheinlich eine Laufzeit von über einem Jahr benötigen und mehrere Genehmigungsinstanzen beschäftigen – von den Kosten ganz zu schweigen. Die Daten-Nerds erledigen das heute im günstigsten Fall in wenigen Stunden oder Tagen. Dabei entsteht eine Transparenz für notwendige Erkenntnisse, die ihresgleichen sucht. Was an höheren mathematischen Kennt-



Abb. 1: Stefan Fischerländer über Trends für 2016 und eigene Analyseergebnisse

nissen benötigt wird, eignet man sich durch Selbststudium und Ausprobieren an – eben weil man es an einigen Stellen in der Praxis einfach benötigt. Und eben nicht, weil ein Lehrer oder Dozent es theoretisch und meist ohne konkreten Anwendungsfall in der Schule oder der Universität erklärte und in einer Prüfung verlangte. Und wenn es an einigen Stellen einmal doch nicht weitergeht, hilft man sich untereinander offen und bereitwillig aus. Hier greift statt wisdom oft he crowds eher wisdom of the Nerds. Solche hoch agilen Verbünde sind von Mitarbeitern in normalen Unternehmen nicht mehr zu schlagen. Moderne Start-ups sind

voll von solchen Tatendrängern und den oft zwar gut ausgebildeten, aber prozessual und strukturell gebundenen Normalo-Mitarbeitern in Normalo-Unternehmen daher oft weiter voraus, als manch einer sich das eingestehen will. Kein Wunder, dass es die echten Highflyer auch nicht gegen übertarifliche Bezahlung in traditionelle Unternehmen zieht. Wie das mittelfristig ausgeht, werden wir sehen.

SEO 2016 – erstellt Google Algorithmen noch manuell?

Stefan Fischerländer weiß, wovon er spricht. Schließlich ist er einer der wenigen, die in der Vergangenheit tatsächlich eine eigene Suchmaschine gebaut haben (Neomo). Er zeigte den Zuhörern auf, warum Trickereien beim SEO künftig immer weniger funktionieren werden. Schon seit vielen Jahren kam der Einsatz von sog. Machine Learning prinzipiell zu besseren Ergebnissen als rein manuell programmierte Such- bzw. Bewertungsalgorithmen, so Fischerländer. Googles „... best machine-learned model is now as good as, and sometimes better than, the hand-tuned formula on the results quality metrics that Google uses“, zitierte er Peter Norvig, Director of Search bei Google. Das war im Jahre 2008, und wer verfolgte, wie rasant sich die Technologien seither entwickelten, kann erahnen, was hier heute bereits möglich ist. Die Möglichkeiten der automatisierten Klassifizierung und einer feingliedrigen Clusterbildung

„Versucht gar nicht erst, Googles Analyse des Nutzerverhaltens auszutricksen“, Stefan Fischerländer

schreiten mit hoher Geschwindigkeit voran.

Fischerländer interessierte, wie sich aktuell die Suchergebnisse auf Smartphones von denen eines Desktops unterscheiden. Mit einer Stichprobe von 10.000 Anfragen hatte er herausgemessen, dass es mit 92 % eine noch recht hohe Übereinstimmung bei den Top 100 gibt. Dabei fiel ihm auf, dass die Anzahl der unterschiedlichen Domains in den Suchergebnissen auf iPhone, iPad und Desktop in etwa gleich waren, die auf Android-Geräten (hier dem Nexus 5) jedoch deutlich abweichen. Der Grund dafür war noch nicht abschließend zu klären, aber er versprach, dranzubleiben und weiter darüber zu berichten.

Er wies weiter darauf hin, dass das Thema App-Indexing bereits große Fortschritte gemacht habe. Dabei werden auf einem Smartphone auch die Inhalte der installierten (!) Apps mit durchsucht und in den Suchergebnissen des betroffenen Geräts mit ausgegeben. Laut Fischerländer seien solche Ergebnisse aus Apps bereits in 40 % der Websuchen auf Android-Geräten mit enthalten. Keine oder keine durchsuchbare App zu haben, kann für ein Unternehmen daher möglicherweise bereits einen spürbaren Wettbewerbsnachteil darstellen. Die Selbstverständlichkeit der fast nur noch mobilen Webnutzung könnte durchaus mit den nachkommenden Generationen steigen. Bereits heute haben 26 % der unter 13-jährigen nach eigenen Angaben einen eigenen YouTube-Account und 21 % nutzen mobil soziale Netzwerke. Unter 13 Jahre – das sollte weder überlesen noch unterschätzt werden.

Und während einige ständig versuchen, auf dem neuesten Stand hinsichtlich SEO zu bleiben, machen große Unternehmen noch immer nicht die allernötigsten Hausaufgaben. Fischerländer zeigte dazu einige recht amüsante, aber auch erschreckende

Original Version	Anonymised Version
Context The BBC producer allegedly struck by Jeremy Clarkson will not press charges against the "Top Gear" host, his lawyer said Friday. Clarkson, who hosted one of the most-watched television shows in the world, was dropped by the BBC Wednesday after an internal investigation by the British broadcaster found he had subjected producer Oisin Tymon "to an unprovoked physical and verbal attack." ...	the <i>ent581</i> producer allegedly struck by <i>ent512</i> will not press charges against the " <i>ent152</i> " host, his lawyer said Friday. <i>ent512</i> , who hosted one of the most - watched television shows in the world, was dropped by the <i>ent581</i> Wednesday after an internal investigation by the <i>ent150</i> broadcaster found he had subjected producer <i>ent193</i> "to an unprovoked physical and verbal attack." ...
Query Producer X will not press charges against Jeremy Clarkson, his lawyer says.	producer X will not press charges against <i>ent512</i> , his lawyer says.
Answer Oisin Tymon	<i>ent193</i>

Abb. 2: Wie Maschinen Antworten aus Kontextinformationen generieren (Forschungspaper Teaching Machines to Read and Comprehend, Nov. 2015 von Google, <http://einfach.st/gcon8>)



Abb. 3: Beispiel BMW: Die zentralen Startseiten der Modelle aus früheren Jahren werden einfach gelöscht

Beispiele auf. Gerade die deutschen Autobauer tun sich offenbar immer noch schwer, ihren Kunden notwendige Informationen zur Verfügung zu stellen. Da ranken z. B. Pressemeldungen für Fahrzeuge besser als die eigentlichen Produktseiten. Oder zentrale Seiten werden wie bei BMW einfach gelöscht, statt sie für Suchmaschinen und Menschen auf aktuelle Versionen weiterzuleiten. Hier wird nicht nur aus SEO-Gesichtspunkten unnötig Potenzial verschenkt.

Blogspring: Ein weiteres Spreadsheet-Tool

Thomas Nething stellte mit Blockspring (www.blockspring.com) ein weiteres Tool zur Datenanreicherung für MS-Excel und Google Sheets vor. Im Kern funktioniert Blockspring ähnlich wie das deutlich bekanntere „SEOtools für Excel“ von Niels Bosma (seotoolsforexcel.com). Über ein Plug-in lassen sich die APIs (Datenschnittstel-

len) vieler bekannter Dienste, so u. a. Alexa, Amazon, Doubleclick, Facebook, Foresquare, Google Analytics und Webmasterconsole, Instagram, LinkedIn, Mailchimp, MOZ, Salesforce, SEMrush, SimilarWeb, Twitter, Wikipedia, You-

„Natürliche Sprachanfragen an Smartphones werden vor allem wegen des automatisch bekannten räumlichen und zeitlichen Kontexts weniger und andere Keywords enthalten als bisher.“



Abb. 4: Das Integrationstool Blockspring hält viele API-Schnittstellen zu bekannten Diensten vor



Abb. 5: SEO-Experten im Gespräch (v. l. n. r: Niels Dahnke, Björn Tantau, Moderator Florian Stelzner, Gero von Wenderholm, Florian Elbers)

Tube und viele andere ohne Programmierkenntnisse anzapfen und die entsprechenden Daten in Excel oder auch gleich online in Google Spreadsheets übertragen. Angelegte Verbindungen lassen sich dann manuell oder permanent automatisiert aktualisieren. Der Programmierer von Blockspring kündigte an, noch weitere Schnittstellen zu integrieren. Seit Neuestem ist auch eine Verbindung mit Tableau, Slack und Bubble möglich. Auf der Website sind auch einige Beispieltemplates zum Ausprobieren bzw. als Starthilfe verfügbar. Laut Nothing erlaubt die Free-Version bis zu 100 Abfragen pro Minute. Eine Team- (7 US\$/Monat, bis 100.000

Abfragen/Monat) und eine Enterprise-Variante (50 US\$, mit zusätzlichem Support innerhalb von zwei Stunden) stehen ebenfalls zu Verfügung, wenn auch etwas versteckt unter „about“ in der Fußzeile. Pro Account erhöht sich der Preis entsprechend.

Bei der Anzahl und Fülle der Änderungsfrequenz der API-Strukturen der Datenplattformen bleibt es allerdings nicht aus, dass der eine oder andere Datenzugang nicht mehr funktioniert. Vor dem Einsatz sollte man sich daher der Funktionsfähigkeit der gewünschten APIs versichern und bei Änderungen auf kleinere Wartezeiten gefasst machen. Wer sensibel mit dem Thema

Datenschutz umgeht, muss sich auch vergegenwärtigen, dass alle über solche Schnittstellen gezogenen Daten über die USA laufen und dies fallweise nach deutschem Recht oder auch durch individuelle Kundenvereinbarungen kritisch hinterfragt werden sollte.

„Wir müssen reden“ – Expertentreff

Mit diesem doppeldeutigen Sessiontitel moderierte Florian Stelzner ein Expertenpanel mit Björn Tantau, Niels Dahnke, Gero von Wenderholm und Florian Elbers. Über das aktuelle Thema AMP (Accelerated Mobile Pages, www.ampproject.org) kam die Diskussion schnell auf den PageSpeed. Seiten können nach allgemeiner Ansicht gerade für den mobilen Abruf gar nicht schnell genug laden. Niels Dahnke betonte, dass man dabei aber immer auch auf das Gesamtprojekt schauen müsse. Statt Server schneller zu machen, reiche es z. B. oft schon aus, üppige Bilder – oft mehrere in einem Slider – auf Startseiten zu entschlacken, um die Ladegeschwindigkeiten spürbar zu erhöhen. Bei großen Websites reicht das wegen der Komplexität oft nicht. Dort sind dagegen meist die Entscheidungswege für Veränderungen deutlich zu lang. Ein Panelteilnehmer erwähnte, dass es bei einem Ticket für einen Change Request einmal sogar ganze vier Jahre gedauert hätte, bis eine Antwort kam.

Das ist sicher eine unruhliche Ausnahme, aber ganz generell tut man sich in größeren Organisationen, so die Meinung der Experten, oft sehr schwer, notwendige Arbeiten auf die Prioritätenlisten zu bekommen. Aber gerade dort ist das Thema Ladezeit extrem wichtig, denn die Häufigkeit, mit der z. B. Google die einzelnen Seiten crawlt bzw. auf Aktualisierungen prüft, hängt nicht unerheblich von dieser Kennzahl ab. Je länger es für Google dauert, Seiten zu holen, desto weniger werden pro

Zeitspanne auch tatsächlich abgerufen. Die Folge: Trotz geänderter Inhalte ranken Seiten noch mit alten Inhalten oder sie ranken nach einer Optimierung eben erst stark zeitverzögert. Zudem werden in Summe weniger Seiten für den Index abgeholt, als möglich wäre.

Recht oft ist auch eine gewisse Beratungsresistenz bei Kunden anzutreffen, darüber waren sich alle Experten einig. Je höher man in der Hierarchie kommt, desto schwieriger wird auch die Kommunikation. Manager können mit Begriffen wie Ladezeit oder anderen technischen Begriffen nicht so viel anfangen oder schalten nach kurzer Zeit gedanklich auf Durchzug. Diskutiert der SEO mit dem Webdesigner, kann es ähnliche Probleme geben. Während der erste wegen bestimmter technischer Probleme Bauchschmerzen hat, muss der Webdesigner das operativ umsetzen bzw. die Arbeit machen. Beim Erfolg wird der SEO „gefeiert“, aber derjenige, der die Arbeit machte, bleibt oft ungenannt im Hintergrund. Das kann durchaus frustrieren und in eine eher ablehnende Haltung münden. Die Beratungsresistenz nehme spürbar zu, so Tantau.

Niels Dahnke erläuterte, dass eine ablehnende Haltung gegenüber SEO-induzierten Änderungen auch manchmal mit dem Argument gestützt werde, es ändere sich dauernd etwas bei Google und man könne daher nicht ständig nacharbeiten. Dem entgegnete er, dass es meist aber um echte Basics gehe. Daher habe er sich einige alte Exemplare der ersten Auflage des SEO-Buchs von Sebastian Erlhofer über ebay besorgt und zeige den Betroffenen dann jeweils, dass die geforderten Anpassungen nicht etwas neu seien, sondern zum Teil als Basics schon vor acht Jahren bekannt waren. Wichtig sei vor allem, die Begriffe aus dem jeweiligen Umfeld der Gesprächspartner zu kennen und aktiv zu verwenden. Einem Kaufmann gegenüber müsse

„Die Leute sind oft so überzeugt von ihren Inhalten, dass sie gar nicht auf die Idee kommen, dass man hier noch etwas verbessern könnte.“

man eben eher mit Kosten und Erlösen argumentieren, nicht mit Klicks, Impressions, Verweildauer oder gar Sichtbarkeitsindizes. Manche Menschen in Unternehmen mögen auch das Wort „Problem“ nicht. Dahnke spricht daher oft lieber davon, Wachstumshemmnisse zu beseitigen. Selbst die Empfänger einer E-Mail in der falschen Reihenfolge zu adressieren, könne ggf. schon zu Widerständen führen. SEOs müssen sich daher wohl noch stärker auf die Bedürfnisse, die Denkwelt und -art und die Begriffe ihrer Kunden einstellen, wenn sie erfolgreich(er) werden wollen. Oft hilft es auch enorm, die tatsächlichen Umsetzer nicht zu vergessen, bewusst zu loben und auch danach die nötige Wertschätzung zu zeigen.

Florian Elbers wies im Übrigen noch darauf hin, dass laut Google das AMP-Projekt künftig auf weitere Contentarten (bisher nur News) ausgeweitet werden soll, so z. B. auf Rezepte und Videos. Dies sollte man sicherlich im Auge behalten, sofern man hiervon betroffen ist bzw. profitieren kann.

SEO-Live-Audits

Marcus Tandler und Danny Linden gaben in ihrer Session gute Einblicke, wie man mit SEO-Tools Rankingprobleme bei Websites aufspüren kann. Dabei wurde unter anderem deutlich, dass viele Sites noch immer zu viel Ballast mit sich herumschleppen. Der Großteil an Webseiten ist veraltet, schlecht gepflegt oder hatte noch niemals bzw. viel zu wenige Besucher. Entgegen dem alten SEO-Mantra „Mehr ist besser“ solle man solche alten Zöpfe in der heutigen Zeit besser abschneiden, so Tandler. Dabei stelle man sich am besten drei Fragen: Brauche ich diese Seite für den Besucher? Brauche ich sie für den Index? Brauche ich sie für das Ranking? Bei ehrlicher Verneinung sei es besser, „altes Fett“ wegzuschneiden und damit das Crawlbudget für Google besser zu nutzen. Zudem verwässern solche Seiten oft die thematische



Abb. 6: Hochkonzentriert bei der Arbeit: Danny Linden und Marcus Tandler

„Einmal mit dem Stahlbesen durchkehren!“, Marcus Tandler

Schärfe der für ein gutes Ranking wichtigen Seiten. Bei einer der analysierten Domains kamen 1.400 Seiten zutage, für die zu erreichen man sieben oder mehr Klicks braucht. Sie bekommen weder intern noch extern Traffic. Hier muss man tatsächlich die Frage nach der Existenzberechtigung stellen.

Oft schlägt bei solchen Sites auch der sog. Slow Death zu (Abbildung 7). Im Zeitverlauf verschlechtert sich das gesamte Ranking einer Domain durch Panda-Algorithmen, die nicht nur an bestimmten Update-Tagen ihr Aufräumarbeit im Google-Index verrichten, so Tandler. Hier solle man als Erstes die Webanalyticsdaten eines gesamten Jahres hernehmen und alles „mit dem Stahlbesen auskehren“, was nicht wirklich nötig ist bzw. keine Besucher abbekommt. Oft entstehe auch durch Blogs und Foren bzw. dort hinterlassene Kommentare so viel „user generated Mist“



Abb. 7: Slow Death – ein typisches Sterbemuster ungepflegter Sites (Quelle: Sistrix)

(O-Ton Tandler), dass Google immer misstrauischer werde, was die qualitative Beurteilung angehe.

Eigene Analysen durchführen

Mario Träger zeigte mittels der Tools „Screaming Frog“ (siehe auch die letzte Ausgabe 36 der Website Boosting) und „URL-Profiler“ auf, wie man wichtige SEO-Fragen für die eigene Website kostengünstig selbst durchführen kann.

Das SEO-Tool Screaming Frog verfügt über einen sehr leistungsfähigen Crawler, dessen Stärke nicht nur im Einholen aller Seiten einer Domain liegt, sondern vor allem in der benutzerspezifischen Parametrisierbarkeit. Träger zeigte dies u. a. an dem Beispiel des Aufspürens doppelter Inhalte in Produkttexten eines Shops. Das Kernproblem vieler Tools liege hier darin,

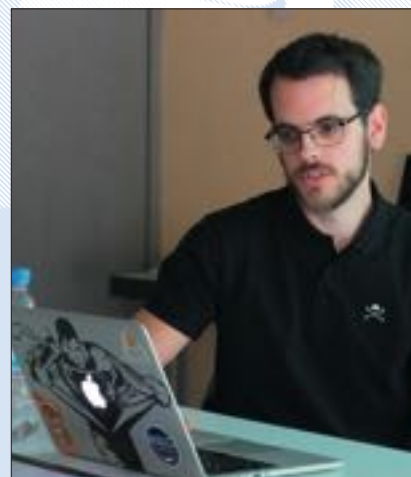


Abb. 8: Mario Träger gab einen guten Einblick, wie man Fehler auf eigenen Websites aufspürt

dass meist der gesamte Text einer Seite verglichen werde, also eben auch Footertexte oder andere Textkategorien, und damit der Vergleich des reinen Produkttextes verwässert werde.

Hierzu öffnet man den Browser (hier Chrome) und sucht eine typische Produktseite. Dann öffnet man die Entwicklerkonsole (ganz rechts oben im Browser, das Menü mit den drei waagerechten Strichen, im Aufklappenmenü dann „weitere Tools“/„Entwicklertools“) und sucht den Abschnitt mit der Produktbeschreibung. Im Codefenster unten unter „Elements“ lässt sich z. B. über ein eindeutiges Suchwort der div-Container finden. Im Beispiel in Abbildung 9 ist dieser mit `<div class=`“description_block“> ausgezeichnet. Über diese Zeile setzt man nun einfach den Mauszeiger und kopiert via rechte Maustaste (Markierung 1) den sog. XPath.

Jetzt öffnet man den Screaming Frog und ruft über das Menü „Configuration“/„Custom“/„Extraction (Markierung 2) das Formular auf. Man wählt als Methode „XPath“ im Pull-down-Menü und kopiert aus der Zwischenablage

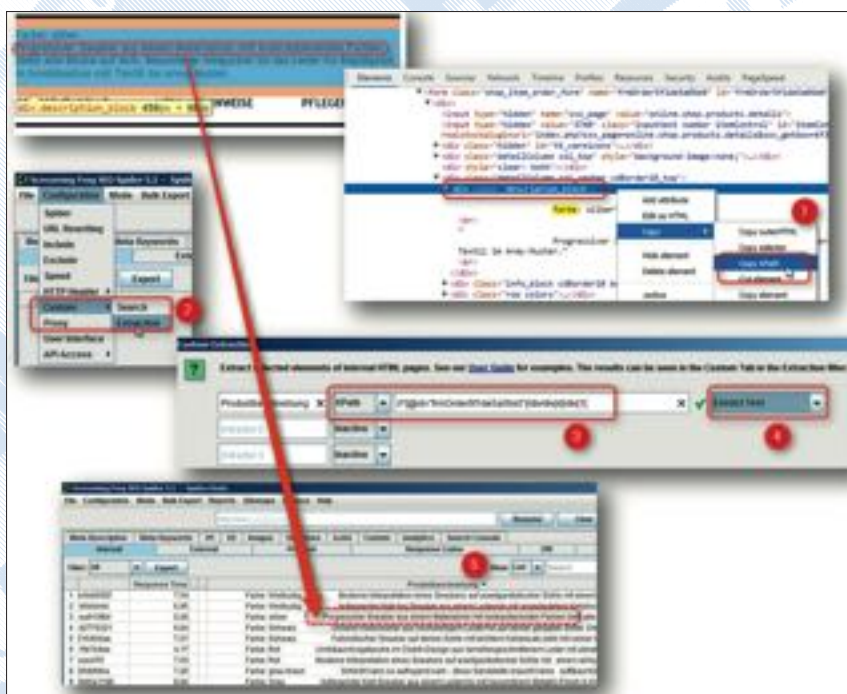


Abb. 9: Echte Produkttextdubletten mit Screaming Frog finden

den XPath in das Formularfeld (Markierung 3). Anschließend wählt man noch im rechten Pull-down „Extract Text“ (Markierung 4) aus. Dies weist Screaming Frog an, aus allen Seiten neben den anderen Kennzahlen eben genau den textlichen Inhalt dieses Pfades bzw. div-Containers zusätzlich mit auszulesen. Anschließend startet man das Tool nach Eingabe des Domainnamens. Unter der Spalte „Produktbeschreibung“ bzw. dem Namen, den man im Formularfeld vorher vergeben hat (Markierung 3, links vom Pull-down „XPath“, findet man nun sauber alle Texte der Produktbeschreibungen und über die Sortierfunktion alle Textkopien – aber eben nur auf Produktbeschreibungen bezogen.

Natürlich lassen sich auf diese Art auch alle anderen Textelemente einer Domain extrahieren. Einfach den jeweiligen XPath kopieren, in Screaming Frog übertragen und diesen entsprechend crawlen lassen. Zusätzliche Filtermöglichkeiten beim Crawl z. B. nach Verzeichnissen helfen ggf., noch fokussierter zu arbeiten.

Eine andere wichtige Frage ist die, welche der vorhandenen Seiten auch tatsächlich in Google indexiert sind. Hier kann das weniger bekannte Tool „URL-Profiler“ helfen. Man wählt dort die Funktion „Google Indexation“ und übergibt im rechten Fenster eine URL-Liste an, die zu prüfen ist. Im einfachsten Fall hat man eine solche Liste bereits vorliegen oder erzeugt sie z. B. mit dem Screaming Frog. URL-Profiler warnt, dass bei größeren Listen eine Abfragesperre von Google ausgelöst werden kann. Hier empfiehlt es sich, sog. Proxy-Server einzusetzen. Eine Empfehlung dazu findet man unter Settings im Programm selbst. Ansonsten macht das Tool entsprechende Pausen von 30 Sekunden vor jeder Abfrage. Im am Ende erzeugten Ergebnisfile findet man dann u. a. den Indexierungsstatus und das letzte Datum der Erfassung

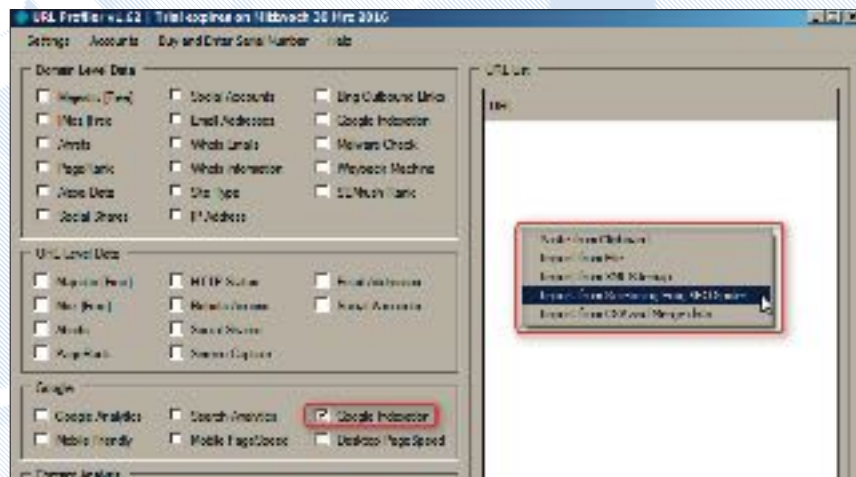


Abb. 10: Der URL-Profiler erlaubt URL-bezogen umfangreiche Datenanreicherungen

URL	Google Indexed	Google Index	Google Cache Date
http://www.	Yes	Base	Mar 12, 2016 16:19:1
http://www.	Yes	Base	Not Cached
http://www.	Yes	Base	Dec 30, 2015 14:53:4
http://www.	Yes	Base	Feb 27, 2016 04:51:1
http://www.	Yes	Base	Feb 19, 2016 10:45:3
http://www.	Yes	Base	Mar 6, 2016 11:37:05
http://www.	Yes	Base	Jan 29, 2016 10:25:4
http://www.	Yes	Base	Mar 9, 2016 18:19:00
http://www.	Yes	Base	Feb 27, 2016 00:10:2
http://www.	No	None	Not Cached
http://www.	Yes	Base	Not Cached
http://www.	No	None	Not Cached
http://www.	Yes	Base	Feb 24, 2016 01:12:4

Abb. 11: URL-Profiler: Welche URL ist auch im Google-Index vorhanden und wann wurde sie zuletzt aktualisiert?

jeder URL, die man zur Prüfung vorgehen hat.

Der URL-Profiler (urlprofiler.com) kostet in der Startversion knapp 16 Euro pro Monat (5.000 URLs pro Import) und kann bis zur Server-Lizenz mit bis zu 250.000 URLs pro Abfrage (ca. 22 Euro) erweitert werden. Eine zweiwöchige Testversion ist verfügbar. Screaming Frog (www.screamingfrog.co.uk) kostet ca. 130 € pro Jahr und kann für kleine Domains (max. 500 URLs) und mit funktionellen Abstrichen kostenlos genutzt werden.

Interne Verlinkung: Betweenness-Zentralität

SEO-Buch-Autor Sebastian Erlhofer begeisterte die Zuhörer mit dem Einsatz eines für SEO-Zwecke eher ungewöhnlichen Tools: Gephi (gephi.org). Dieses Tool ist kostenlos von jedermann nutzbar und visualisiert Zusammenhänge in Daten auf mathematischer Basis. Doch



Abb. 12: Nicht selten weicht der Google-Index durch Programmierfehler auf einer Site deutlich von der Realität ab (Quelle: Mario Träger)

von Anfang an. Die üblichen Verdächtigen bei der Linkanalyse aus SEO-Sicht sind meist Ankertexte, die Linkplatzierung innerhalb einer Seite, die Anzahl Links und/oder die Linkpower, auch Linkjuice oder PageRank genannt. Über Links werden Beziehungen zwischen Webseiten hergestellt. Man spricht in der Graphentheorie hier von Knoten (Dokumente) und Kanten (Links). Liegen alle (Link-)Verbindungen von Dokumen-



Abb. 13: Sebastian Erlhofer erklärte interessante Zusammenhänge über interne Verlinkung via Graphentheorie

ten untereinander vor, lässt sich die Netzwerktopologie berechnen. Daraus ableitbar wird dann die Wichtigkeit oder die Zentralität von Knoten. Hier gibt es verschiedene Betrachtungsmöglichkeiten. Über die Anzahl der eingehenden Kanten auf ein Dokument wird die sog. „Indegree-Zentralität“ berechenbar. Entsprechend beschreibt die „Outdegree-Zentralität“ die Anzahl der ausgehenden Kanten. Weiterhin gibt es noch die sog. Eigenvektor-Zentralität, der auch prinzipiell der PageRank folgt. Der Wert eines Knotens (Dokument) hängt von den Werten der verbundenen umgebenden Knoten ab. Mit anderen Worten werden „beliebte“ Knoten herausgestellt.

Erlhofer wies dann auf die Besonderheit der sog. „Betweenness-Zentralität“ hin. Diese beschreibt die Wichtigkeit eines Knotens als Verteiler und Durchlauf für das gesamte betrachtete Netzwerk. Hierbei stellt man sich vereinfacht die Frage, welcher Schaden im Netzwerkgefüge entstünde, wenn man diesen Knoten herausnehmen würde. Noch schärfer formuliert: Bei welchem Knoten richtet die Herausnahme den größten Schaden an? Um solche Netzwerkanalysen fahren zu können, setzt man z. B. wie oben erwähnt Gephi ein. Dort liest man über eine vorbereitete Tabelle alle über Links bzw. Kanten verbundene Knoten bzw. Dokumente ein, fallweise angereichert mit zusätzlichen

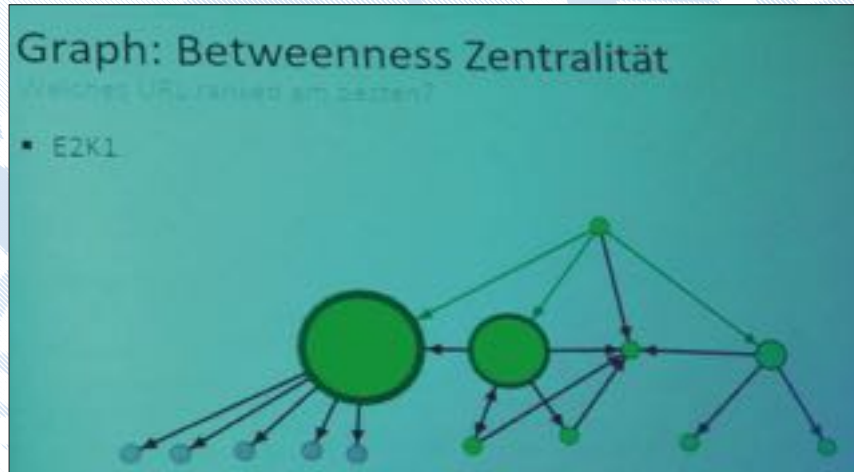


Abb. 14: Welche URL rankt am besten? Ein Experiment sollte Aufschluss bringen (Quelle: Sebastian Erlhofer)

Daten, die später in der Auswertung berücksichtigt werden können. Über eine Force-Atlas-Visualisierung und Gewichtung löst sich dann ein zunächst kompakter Quader langsam in ein sichtbares Netz auf, in dem man je nach Fragestellung analytisch verschiedenen topologischen Aspekten dieses Netzwerks nachgehen kann. Nicht-Mathematiker seien aber vor vorschnellem Ausprobieren dieses Tools gewarnt. Es ist weder ein typisches SEO-Tool noch kommt man mit normaler Schulmathe besonders weit. Anleitungen für gewillte Einsteiger gebe es aber bei YouTube genügend, gab Erlhofer als Tipp mit auf den Weg.

Die eigentlich spannenden Learnings folgten aber erst anschließend. Er zeigte ein eigenes und sauber aufgesetztes SEO-Experiment, in dem er versuchte, das erzeugte Ranking einzelner Seiten einer Domain mit den verschiedenen Zentralitätsberechnungen zu vergleichen. Dabei zeigte sich anhand des Testgraphen, dass das Ranking weder mit der Indegree-, der Outdegree- noch mit der PageRank-Zentralität in Übereinstimmung zu bringen war. Der „Gewinner“ war tatsächlich die Betweenness-Zentralität. Es rankte also diejenige Webseite, die beim Herausnehmen aus der Domain den größten „Schaden“ angerichtet hätte. Einschränkend fügte Erlhofer hinzu, dass noch nicht alle Seiten der Domain

gecrawlt seien und man noch einige Zeit abwarten müsse für eine abschließende Bewertung. Die Methode eigne sich seiner Meinung nach bestens zur visuellen Erkennung von Zusammenhängen, auch Wettbewerber betreffend. Es lassen sich starke Knoten (letztlich: URLs) identifizieren und prüfen, ob diese durch Verlinkung hergestellte Wichtigkeit auch tatsächlich so gewollt und gewünscht sei. Bei großen Netzwerken stoße man allerdings schnell an Berechnungskapazitätsgrenzen. Aufgrund des erforderlichen mathematischen Hintergrund-Know-hows seien solche Analysen leider auch nicht mal eben schnell durchführbar.

Save the date: Die SEO Campixx 2017 selbst wird voraussichtlich am 11. und 12. März stattfinden. Die inhaltliche und organisatorische Zusammensetzung der Campixx:Week davor steht noch nicht fest, wird aber vom Veranstalter Marco Janck rechtzeitig bekannt gegeben. Er verriet nur soviel, als dass er wohl am bisherigen Konzept der Campixx:Week einiges verändern wird. Wer also einzelne Thementage besuchenswert fand, aber wem die Streuung über eine ganze Woche zeitlich zu breit war, der wartet sicher gespannt, was Janck hier optimieren wird.