



Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

DEPARTMENT INFORMATION

Bachelorarbeit

Wenn Google doppelt sieht – Identifikation einer Toleranzgrenze für internen und externen Duplicate Content

vorgelegt von

Johanna Hartje

Studiengang Medien und Information

erster Prüfer: Prof. Dr. Dirk Lewandowski

zweite Prüferin: Prof. Dr. Ulrike Spree

Hamburg, August 2013

Abstract

Wenn Google doppelt sieht – Identifikation einer Toleranzgrenze für internen und externen Duplicate Content

von Johanna Hartje

Guter Content ist heute ein sehr wichtiger Faktor für eine Topposition in einer Suchmaschine wie Google. Doch was geschieht, wenn Content in großem Stil kopiert und auf anderen Seiten wiederverwendet wird, wenn also Duplicate Content entsteht? Genau das soll in der vorliegenden Arbeit untersucht werden. Ziel ist es, eine Toleranzgrenze von internem und externem Duplicate Content bei der Suchmaschine Google zu identifizieren. Dazu wurde ein umfangreicher Website-Test durchgeführt. Es wurden 30 Domains erstellt und mit je drei individuellen Texten bestückt. Diese sind auf drei verschiedene fiktive Keywords optimiert. Dadurch konnte in einem isolierten Raum ohne fremde unkontrollierbare Einflüsse gearbeitet und die gewünschten Faktoren gezielt manipuliert werden. Als die Rankings stabil waren, wurden bestimmte Stufen an internem und externem Duplicate Content eingebaut und zwar 10 %, 30 %, 50 %, 70 % und 90 %. Danach wurde die Rankingentwicklung der Seiten über 21 Tage beobachtet und die Ergebnisse entsprechend ausgewertet. Nach der Auswertung konnte bestätigt werden, dass Duplicate Content in Bezug auf das Ranking der Suchmaschine Google durchaus eine Rolle spielt. Außerdem wurde gezeigt, dass sich interner Duplicate Content stärker auf das Ranking auswirkt als externer, und dass die Toleranzgrenze für internen und externen Duplicate Content vermutlich bei mehr als 50 % liegt. Diese Ergebnisse könnten besonders für SEO-Agenturen interessant sein. Darum wird am Ende der Arbeit ein idealer Test entwickelt und ein Framework zur Wiederholung des Testes präsentiert. So können noch weitere Aspekte untersucht werden, die im Rahmen der Bachelorarbeit aufgrund der kurzen Bearbeitungszeit nicht betrachtet werden konnten, wie zum Beispiel der Einfluss von Backlinks oder von anderen Content-Formen wie Bildern und Videos. Außerdem können die Ergebnisse nach einem neuen großen Update von Google erneut auf seine Gültigkeit überprüft werden. Mit den Ergebnissen des Testes soll Licht ins Dunkel gebracht werden und die ewig andauernde Diskussion um die Behandlung von Duplicate Content bei Google beendet werden.

Inhaltsverzeichnis

Abbildungsverzeichnis.....	4
Tabellenverzeichnis	7
Verzeichnis der Anhänge	8
1 Einleitung	9
2 Die Bedeutung von Content in den Suchmaschinen und andere Rankingfaktoren	11
2.1 Wie Suchmaschinen arbeiten und welche Rankingfaktoren eine Rolle spielen.....	11
2.2 Historische Entwicklung von Content als Rankingfaktor mit Bezug auf die vergangenen Updates.....	14
2.3 Die Contentguidelines von Google	18
3 Duplicate Content in den Suchmaschinen	20
3.1 Die verschiedenen Arten von Duplicate Content mit ihren Eigenschaften.....	20
3.2 Das Problem von Duplicate Content und die Sichtweise von Google	23
3.3 Typische Beispiele zur Entstehung von Duplicate Content und Möglichkeiten zur Vermeidung.....	24
4 Der Website-Test	34
4.1 Die Grundidee und die Hypothesen	34
4.2 Bisherige Tests und aktuelle Situation.....	35
4.3 Die gewählte Methode	41
4.4 Aufbau und Rahmenbedingungen.....	43
4.5 Durchführung des Testes.....	53
4.5.1 Die Domains vom Keyword „Elbwelpe“	53
4.5.2 Die Domains vom Keyword „Ranballsport“.....	59
4.5.3 Die Domains vom Keyword „Jahreszeitentechnologie“	63
4.6 Beobachtung und Auswertung der Rankingentwicklung.....	67
4.7 Zusammenfassung der Ergebnisse	90
5 Tipps und Handlungsempfehlungen auf Basis der Testergebnisse	91
6 Framework für Agenturen mit Hinweisen für den „idealen Test“	93
7 Fazit	106
Literatur.....	108
Anhänge	117

Abbildungsverzeichnis

Abb. 01: Aufbau einer Suchmaschine.....	12
Abb. 02: Faktoren einer Website, die sich negativ auf deren Sichtbarkeit in den Suchmaschinen auswirken.....	16
Abb. 03: Vergleich der Korrelationen der untersuchten Content-Faktoren von Searchmetrics 2012 und 2013.....	18
Abb. 04: Interner Duplicate Content – über mehrere URLs der gleiche Domain erreichbar.....	21
Abb. 05: Externer Duplicate Content – gleiche Content über verschiedene Domains erreichbar.....	21
Abb. 06: Near Duplicate Content – gleicher Content mit nur kleinen Veränderungen.....	22
Abb. 07: Ergebnisseite vom Dupli Checker.....	32
Abb. 08: Angabe der Ähnlichkeit zwischen zwei URLs beim Similar Page Checker.....	33
Abb. 09: Der Einbau von internem Duplicate Content.....	44
Abb. 10: Verteilungsplan bei der Optimierung auf drei Keywords.....	47
Abb. 11: Struktur für die Texte.....	51
Abb. 12: Screenshot der Seite barkasse-im-hafen.de/elbwelpe-auf-grosser-fahrt/ als Beispiel für die Umsetzung der Kriterien, optimiert wurde auf das Keyword „Elbwelpe“.....	52
Abb. 13: Footer der Seite barkasse-im-hafen.de/elbwelpe-auf-grosser-fahrt/ als Beispiel für interne Verlinkungen auf alle anderen URLs der Domain.....	53
Abb. 14: Entwicklung der Rankings beim Keyword „Elbwelpe“ von der Onlinestellung bis zum Einbau von Duplicate Content am 04.07.2013.....	56
Abb. 15: Screenshot der Seite http://mein-mannschaftssport.de/author/admin	60
Abb. 16: Entwicklung der Rankings beim Keyword „Ranballsport“ von der Onlinestellung bis zum Einbau von Duplicate Content am 10.07.2013.....	61

Abb. 17: Entwicklung der Rankings beim Keyword „Jahreszeitentechnologie“ von der Onlinestellung bis zum Einbau von Duplicate Content am 28.07.2013.....	65
Abb. 18: Entwicklung der Rankings im Zeitraum von 21 Tagen nach dem Einbau des Duplicate Contents beim Keyword „Elbwelpe“.....	68
Abb. 19: Anzahlen der URLs, die jeweils im Vergleich zum Vortag ihre Position verändert haben.....	69
Abb. 20: Durchschnittspositionen der Typ-B-URLs.....	71
Abb. 21: Rankingentwicklung der URLs mit 10 % internem Duplicate Content...	72
Abb. 22: Rankingentwicklung der URLs mit 30 % internem Duplicate Content...	72
Abb. 23: Rankingentwicklung der URLs mit 50 % internem Duplicate Content...	73
Abb. 24: Rankingentwicklung der URLs mit 70 % internem Duplicate Content...	74
Abb. 25: Rankingentwicklung der URLs mit 90 % internem Duplicate Content...	74
Abb. 26: Durchschnittspositionen der beklauten Typ-B-URLs.....	75
Abb. 27: Rankingentwicklung der URLs, von denen 10 % Content intern geklaut wurde.....	76
Abb. 28: Rankingentwicklung der URLs, von denen 30 % Content intern geklaut wurde.....	76
Abb. 29: Rankingentwicklung der URLs, von denen 50 % Content intern geklaut wurde.....	77
Abb. 30: Rankingentwicklung der URLs, von denen 70 % Content intern geklaut wurde.....	78
Abb. 31: Rankingentwicklung der URLs, von denen 90 % Content intern geklaut wurde.....	78
Abb. 32: Durchschnittspositionen der Typ-C-URLs.....	79
Abb. 33: Rankingentwicklung der URLs mit 10 % externem Duplicate Content...	80
Abb. 34: Rankingentwicklung der URLs mit 30 % externem Duplicate Content...	80
Abb. 35: Rankingentwicklung der URLs mit 50 % externem Duplicate Content...	81
Abb. 36: Rankingentwicklung der URLs mit 70 % externem Duplicate Content...	81
Abb. 37: Rankingentwicklung der URLs mit 90 % externem Duplicate Content...	82
Abb. 38: Durchschnittspositionen der Typ-D-URLs.....	83

Abb. 39: Rankingentwicklung der URLs, von denen 10 % Content extern geklaut wurde.....	84
Abb. 40: Rankingentwicklung der URLs, von denen 30 % Content extern geklaut wurde.....	85
Abb. 41: Rankingentwicklung der URLs, von denen 50 % Content extern geklaut wurde.....	85
Abb. 42: Rankingentwicklung der URLs, von denen 70 % Content extern geklaut wurde.....	85
Abb. 43: Rankingentwicklung der URLs, von denen 90 % Content extern geklaut wurde.....	86
Abb. 44: Durchschnittspositionen der URLs, bei denen 100 % externer Duplicate Content eingebaut wurde (C-100) und von denen 100 % Content geklaut wurde (D5-100).....	87
Abb. 45: Rankingentwicklung der URLs, bei denen 100 % externer Duplicate Content eingebaut wurde.....	87
Abb. 46: Rankingentwicklung der URLs, von denen 100 % Content extern geklaut wurde.....	88
Abb. 47: Einfaches Phasenmodell zur Durchführung eines Testes.....	99
Abb. 48: Prozessplan zur Wiederholung des Testes.....	101

Tabellenverzeichnis

Tab. 01: Domains, die für den Test registriert wurden.....	48
Tab. 02: Indexierungsstatus der Domains beim Keyword „Elbwelpe“ im zeitlichen Verlauf.....	55
Tab. 03: Übersicht der URLs mit verschiedenen Anteilen an internem und externem Duplicate Content beim Keyword „Elbwelpe“.....	57
Tab. 04: Übersicht der Domains mit ihren Typen zum Keyword „Elbwelpe“.....	58
Tab. 05: Indexierungsstatus der Domains beim Keyword „Randballsport“ im zeitlichen Verlauf.....	59
Tab. 06: Übersicht der URLs mit verschiedenen Anteilen an internem und externem Duplicate Content beim Keyword „Randballsport“.....	62
Tab. 07: Übersicht der Domains mit ihren Typen zum Keyword „Randballsport“.....	63
Tab. 08: Indexierungsstatus der Domains beim Keyword „Jahreszeitentechnologie“ im zeitlichen Verlauf.....	63 f.
Tab. 09: Übersicht der URLs mit verschiedenen Anteilen an internem und externem Duplicate Content beim Keyword „Jahreszeitentechnologie“.....	66
Tab. 10: Übersicht der Domains mit ihren Typen zum Keyword „Jahreszeitentechnologie“.....	67
Tab. 11: Bestätigte und widerlegte Hypothesen der Bachelorarbeit.....	90
Tab. 12: Benötigte Domains bei der Optimierung auf drei Keywords.....	100

Verzeichnis der Anhänge

Anh. 01: Anfrageabhängige Faktoren im Ranking einer Web-Suchmaschine.....	117
Anh. 02: Anfrageunabhängige Faktoren im Ranking einer Web-Suchmaschine....	118
Anh. 03: Searchmetrics Rankingfaktoren2013.....	119
Anh. 04: Korrelation der Content-Faktoren 2013.....	120
Anh. 05: Entwicklung des Rankings im Zeitraum von 21 Tagen nach dem Einbau des Duplicate Contents beim Keyword „Randballsport“.....	121
Anh. 06: Entwicklung des Rankings im Zeitraum von 21 Tagen nach dem Einbau des Duplicate Contents beim Keyword „Jahreszeitentech- nologie“.....	122

1 Einleitung

Die Bedeutung von Duplicate Content für das Ranking in der Suchmaschine Google ist in der SEO-Szene ein sehr umstrittenes Thema. Einige sind der Meinung, dass es eine Duplicate Content Penalty gibt, also, dass eine Seite gezielt wegen Duplicate Content im Ranking zurückgesetzt werden kann, andere sind vom Gegenteil überzeugt.

Erst im April dieses Jahres, genau am 22.04.2013, stellte der deutsche Bilder-SEO-Spezialist Martin Mißfeldt in seinem Artikel „Duplicate Content Penalty nur ein Seo-Gespens?!“ die Behauptung auf, dass es keine Duplicate Content Penalty gibt (vgl. MISSFELDT 2013). Er bezieht sich hier auf verschiedene Aussagen von Matt Cutts, dem Head of Webspam bei Google, und dem Google Webmaster Beauftragten John Müller. Schon in den Kommentaren wird deutlich, dass einige SEOs (Suchmaschinenoptimierer) hier durchaus anderer Meinung sind und sehr wohl schon Erfahrungen mit einer Duplicate Content Penalty gemacht haben. Auch Matt Cutts selbst äußerte sich im Juli 2013 noch einmal zu diesem Thema und versucht für Klarheit zu sorgen (vgl. CUTTS 2013). Doch auch hier sind einige SEOs skeptisch und glauben, dass Google nicht immer so handelt, wie sie selbst sagen. Die Diskussion um die Bedeutung von Duplicate Content ist also auch 2013 immer noch in vollem Gange und wird immer wieder neu angefacht.

Das Ziel dieser Bachelorarbeit ist es nun, Licht ins Dunkel zu bringen und vor allem folgende Fragen zu klären: Wie genau reagiert Google genau auf Duplicate Content? Reagiert Google überhaupt? Gibt es eine Toleranzgrenze bis wohin Duplicate Content akzeptiert wird? Was passiert, wenn sich zu viel Duplicate Content auf einer Seite befindet und die Toleranzgrenze überschritten wird? Welche Auswirkungen haben die verschiedenen Arten von Duplicate Content, also interner und externer Duplicate Content?

Um all diese Fragen beantworten zu können, wurde ein umfangreicher Website-Test durchgeführt. Hierzu wurden insgesamt 30 Domains aufgesetzt und auf drei verschiedene Keywords optimiert. Zunächst enthielt jede Domain drei Seiten mit individuellen Texten. Diese wurden, wenn stabile Rankings erreicht waren, durch verschiedene Anteile an Duplicate Content ersetzt und zwar 10 %, 30 %, 50 %, 70% und 90 %. So sollte es möglich sein, eine Toleranzgrenze für Duplicate Content zu identifizieren. Die Entwick-

lung der Rankings wurde über 21 Tage beobachtet und anschließend gründlich ausgewertet. Die ausführliche Beschreibung und Auswertung des Website-Testes findet sich in Kapitel 4 dieser Arbeit.

Zunächst wird in Kapitel 2 aber erst einmal ein Einblick in die Funktionsweise von Suchmaschinen gegeben und aufgezeigt, welche Rolle Content überhaupt für das Ranking spielt. Dabei wird Bezug auf die vergangenen großen Updates von Google genommen und so gezeigt, dass guter Content für das Ranking immer wichtiger wird.

Danach werden in Kapitel 3 die verschiedenen Arten von Duplicate Content vorgestellt und erklärt, warum das Ganze überhaupt ein Problem für die Suchmaschine darstellt. Schließlich werden verschiedene Beispiele zur Entstehung von Duplicate Content mit entsprechenden Lösungsmöglichkeiten aufgezeigt und einige Tools vorgestellt, die bei der Identifikation von Duplicate Content-Problemen helfen können.

Nach dem der Website-Test, also das Herzstück dieser Arbeit, ausführlich behandelt wurde, werden in Kapitel 5 Tipps und Handlungsempfehlungen für Webseitenbetreiber auf Basis der Testergebnisse gegeben.

Kapitel 6 richtet sich abschließend speziell an SEO-Agenturen. Hier wird ein idealer Test vorgestellt, mit dem weitere Faktoren untersucht werden können, die im Website-Test im Rahmen der Bachelorarbeit aufgrund der kurzen Bearbeitungszeit nicht behandelt werden konnten. Außerdem wird ein Framework präsentiert, mit dem der Website-Test dieser Arbeit leicht wiederholt werden kann. Dies bietet sich zum Beispiel nach einem neuen großen Update von Google an.

2 Die Bedeutung von Content in den Suchmaschinen und andere Rankingfaktoren

Suchmaschinen sind aus dem Internet heute nicht mehr wegzudenken. Laut der ARD/ZDF-Onlinestudie 2012 nutzen 83 % der Befragten Suchmaschinen mindestens einmal wöchentlich. Damit liegen sie noch über dem Senden und Empfangen von E-Mails, was 79 % der Befragten mindestens einmal wöchentlich machen (vgl. FREES / VAN EIMEREN 2012, S. 369). Aber nicht nur die Internetuser, sondern auch die Webseitenbetreiber wissen den Nutzen der Suchmaschinen zu schätzen. So kamen laut einer Auswertung von shareaholic im Oktober 2012 bei 200.000 untersuchten Seiten knapp 38 % der Visits von der organischen Googlesuche (vgl. ARONICA 2012). Welche Suchmaschine dabei von den Deutschen präferiert wird, ist eindeutig: Google stellte im Juni 2013 in Deutschland einen Marktanteil von 90,7 % dar (vgl. SEO-UNITED I 2013). Die anderen Suchmaschinen werden kaum genutzt und nur am Rande wahrgenommen. Daher wurde auch Google ausgewählt, um die Auswirkungen von Duplicate Content auf das Ranking zu untersuchen. Doch wie kommt das Ranking in der Suchmaschine überhaupt zustande? Angeblich sollen 200 Faktoren beachtet werden (vgl. DEAN 2013). Welche sind das? Wie haben sie sich in den vergangenen Jahren verändert und welche Rolle spielt der Content heute? Diese Fragen sollen im folgenden Kapitel beantwortet werden.

2.1 Wie Suchmaschinen arbeiten und welche Rankingfaktoren eine Rolle spielen

„Eine Suchmaschine (auch: Web-Suchmaschine; Universalsuchmaschine) ist ein Computersystem, das Inhalte aus dem *World Wide Web* mittels *Crawling* erfasst und über eine Benutzerschnittstelle durchsuchbar macht, wobei die Ergebnisse in einer nach systemseitig angenommener Relevanz geordneten Darstellung aufgeführt werden“ (LEWANDOWSKI 2013, S. 495).

Bereits in der Definition wird das Crawling als Methode zur Erfassung der Webinhalte genannt. Der Aufbau einer Suchmaschine wird aber noch durch weitere Elemente ergänzt (siehe Abbildung 01). Zu Beginn verfolgt der Crawler Hyperlinks aus bereits bekannten Dokumenten und versucht so, möglichst viele neue Dokumente zu finden. Danach zerlegt das Parsing Module die gefundenen Dokumente in indexierbare Einheiten, die dann vom Indexer indexiert werden (vgl. LEWANDOWSKI 2005, S. 27). Es wird ein invertierter Index erstellt, „der für jede indexierte Einheit diejenigen Dokumente verzeichnet, in denen sie vorkommt“ (LEWANDOWSKI 2013, S. 500). Wenn ein Benutzer eine Suchanfrage stellt, wird der Index abgefragt. Dies geht dann schneller, als alle Dokumente einzeln zu durchsuchen. Die eingegebene Suchanfrage wird dabei vom Searcher in eine weiterverarbeitbare Form umgesetzt, in der beispielsweise Operatoren entsprechend aufgelöst werden (vgl. LEWANDOWSKI 2013, S. 501).

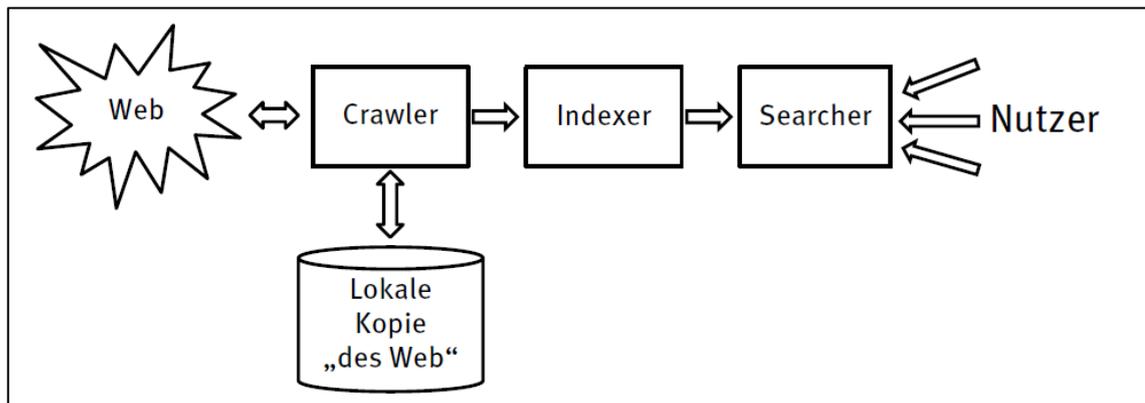


Abbildung 01: Aufbau einer Suchmaschine (Quelle: LEWANDOWSKI 2013, S. 498)

Die Ergebnisse, die der Nutzer erhält, werden dann in einem bestimmten von den Suchmaschinen festgelegten Ranking ausgegeben. Dies ist auch nötig, da es viel zu viele zu den meist einfachen Suchanfragen der Nutzer passende Dokumente im Web gibt. Diese unsortiert auszuspielen, wäre für die Nutzer nicht befriedigend und es wäre sehr aufwändig, alle Dokumente selbst nach dem passenden zu durchsuchen. Also übernimmt die Suchmaschine diese Aufgabe und sortiert die Dokumente nach einer angenommenen Relevanz (vgl. DOPICHAJ 2009, S. 101). Dabei spielen verschiedene Rankingfaktoren eine Rolle. Diese lassen sich grob in vier Bereiche unterteilen:

- textspezifische Faktoren: Hier wird abgeglichen, „welche Wörter der Suchanfrage in den zu durchsuchenden Dokumenten vorkommen und daher in der Treffermenge mit aufgenommen werden sollen“ (LEWANDOWSKI 2013, S. 501).

Es können außerdem Begriffe, die etwa in Überschriften oder zu Beginn des Dokuments genannt werden, höher gewichtet werden.

- Messung von Popularität: Die Popularität eines Dokuments wird entweder durch deren Verlinkungen oder auch durch das Klickverhalten der User bestimmt. Dies ist der zweite bestimmende Bereich der Rankingfaktoren (vgl. LEWANDOWSKI 2013, S. 501).
- Aktualität: Je nach Suchanfrage kann es sinnvoll sein, besonders aktuelle Dokumente weit oben zu ranken. „Da linktopologische Algorithmen tendenziell ältere Dokumente bevorzugen, wird Aktualität auch als Ausgleichsfaktor herangezogen“ (LEWANDOWSKI 2013, S. 501).
- Lokalität: Hierbei wird der Standort des Nutzers berücksichtigt und beispielweise Dokumente aus Deutschland bevorzugt, wenn der Nutzer von dort kommt (vgl. LEWANDOWSKI 2013, S. 501 f.).

Darüber hinaus lassen sich die Rankingfaktoren in anfrageabhängige und anfrageunabhängige Faktoren unterteilen. Die anfrageabhängigen Faktoren beinhalten dabei zum Beispiel Worthäufigkeiten oder die Position der Suchbegriffe im Dokument. Sie orientieren sich an den im klassischen Information Retrieval verwendeten Verfahren. Dem gegenüber wird mit den anfrageunabhängigen Faktoren die Qualität bzw. Autorität eines Dokuments unabhängig von der Suchanfrage bestimmt (vgl. LEWANDOWSKI 2005, S. 90). Bei klassischen Information-Retrieval-Systemen werden in der Regel nur die anfrageabhängigen Faktoren für das Ranking verwendet. Dies reicht bei Web-Dokumenten jedoch nicht aus. Wegen der großen qualitativen Unterschiede der Dokumente im Web werden die anfrageunabhängigen Faktoren als weiteres Rankingkriterium hinzugezogen (vgl. LEWANDOWSKI 2005, S. 93). Alle Suchmaschinen arbeiten also mit einer Kombination der anfrageabhängigen und anfrageunabhängigen Faktoren. Jedoch werden sie je Suchmaschine vermutlich anders gewichtet, wodurch Unterschiede in den Suchergebnissen zustande kommen können. Eine Übersicht über die anfrageabhängigen und anfrageunabhängigen Rankingfaktoren findet sich im Anhang 01 und 02 dieser Arbeit.

Auch viele SEOs beschäftigen sich immer wieder mit der Frage, welche Faktoren für das Ranking eine Rolle spielen. Dazu wurde beispielsweise von Searchmetrics 2013 eine Korrelationsanalyse durchgeführt. Hierbei wurde geprüft, welche Eigenschaften hoch rankende Seiten im Gegensatz zu niedrig rankenden Seiten aufweisen. Wichtige

Ergebnisse in der diesjährigen Analyse waren, dass Keyword-Links und –Domains an Relevanz verloren haben, für Domains von starken Marken anscheinend Ausnahmen gelten, Social Signals weiterhin mit sehr guten Rankings korrelieren, guter Content immer wichtiger wird, die Anzahl von Backlinks weiterhin sehr wichtig bleibt und gewisse On-Page-Techniken Grundvoraussetzung für ein gutes Ranking sind (vgl. SEARCHMETRICS 2013). Die Ergebnisse der Analyse in der Darstellung einer Spearman-Korrelation finden sich im Anhang 03 dieser Arbeit.

Ähnliche Analysen wurden 2013 auch von MOZ und Netmark veröffentlicht. Hierbei kamen teilweise andere Ergebnisse zustande, weil unterschiedliche Methoden verwendet wurden. Im Großen und Ganzen sind sich aber alle einig, dass Backlinks weiterhin einen der größten Einflüsse auf das Ranking haben und auch Social Signals mit hohen Positionen im Ranking korrelieren (vgl. RIXECKER 2013 und NETMARK 2013).

2.2 Historische Entwicklung von Content als Rankingfaktor mit Bezug auf die vergangenen Updates

Google arbeitet ständig an der Weiterentwicklung des Algorithmus und der Verbesserung des Rankings. 500 bis 600 Mal im Jahr wird dazu der Algorithmus angepasst. Doch diese Anpassungen sind meist so klein, dass sie gar nicht bemerkt werden. Nur wenn die Veränderungen die Suchergebnisse signifikant beeinflussen, spricht man von einem Update. Dies kommt mehrmals im Jahr vor (vgl. SEEFELD 2012).

Google ging im März 1998 als erste Suchmaschine, die Links zur Berechnung des Rankings einbezogen hat, online. Seither gab es bis Mitte Juni 2013 insgesamt 102 registrierte Updates mit unterschiedlich großen Auswirkungen (vgl. MOZ 2013). Im September 2002 wurde das erste Update dokumentiert, bei dem die Veränderungen allerdings unklar blieben (vgl. SEEFELD 2012). 2003 folgten mit „Boston“ im Februar und „Cassandra“ im April zunächst Updates mit Bezug auf den Index oder die Bewertung von Links. Das erste Content-bezogene Update kam mit „Florida“ im November 2003. Hierbei verloren viele Seiten mit Keyword-Stuffing an Positionen im Ranking (vgl. SEEFELD 2012). Google ging so gegen gängige SEO-Methoden aus den 90er Jahren

vor, was mit den Updates „Austin“ (Januar 2004) und „Brandy“ (Februar 2004) fortgeführt wurde. Mit „Bourbon“ wurde im Mai 2005 das erste Update mit Bezug auf Duplicate Content ausgerollt. Es wurde von Google als „something like 3.5 improvements in search quality“ (BATELLE 2005) angekündigt. Zunächst war nicht klar, was damit gemeint war. Doch schnell stellte sich heraus, dass die Art und Weise, wie Duplicate Content und non-canonical URLs (also mit und ohne www. erreichbar) behandelt werden, geändert wurde (vgl. MOZ 2013). In den nächsten Jahren wurde es stiller um den Content und es wurden überwiegend Updates zu allgemeinen Algorithmusänderungen oder zur Darstellung in den SERPs (Suchergebnisseiten) ausgespielt.

Dies änderte sich am 23.02.2011 als mit dem Panda-Update eines der größten Google-Updates überhaupt ausgerollt wurde. „Ziel dieses Updates ist es, die Rankings für Websites von geringer Qualität zu reduzieren, d. h. für Websites, die einen geringen Mehrwert für Nutzer darstellen, die Inhalte von anderen Websites kopieren oder die einfach nicht hilfreich sind“ (PACHER 2011). Auf der anderen Seite sollten qualitativ hochwertige Seiten höher gerankt werden. So soll die Qualität in den Suchergebnissen verbessert werden. Im Februar 2011 wurde das Update zunächst nur in den USA ausgerollt und traf hier etwa 12% aller englischsprachigen Suchanfragen spürbar. Am 11.04.2011 ging es als Panda 2.0 für alle englischsprachigen Suchanfragen weltweit live und wurde am 12.08.2011 schließlich für alle Sprachen außer Chinesisch, Japanisch und Koreanisch übernommen. Von den Suchanfragen für alle übrigen Sprachen wurden circa 6-9% getroffen, was ebenfalls zu einem deutlich spürbaren Effekt führte (vgl. SISTRIX I 2013). Guter und einzigartiger Content spielt nun für das Ranking also eine wesentlich höhere Rolle als zuvor. „Google baut auf die qualitativ hochwertigen Inhalte guter Websites aus aller Welt und fühlt sich für die Erhaltung eines gesunden Web-Ökosystems verantwortlich“ (PACHER 2011). Wer also qualitativ schlechte Inhalte auf seiner Seite veröffentlicht, verliert viele Positionen im Ranking und somit Klicks und damit möglicherweise verbundene Werbeeinahmen. Doch wie bewertet Google die Qualität einer Seite? Laut den Auswertungen von SISTRIX wirken sich folgende Faktoren negativ auf die Sichtbarkeit einer Website in den Suchmaschinen aus (Abbildung 02):

- ▶ eine geringe Menge an originalen Inhalten (unique content)
- ▶ wenige oder keine qualitativen Backlinks
- ▶ geringe Verweildauer (time on page & time on site)
- ▶ ein hoher Prozentsatz an doppelten Inhalten (duplicate content)
- ▶ hohe Absprungrate (bounce rate)
- ▶ ein hoher Prozentsatz an Seiten mit geringen original/einzigartigen Inhalten
- ▶ ein hoher Prozentsatz an sog. boilerplate Inhalten (z.B. gleiche Texte oder Links auf jeder Seite)
- ▶ eine hohe Anzahl an irrelevanten Anzeigen (Werbung), (generell: zu viel Werbung)
- ▶ Seiteninhalt und Seitentitel stimmen nicht mit der Suchanfrage überein
- ▶ unnatürlich häufiges Vorkommen eines Wortes auf einer Seite

Abbildung 02: Faktoren einer Website, die sich negativ auf deren Sichtbarkeit in den Suchmaschinen auswirken (Quelle: SISTRIX I 2013)

Nicht nur die großen Veränderungen in der Bewertung der Qualität des Contents und der große prozentuale Anteil der getroffenen Seiten zeigt die Bedeutung des Panda-Updates. Auch der Zeitraum über den es ausgerollt wurde, ist länger als üblich. Version 1.0 des Pandas ging am 23.02.2011 live. Über zwei Jahre später, am 11.07.2013, gab Matt Cutts (Google's Head of Webspam) bekannt, dass der Panda Algorithmus noch immer jeden Monat durch ein Update angepasst wird. Dieses wird jeweils über einen Zeitraum von etwa zehn Tagen ausgerollt (vgl. SCHWARTZ I 2013). Dadurch zeigt Google, wie viel Wert auf die Qualität des Contents gelegt wird und dass dies auch dauerhaft so bleiben soll.

Etwa ein Jahr und zwei Monate nach dem Panda-Update folgte ein weiteres Update zur Stärkung der Qualität in den SERPs: Penguin wurde am 24.04.2012 für alle Sprachen ausgerollt. Diesmal ging es um die Bekämpfung von Webspam. Darunter versteht Google Techniken, die einzig und allein dem Zweck dienen, eine möglichst hohe Position in den Suchergebnissen zu erzielen und keinerlei Nutzen für den User haben, wie zum Beispiel Keyword-Stuffing oder Cloaking (vgl. SISTRIX II 2013). Google verfolgt mit diesem Update folgendes Ziel: „The goal of many of our ranking changes is to help searchers find sites that provide a great user experience and fulfill their information needs“ (CUTTS 2012). Darum sollen Seiten, die gegen die Google Quality Guidelines verstoßen, abgestraft werden und qualitativ hochwertige Seiten höher ranken (mehr zu den Quality Guidelines in Kapitel 2.3). Obwohl das Penguin-Update nur 3,1% aller englischsprachigen Suchanfragen getroffen hat und somit weit geringere Auswirkungen als das Panda-Update hat, zeigt sich mit diesen beiden Updates die deutliche Ausrichtung Google's weg von unlauteren Tricks hin zu qualitativ hochwertigem Content, der den Usern auch wirklich nützt.

Besonders durch das Panda-Update, aber auch in Kombination mit dem Penguin-Update, wurden die Aufgaben der Webmaster in gewisser Weise neu definiert. Es kommt nun viel mehr darauf an, wirklich guten und einzigartigen Content zu veröffentlichen, als denn die Suchmaschinen mit verschiedenen Tricks hinters Licht führen zu wollen. Google postete dazu etwa sechs Wochen nach Panda 1.0 Folgendes in seinem Webmaster Central Blog: „Search is a complicated and evolving art and science, so rather than focusing on specific algorithmic tweaks, we encourage you to focus on delivering the best possible experience for users“ (SINGHAL 2011).

Die Bedeutung von Content hat in großem Maße zugenommen und Danny Sullivan, Gründer vom Search Engine Land, spricht sogar von Panda als neuem Rankingfaktor und nicht „nur“ von einem Update (vgl. SULLIVAN 2011). Auch in der bereits in Kapitel 2.1 angesprochenen Korrelationsanalyse der Rankingfaktoren von Searchmetrics spielen Content-Faktoren eine weitaus größere Rolle als noch im Jahr zuvor. Laut den Ergebnissen 2013 korrelieren fast alle Content-Faktoren positiv und wurden im Vergleich zum Vorjahr deutlich aufgewertet. Gut rankende Seiten weisen bis zu einem gewissen Umfang mehr Text und eine höhere Anzahl von Medienintegrationen auf. Außerdem scheint eine gute interne Linkstruktur ein weiteres Qualitätsattribut zu sein (vgl. SEARCHMETRICS 2013). Abbildung 03 zeigt den Vergleich der Korrelation von einigen Content-Faktoren aus den Untersuchungen der Jahre 2012 und 2013. Die Anzahl der Worte im Text beispielsweise hatte 2012 mit einem Korrelationswert von -0,06 noch einen negativen Einfluss. Dieser hat sich 2013 umgekehrt und mit einem Wert von 0,08 hat die Anzahl der Worte nun einen deutlichen positiven Einfluss. 2013 wurden außerdem noch einige weitere Content-Faktoren untersucht, die fast ausschließlich positiv korrelieren. Eine Übersicht hierzu findet sich im Anhang 04 dieser Arbeit.

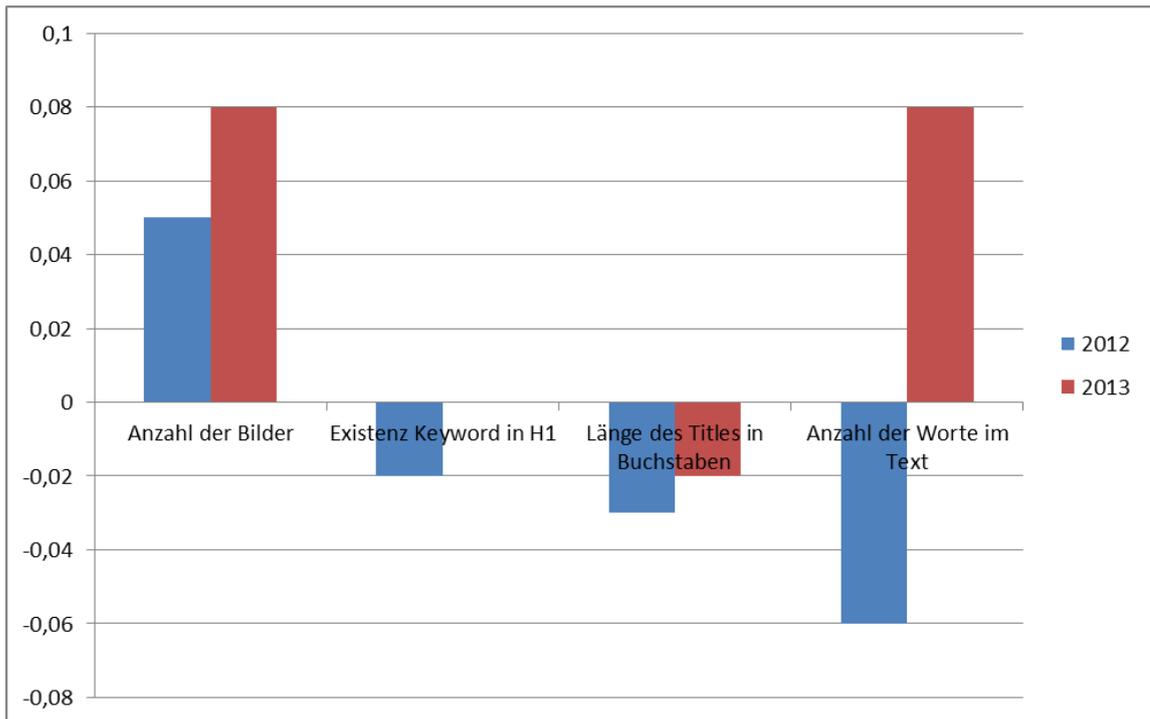


Abbildung 03: Vergleich der Korrelationen der untersuchten Content-Faktoren von Searchmetrics 2012 und 2013 (eigene Darstellung in Anlehnung an SEARCHMETRICS 2012 und SEARCHMETRICS 2013)

Wer also hoch ranken will, kommt an gutem Content heute nicht mehr vorbei. Auch aufgrund dieser Entwicklung ist es sehr interessant im Test herauszufinden, wie Google auf Duplicate Content reagiert und ab welchem Anteil von Duplicate Content die Seiten abgestraft werden.

2.3 Die Contentguidelines von Google

Google setzt nicht nur mit den Updates neue Maßstäbe, sondern auch mit seinen Richtlinien für Webmaster. Diese wurden am 02.10.2012, also im Jahr des Penguin-Updates, aktualisiert. Google möchte mit den Richtlinien transparent kommunizieren, welche Techniken in den Rankings bevorzugt bzw. gewünscht und welche möglicherweise abgestraft werden. Außerdem ging Google bei der Aktualisierung auch auf aktuelle Entwicklungen und negative Trends aus der SEO-Szene ein (vgl. DZIKI 2012).

In den Richtlinien für Webmaster sind Design- und Inhaltsrichtlinien, technische Richtlinien und Qualitätsrichtlinien enthalten. Werden die Richtlinien befolgt, „ist es für Google einfacher, Ihre Website zu finden, in den Index aufzunehmen und zu sortieren“ (GOOGLE I 2013). In den Design- und Inhaltsrichtlinien wird immer wieder darauf hingewiesen, dass die Seite klar strukturiert sein und der Content eindeutig beschrieben werden soll. Bei der Strukturierung sollen interne Verlinkungen helfen, aber es sollen nicht zu viele Links auf einer Seite gesetzt werden. Es ist durchaus gewünscht, dass Suchbegriffe auf der Seite verwendet werden, aber eben nur sinnvolle und nützliche. Es sollen aussagekräftige und eindeutige Title-Tags und Alt-Attribute benutzt werden und es soll vermieden werden, wichtigen Text in Bildern zu speichern, weil diese nicht erfasst werden können. Zur schnelleren Indexierung wird außerdem dazu geraten, eine Sitemap zu erstellen. Zum Schluss wird noch einmal auf die korrekte Verwendung der HTML-Attribute hingewiesen und dazu geraten nur kurze Parameter in dynamischen URLs zu verwenden (vgl. GOOGLE I 2013). Im Allgemeinen wird immer wieder darauf hingewiesen, dass die Website nützlich und informativ sein soll. Es sollen aussagekräftige und eindeutige Informationen zur Verfügung gestellt werden. In den Qualitätsrichtlinien wird darüber hinaus noch einmal deutlich gemacht, dass die Seiten in erster Linie für Nutzer und nicht für Suchmaschinen erstellt werden sollen. Bei der Entwicklung sollen sich die Webmaster stets die Fragen stellen, ob dies für die Nutzer von Vorteil ist und ob man genauso handeln würde, wenn es keine Suchmaschinen gäbe. Die Nutzer sollen nicht getäuscht werden und Webseiten sollen einzigartig und wertvoll gestaltet werden. Es heißt hier: „Überlegen Sie, was Ihre Website einzigartig, wertvoll oder einnehmend macht. Gestalten Sie Ihre Website so, dass sie sich von anderen in Ihrem Bereich abhebt“ (GOOGLE I 2013).

Google versucht also auch durch die Richtlinien für Webmaster die Qualität der Websites und somit der Suchergebnisse zu steigern. Dabei spielt einzigartiger und hochwertiger Content ebenfalls eine sehr große Rolle. Außerdem wird in den Richtlinien auch noch einmal auf verbotene Techniken wie zum Beispiel Cloaking, Brückenseiten oder Linktausch hingewiesen.

3 Duplicate Content in den Suchmaschinen

Interner Duplicate Content wurde 2012 bei einer Umfrage von seo-book.de unter 250 SEO-Spezialisten zum häufigsten SEO-Fehler der Welt gewählt. Zum schädlichsten wurden Fehler mit dem Canonical Tag gewählt, was in direktem Zusammenhang mit dem Duplicate Content steht. Bezogen auf die Schädlichkeit belegt interner Duplicate Content Platz 3 der Umfrage (vgl. KUBITZ I 2012). Die Umfrage zeigt, wie wichtig Duplicate Content in Bezug auf die SEO-Optimierung einer Seite ist und wie viele Fehler hier gemacht werden.

Doch was genau ist Duplicate Content? Wie kann er entstehen und vor allem wie vermieden werden? Diese und weitere Fragen sollen im folgenden Kapitel beantwortet werden. Es werden verschiedene Arten von Duplicate Content ausführlich erläutert und die Sichtweise von Google dargestellt. Außerdem werden typische Beispiele zur Entstehung aufgezeigt und verschiedene Maßnahmen zur Vermeidung und Bekämpfung von Duplicate Content erklärt.

3.1 Die verschiedenen Arten von Duplicate Content mit ihren Eigenschaften

Googles offizielle Definition von Duplicate Content lautet wie folgt:

„Duplizierter Content“ sind umfangreiche Contentblöcke, die anderem Content auf derselben oder einer anderen Domain entsprechen oder auffällig ähneln“ (GOOGLE II 2013).

Es wird also zwischen internem und externem Duplicate Content unterschieden.

Interner Duplicate Content beschränkt sich dabei auf die eigene Domain (vgl. SISTRIX III 2013). Der gleiche Content ist unter mehreren URLs der gleichen Domain erreichbar (siehe Abbildung 04).



Abbildung 04: Interner Duplicate Content – über mehrere URLs der gleichen Domain erreichbar (Quelle: MEYERS I 2011)

Externer Duplicate Content tritt hingegen domainübergreifend auf (vgl. SISTRIX III 2013). Hier ist der gleiche Content also unter verschiedenen Domains erreichbar (siehe Abbildung 05).



Abbildung 05: Externer Duplicate Content – gleicher Content über verschiedene Domains erreichbar (Quelle: MEYERS I 2011)

Neben internem und externem Duplicate Content gibt es weitere Fälle, bei denen Content kopiert oder sogar gestohlen wird. Vier davon werden im Folgenden erläutert.

Von **Near Content** oder **Near Duplicate Content** ist die Rede, wenn sich die Seiten sehr stark ähneln, aber nicht komplett kopiert wurden. Die Seiten können sich etwa durch andere Bilder, einen anderen Absatz oder die Reihenfolge des Contents unterscheiden (vgl. MEYERS I 2011), wie in Abbildung 06 dargestellt. Auch Near Duplicate Content kann zu Problemen führen und von den Suchmaschinen als Duplicate Content identifiziert und so behandelt werden (vgl. ONPAGE.ORG 2013). Daher sollte auch dieser möglichst vermieden oder mit den unter 3.3 beschriebenen Methoden bekämpft werden.



Abbildung 06: Near Duplicate Content – gleicher Content mit nur kleinen Veränderungen
(Quelle: MEYERS I 2011)

Wird Content weitergegeben und an anderer Stelle im Internet genutzt, spricht man von **Content Syndication**. Die Inhalte können dabei kostenpflichtig oder frei mit Erlaubnis des Autors genutzt werden. Ein Beispiel für die kostenlose Nutzung wären die Inhalte von Wikipedia (vgl. SEO-UNITED II 2013). Bei der Weitergabe von Inhalten an fremde Domains kommt es allerdings zu externem Duplicate Content. Es können aber verschiedene Maßnahmen getroffen werden, damit Google die ursprüngliche Quelle des Inhalts trotzdem erkennt. So sollte es beispielsweise schon ausreichen, wenn vom syndicated Content ein Backlink zur Originalquelle gesetzt wird (vgl. ENGE 2010). Grundsätzlich sollte aber nur hochqualitativer Content zur Weitergabe genutzt werden. Je schlechter der Inhalt ist, desto negativer sind auch die Auswirkungen im Ranking (vgl. MARSHALL 2011).

Wenn der Content ohne Erlaubnis von anderen höherwertigen Seiten gestohlen wird, ist dies **Scraped Content**. Diese Methode ist illegal. Die Webmaster erhoffen sich davon selbst eine höhere Platzierung in den Rankings. Die Verwendung von Scraped Content verstößt meistens gegen das Urheberrecht, da die Scrapper gestohlene Texte als ihre eigenen ausgeben. Es ist aber schwer gegen die Scrapper vorzugehen, da sie meist ein weltweit vernetztes System mit sehr vielen verschiedenen Domains besitzen (vgl. JANSSEN 2013). Google gibt an, dass Scraped Content genauso wie Duplicate Content behandelt wird, da er ohne weiteres Hinzufügen von Inhalt keinen Mehrwert für die User darstellt (vgl. GOOGLE III 2013). Schon allein aus moralischen Gründen ist dringend von der Verwendung von Scraped Content abzuraten.

Besteht eine Seite nur aus syndicated oder scraped Content ohne weiteren Mehrwert, entsteht **Thin Content**. Dieser wurde von Google im Panda-Update abgestraft und seit dem stark bekämpft. Thin Content ist allgemein gesagt, „Inhalt von geringer Qualität mit geringem oder gar keinem Mehrwert“ (GOOGLE IV 2013). Beispiele sind etwa

Partnerwebseiten, Brückenseiten oder automatisch generierte und kopierte Inhalte. Diese Seiten bieten keinen Mehrwert für die Nutzer und sollen darum auch nicht in den Suchergebnissen angezeigt werden. Um das Problem zu beheben, sollte möglicher Duplicate Content von den Seiten entfernt und der restliche Inhalte durch individuellen Content mit Mehrwert für die Nutzer angereichert werden (vgl. GOOGLE IV 2013).

3.2 Das Problem von Duplicate Content und die Sichtweise von Google

Google möchte dem Nutzer vielfältige und relevante Ergebnisse präsentieren. Auf der ersten Suchergebnisseite sollen möglichst zehn verschiedene Treffer erscheinen und nicht derselbe Content unter verschiedenen URLs. Daher versucht Google Duplicate Content zu identifizieren und so dem User nur einzigartige Ergebnisse zu liefern (vgl. MOSKWA 2008).

Um diese Vielfalt zu gewährleisten, werden alle URLs mit dem gleichen Inhalt, also mit Duplicate Content, von Google zu einer Gruppe zusammengefasst. Aus dieser wird dann eine URL ausgewählt, die die Gruppe aus Googles Sicht am besten repräsentiert. Diese erscheint dann in den SERPs. Die anderen URLs der Gruppe treten nicht in den Suchergebnissen auf oder werden wesentlich niedriger eingestuft. Als dritter Schritt werden noch bestimmte Eigenschaften der URLs der Gruppe, wie etwa die Link-Popularität, auf die ausgewählte URL übertragen (vgl. MOSKWA 2008).

Google geht zunächst immer davon aus, dass Duplicate Content zufällig und ohne böse Absichten entstanden ist. Dann wird wie oben beschrieben verfahren. Duplicate Content allein führt zu keinen Abstrafungen oder direkten Maßnahmen gegen die Seite. Eine reine Duplicate Content-Penalty gibt es nicht (vgl. MOSKWA 2008). Wenn allerdings herausgefunden wurde, dass mit dem Duplicate Content das Ranking manipuliert oder Nutzer getäuscht werden sollen, werden Korrekturen am Index und Ranking der betroffenen Seite durchgeführt. „Infolgedessen werden diese Websites unter Umständen in den Suchergebnissen niedriger eingestuft oder sogar aus dem Google-Index entfernt und damit nicht mehr in den Suchergebnissen angezeigt“ (GOOGLE II 2013).

Duplicate Content führt neben der Beeinträchtigung der Vielfalt in den SERPs noch zu einem zweiten Problem. Die Crawling-Kapazität von Google ist begrenzt. Wenn nun sehr viel Duplicate Content gecrawlt wird, können weniger hochwertige und einzigartige Inhalte erfasst werden. Dies ist aus Sicht von Google ein Problem, da Ressourcen verschwendet werden. Auch die Webmaster sind unzufrieden, wenn möglicherweise nicht alle Seiten gecrawlt werden (vgl. MOSKWA 2008). Somit ist die Empfehlung von Google eindeutig:

„Erstellen Sie keine doppelten Seiten, Sub-Domains oder Domains, die im Grunde denselben Inhalt haben“ (MOSKWA 2008).

Es sollen einzigartige Seiten mit individuellen Inhalten erzeugt werden, die für den User nützlich und attraktiv sind. Webseiten sollen sich von anderen abheben und in erster Linie für Nutzer und nicht für Suchmaschinen erstellt werden (vgl. GOOGLE I 2013). Dies geht auch schon aus den Qualitätsrichtlinien von Google hervor (siehe dazu Kapitel 2.3). Wenn es dann doch zu Duplicate Content kommt, sollten die im folgenden Kapitel 3.3 vorgestellten Maßnahmen ergriffen werden, um ihn zu bekämpfen und Google bei der Identifikation der relevantesten Seite zu helfen.

3.3 Typische Beispiele zur Entstehung von Duplicate Content und Möglichkeiten zur Vermeidung

Duplicate Content entsteht oft unbeabsichtigt aufgrund von technischen Gegebenheiten und beispielsweise einer falschen Konfiguration des Content Management Systems (vgl. GOOGLE II 2013). Im Folgenden sollen nun einige typische Beispiele zur Entstehung von Duplicate Content aufgezeigt werden.

Fall #1: Mit und ohne WWW.

Wenn die Domain mit und ohne WWW. erreichbar ist, entsteht seitenweise Duplicate Content. Dies ist einer der ältesten und häufigsten Fälle von Duplicate Content. Das

Problem wird noch ergänzt, wenn die Seiten auch mit `http://` und `https://` aufgerufen werden können (vgl. DE VALK 2013).

Fall #2: Session IDs und Tracking

Um das Verhalten der Nutzer auf den Seiten zu analysieren, werden Session IDs und Tracking-Codes verwendet. Jeder neue Besucher bekommt eine eigene ID, die an jede besuchte URL angehängt wird. Wenn all diese neu entstandenen URLs indexiert werden, kann es zu einer Vielzahl an Duplicate Content kommen (vgl. MEYERS I 2011).

Fall #3: länderspezifische Versionen

Es kann außerdem zu Duplicate Content kommen, wenn es verschiedene Versionen einer Seite in der gleichen Sprache aber mit kleinen regionalen Unterschieden gibt. Beispielsweise kann eine englische Seite einmal für die USA (`abc.com`) und mit kleinen Abweichungen für Groß Britannien (`abc.co.uk`) und für Australien (`abc.au`) zur Verfügung stehen (vgl. STABENOW 2013).

Fall #4: Homepage Duplikate

Duplicate Content der Homepage kann nicht nur entstehen, wenn sie mit und ohne `WWW.` erreichbar ist, sondern auch, wenn sowohl das Root Verzeichnis als auch der Dokumentname indexiert werden:

`www.abc.de`
`www.abc.de/index.html`

Dies kann vor allem dann zum Problem werden, wenn sowohl intern als auch extern auf beide URLs verlinkt wird (vgl. MEYERS I 2011).

Fall #5: Sortierung und Filter

In vielen Online-Shops wird bei Produktlisten angeboten, dass diese individuell auf- oder absteigend nach Preis oder Name sortiert und nach verschiedenen Farben gefiltert werden können. Dazu wird meist ein Parameter an die URL gehängt. So entsteht Duplicate Content, da der Inhalt der Seite nicht verändert wird, aber unter einer anderen URL erreichbar ist (vgl. DE VALK 2013).

Fall #6: Paginierung

Ein weiteres Problem kann durch Paginierung entstehen, also wenn lange Artikellisten auf mehreren Seiten angezeigt werden. Die Seiten unterscheiden sich nur wenig voneinander und verursachen so Duplicate Content. Dieses Problem wird besonders in Online-Shops größer, wenn hier neben den Produkten noch ein Kategorietext angezeigt wird und sich auch die Produktnamen kaum unterscheiden. Durch die Paginierung kann es außerdem zu Problemen mit dem Crawler der Suchmaschine kommen, da dieser, um Produkte, die weiter hinten in der Liste stehen, zu erreichen, immer mehr Links folgen muss (vgl. SCHMIDT-VOIGT 2012).

Fall #7: doppelte Pfade

Manchmal können Seiten über verschiedene Pfade erreicht werden. Solange dadurch keine doppelten URLs, also Duplicate Content, entstehen, ist dies in Ordnung. Im folgenden Beispiel ist der gleiche Content aber über drei verschiedene URLs erreichbar:

www.abc.de/electronik/ipad
www.abc.de/apple/ipad
www.abc.de/tag/favoriten/ipad

Die Produktseite vom iPad kann also über zwei verschiedene Kategorien (elektronik und apple) sowie über einen user-generated Tag erreicht werden. Hierbei sind user-generated Tags besonders problematisch, da so theoretisch unendlich viele URLs generiert werden können (vgl. MEYERS I 2011).

Fall #8: Druckversion

In vielen Portalen und Nachrichtenmagazinen werden spezielle Druckversionen der Artikel ohne Werbung und Bilder angeboten. Ein schöner Service für die User, doch auch hierbei handelt es sich um internen Duplicate Content, da der Artikel Wort für Wort kopiert wurde und unter einer anderen URL erreichbar ist (vgl. DE VALK 2013).

Fall #9: Produktvariationen

Wenn in einem Online-Shop Produkte in verschiedenen Farben angeboten werden und für jede Farbe eine eigene URL kreiert wird, kann es zu sehr viel Duplicate Content

kommen. Auf den einzelnen Seiten verändert sich die Produktbeschreibung meist nicht, sondern eben nur die Farbe des Produktes. Beispiel:

www.abc.de/produkte/ipod/nano/blau

www.abc.de/produkte/ipod/nano/rot

Bei sehr vielen angebotenen Produkten im Shop können so schnell hunderte oder tausende URLs mit Duplicate Content entstehen (vgl. MEYERS I 2011).

Mit den Beispielen sollen nur die häufigsten Fälle zur Entstehung von Duplicate Content aufgezeigt werden. Die Liste ist in keinem Fall vollständig, denn es kann individuell abhängig vom Aufbau und von der Struktur jeder Seite natürlich weitere Situationen geben, in denen Duplicate Content entstehen kann. Für diese müssen dann individuelle Lösungen gefunden werden.

Liegt bei einer Website dann Duplicate Content vor, versucht Google wie beschrieben die relevanteste URL zu identifizieren und zeigt diese in den Suchergebnissen. Der Seitenbetreiber kann aber auch selbst seine bevorzugte URL festlegen und diese Google mitteilen. So hat er mehr Kontrolle darüber, welche URLs in den Suchergebnissen angezeigt werden und kann beispielsweise gezielt die Linkpopularität einer Seite stärken. Die bevorzugte URL heißt auch kanonische URL (vgl. GOOGLE V 2013). Es gibt verschiedene Möglichkeiten, Google die kanonische URL mitzuteilen, die im Folgenden erläutert werden. Außerdem wird aufgezeigt, für welche der oben beschriebenen Fälle diese Lösungsmöglichkeit angebracht wäre.

Lösungsmöglichkeit #1: Das Canonical-Tag

Mit dem Canonical-Tag kann bestimmt werden, welche URL-Version in den Suchergebnissen bevorzugt angezeigt werden soll. Außerdem können so die eingehenden Links von Duplikaten zusammengeführt werden (vgl. SCOTT 2013). Das Canonical-Tag kann sowohl intern, also innerhalb der gleichen Domain, als auch extern zu einer fremden Domain verwendet werden. „In beiden Fällen sollte es eine einzige „richtige“ und eine oder mehrere „falsche“ Seiten geben“ (KUBITZ II 2012). Zur Verwendung

des Canonical-Tags muss dann folgende Zeile angepasst in den Header der „falschen“ Seiten geschrieben werden:

```
<link rel="canonical" href="http://www.domain.de/richtige-URL"/>
```

 (vgl. KUBITZ II 2012).

Es müssen aber einige Dinge bei der Verwendung des Canonical-Tags beachtet werden. So muss der Inhalt auf den duplizierten Seiten zum Großteil auch dem Inhalt auf der kanonisierten Seite entsprechen. Es kann also nicht einfach auf irgendeine Seite mit dem Canonical-Tag verwiesen werden. Außerdem muss die kanonisierte Seite fehlerfrei erreichbar sein und darf nicht durch einen noindex-Meta-Tag gesperrt sein. Es darf auch nur ein Canonical-Tag pro Seite verwendet werden und dieser muss im <head> der Seite stehen. Wenn mehrere Canonical-Tags auf einer Seite verwendet werden oder sie sich im <body> befinden, werden sie von Google ignoriert (vgl. SCOTT 2013).

Google weist außerdem ausdrücklich darauf hin, dass es sich bei dem Canonical-Tag um keine Anweisung handelt. Mit dem Tag können lediglich Seiten vorgeschlagen werden. „Google wird dies sowie weitere Signale bei der Bestimmung der URLs mit identischen Inhalten berücksichtigen und die relevantesten Seiten berechnen, die in den Suchergebnissen angezeigt werden“ (GOOGLE VI 2013).

Das Canonical-Tag bietet sich bei den oben aufgezeigten Fällen #7, #8 und #9 als Lösungsmöglichkeit zur Vermeidung des Duplicate Contents an. Es wird bei den doppelten Pfaden und Produktvariationen jeweils eine URL als kanonische URL festgelegt, auf die dann von den anderen URLs verwiesen wird. Außerdem sollte von der Druckversion auf die Artikelversion verwiesen werden (vgl. MEYERS I 2011).

Lösungsmöglichkeit #2: 301-Weiterleitungen

Über eine 301-Weiterleitung werden Nutzer und Suchmaschinen automatisch zu einer anderen Seite weitergeleitet. Die Weiterleitung wird serverseitig über die htaccess-Datei eingerichtet und der Statuscode 301 gibt an, dass die Seite dauerhaft an einen neuen Speicherplatz verschoben wurde. Über die 301-Weiterleitung werden auch Links, die auf die alte Seite zeigen, auf die neue Seite weitergeleitet. So geht weder die Linkpower noch der Traffic der alten Seiten verloren. 301-Weiterleitungen bieten sich beispielsweise an, wenn eine Website auf eine neue Domain übertragen wurde und die Nutzer von der alten Adresse hierhin weitergeleitet werden sollen (vgl. GOOGLE VII 2013). In

Bezug auf Duplicate Content können 301-Weiterleitungen verwendet werden, wenn es für eine URL genau eine klare kanonische URL gibt und die Seite mit Duplicate Content ebenfalls Traffic erzeugt oder eingehende Links hat (vgl. MEYERS I 2011).

Bei den oben aufgezeigten Fällen bietet sich eine 301-Weiterleitung beim ersten Fall an. Hier sollte sich für eine Version mit oder ohne WWW entschieden werden und die Seiten entsprechend weitergeleitet werden (vgl. DE VALK 2013). Ebenso kann bei Fall #4 von den anderen Versionen der Homepage per 301 auf die gewünschte weitergeleitet werden und bei Fall #7 könnte statt des Canonical-Tags auch die 301-Weiterleitung verwendet werden (vgl. MEYERS I 2011).

Lösungsmöglichkeit #3: Einstellungen in den Google Webmaster Tools

In den Google Webmaster Tools können außerdem Einstellungen vorgenommen werden, die bei der Lösung von Duplicate Content-Problemen helfen können. So kann im Abschnitt „Crawling“ unter „URL-Parameter“ festgelegt werden, wie bestimmte Parameter in den URLs von Google behandelt werden sollen. Es kann beispielsweise angegeben werden, dass der Parameter „sessionid“ ignoriert werden soll. Dadurch kann Duplicate Content im Google-Index vermieden werden und die Seite kann effizienter gecrawlt werden (vgl. GOOGLE V 2013).

Die Möglichkeit der Ignorierung von Parametern sollte im oben vorgestellten Fall #2 zur Bekämpfung des Duplicate Content genutzt werden. Hier können dann Session-IDs und Tracking-Parameter angegeben werden und das Problem wird gelöst (vgl. MEYERS I 2011).

Lösungsmöglichkeit #4: Das hreflang-Attribut

Mit dem hreflang-Attribut kann den Suchmaschinen gezeigt werden, dass es weitere Sprachversionen der Website mit gleichem Inhalt gibt. So kann dabei geholfen werden, dass den Nutzern eines jeweiligen Landes die richtige Sprachversion der Website ausgespielt wird und zugleich kann Duplicate Content vermieden werden. Um anzugeben, dass der Inhalt einer Seite noch einmal in einer anderen Sprache besteht, kann ein HTML-Link-Element im <head> des HTML-Codes, im http-Header oder in der XML-Sitemap eingefügt werden. Angenommen <http://pt.beispiel.com> beinhaltet die portugiesische Version der deutschen Seite <http://de.beispiel.com>, dann wird auf der deutschen Seite folgender Link eingebunden:

```
<link rel="alternate" hreflang="pt" href="http://pt.beispiel.com" />
```

Im oben beschriebenen Fall #3 liegt Duplicate Content vor, da der gleiche englische Text für verschiedene englischsprachige Länder angeboten wird. Das hreflang-Attribut sollte hier verwendet werden, um die richtige Version für das jeweilige Land auszuspielen und zusätzlich sollten Canonical-Tags von den Seiten von Groß Britannien und Australien zur .com-Version gesetzt werden, um den Duplicate Content zu vermeiden (vgl. STABENOW 2013).

Lösungsmöglichkeit #5: Paginierung mit rel="prev" und rel="next"

Mit den HTML-Attributen rel="prev" und rel="next" können Beziehungen zwischen verschiedenen Seiten angegeben und diese miteinander verbunden werden. „Dieses Markup enthält einen starken Hinweis für Google, dass diese Seiten als eine logische Sequenz behandelt werden sollen“ (GOOGLE VIII 2013). Die Nutzer werden dabei in der Regel auf die erste Seite geleitet. Das Attribut wird wie hreflang auch im <head> einer Seite eingebunden. Dabei wird jeweils auf die vorherige und nachfolgende Seite der Sequenz verwiesen. Auf Seite zwei einer Sequenz würde das Attribut so aussehen:

```
<link rel="prev" href="http://abc.de/Artikel-Seite1.html">  
<link rel="next" href="http://abc.de/Artikel-Seite3.html">
```

Die erste Seite der Sequenz muss entsprechend kein rel="prev" verwenden und die letzte Seite kein rel="next" (vgl. GOOGLE VIII 2013).

Die Paginierung mit rel="prev" und rel="next" bietet sich im oben beschriebenen Fall #6 als Lösung an (vgl. SCHMIDT-VOIGT 2012).

Lösungsmöglichkeit #6: Die Metaangabe noindex

Über die Meta-Robots-Tags im <head> einer Seite kann das Verhalten vom Crawler gesteuert werden. Folgende Angabe besagt beispielsweise, dass die Seite weder indexiert werden darf, noch darf der Crawler den Links der Seite folgen:

```
<meta name="ROBOTS" content="NOINDEX, NOFOLLOW">
```

Es kann aber auch „NOINDEX, FOLLOW“ angegeben werden. Dann darf die Seite zwar nicht indexiert werden, den Links wird aber trotzdem gefolgt (vgl. MEYERS I 2011). Dies bietet sich als Lösung für das Duplicate Content-Problem im Fall #5 an. Die

einzelnen Seiten mit den Filtern dürfen dann nicht indexiert werden, sodass kein Duplicate Content entsteht. Den Links zu den Produktbeschreibungen darf aber trotzdem gefolgt werden, damit diese dann indexiert werden können (vgl. MEYERS I 2011).

Bei all diesen Lösungsmöglichkeiten ist es außerdem wichtig, dass auf eine einheitliche interne und externe Verlinkung geachtet wird. Wenn ein Canonical-Tag auf einer Seite gesetzt wurde, sollte auch nur die kanonische URL verlinkt werden und nicht die mit dem Duplicate Content. Wird manchmal auf die kanonische URL verlinkt und manchmal auf die mit Duplicate Content, werden unterschiedliche Signale an Google gesendet, was die Wirkung des Canonical-Tags beeinflussen kann (vgl. MEYERS I 2011). Am besten ist es natürlich, wenn erst gar kein Duplicate Content auf einer Seite entsteht. Manchmal ist dies aber schon aus technischen Gründen nicht möglich. Dann sollte schnell eine saubere Lösung für das Problem gefunden werden.

Doch bevor der Duplicate Content bekämpft werden kann, muss er erst einmal gefunden werden. Hierzu gibt es zahlreiche Tools, mit denen online nach Duplicate Content und Plagiaten gesucht werden kann. Die meisten sind kostenlos. Bei den Tools gibt es verschiedene Funktionsweisen. So kann entweder eine URL eingegeben werden und es wird nach Duplikaten gesucht oder es wird ein Text eingegeben und es werden dann URLs gesucht, auf denen dieser Text verwendet wird. Die dritte Möglichkeit ist, dass zwei URLs eingegeben werden und diese dann auf Ähnlichkeit geprüft werden. Im Folgenden soll nun jeweils ein Tool von jeder Funktionsweise kurz vorgestellt werden.

Tool #1: URL-Eingabe, Beispiel: Copyscape

URL: <http://www.copyscape.com>

Copyscape ist das wohl bekannteste und am weitesten verbreitete kostenlose Online-Tool zum Finden von Duplicate Content (vgl. SEOUPDATES.INFO 2013). Auf der Startseite wird eine URL eingegeben und das Tool gibt Seiten zurück, die den gleichen Inhalt haben. Beim Prüfen von den Domains aus dem Website-Test konnte Copyscape allerdings nur die URLs mit 100% externem Duplicate Content finden. Um schnell zu prüfen, ob ein bestimmter kompletter Artikel mehrmals vorliegt, ist das Tool aber si-

cherlich sehr gut geeignet. In der kostenpflichtigen Version kann außerdem ein Textabschnitt eingegeben werden, der dann gesucht wird.

Tool #2: Eingabe von Text, Beispiel: Dupli Checker

URL: <http://www.duplichecker.com/>

Bei Dupli Checker kann ein Text eingegeben werden und das Tool gibt alle URLs zurück, die diesen Text verwenden (siehe Abbildung 07). Als zweite Möglichkeit kann auch ein docx- oder txt-Dokument hochgeladen werden. Dann wird nach dem kompletten darin enthaltenen Text gesucht.



Abbildung 07: Ergebnisseite vom Dupli Checker (<http://www.duplichecker.com/>)

Beim Testen mit den URLs vom Website-Test konnte das Tool voll und ganz überzeugen. Es konnten jeweils alle URLs gefunden werden, egal wie lang der eingegebene Text war. Wenn das Tool dauerhaft verwendet wird, sollte ein kostenloser Account angelegt werden, da ohne Account nur eine bestimmte Anzahl an Anfragen durchgeführt werden kann. Als Alternative zum Dupli Checker bietet sich Plagium (<http://www.plagium.com/>) an. Hier werden in der kostenpflichtigen Version einige interessante zusätzliche Funktionen geboten.

Tool #3: Vergleich von zwei URLs, Beispiel: Similar Page Checker

URL: <http://www.webconfs.com/similar-page-checker.php>

Mit dem Similar Page Checker können zwei URLs miteinander verglichen werden. Es wird geprüft und angegeben wie groß die Ähnlichkeit zwischen den URLs ist.



Abbildung 08: Angabe der Ähnlichkeit zwischen zwei URLs beim Similar Page Checker
(<http://www.webconfs.com/similar-page-checker.php>)

Beim Testen mit den URLs des Website-Tests konnte das Tool überzeugen. Es wurde immer fast genau die Prozentzahl des eingebauten Duplicate Contents angegeben. Bei den in Abbildung 08 dargestellten URLs wurde beispielsweise 90% interner Duplicate Content eingebaut. Wenn jeweils die URL mit Originalcontent und mit Duplicate Content bekannt sind, ist der Similar Page Checker zur Prüfung der Ähnlichkeit als zuverlässiges Tool sehr zu empfehlen.

Tool #4: Weitere Möglichkeiten

Außerdem können die Google Webmaster Tools genutzt werden, um Duplicate Content auf der eigenen Seite zu identifizieren. Im Menü „Darstellung der Suche“ unter „HTML-Verbesserungen“ wird angegeben, welche und wie viele Title-Tags und Meta-Descriptions auf der Seite doppelt vorliegen. Dies ist ein deutlicher Hinweis darauf, dass auch der Inhalt der Seite doppelt ist, also Duplicate Content vorliegt.

Darüber hinaus kann beim Verdacht auf Duplicate Content eine Phrasensuche bei Google durchgeführt werden. Dazu werden einige Sätze eines Textes kopiert und mit Anführungszeichen bei der normalen Googlesuche eingegeben. Es werden dann alle Seiten angezeigt, die diesen Textabschnitt enthalten. Wenn nur intern, also auf der Seite selbst, nach dem Textabschnitt gesucht werden soll, kann die Phrasensuche noch durch eine Site-Abfrage ergänzt werden. Hierzu wird zusätzlich der Operator `site:domain.de` bei Google eingegeben. Dann werden nur Seiten der eingegebenen Domain angezeigt.

4 Der Website-Test

In den vorherigen Kapiteln wurde geklärt wie Suchmaschinen arbeiten, welche Rankingfaktoren eine Rolle spielen und welche Probleme durch Duplicate Content hervorgerufen werden können. In diesem Kapitel wird nun ein Website-Test durchgeführt, der das Kernstück der Arbeit behandelt. Mit dem Test soll eine Toleranzgrenze von Duplicate Content in der Suchmaschine Google identifiziert werden. Bisher gibt es nur Spekulationen über eine mögliche Grenze und es wurden kaum Tests in diesem Bereich durchgeführt. Durch diesen ausführlichen, über mehrere Wochen laufenden Test im Rahmen dieser Arbeit, sollen nun die möglichen Grenzen des Duplicate Contents identifiziert werden. Der konkrete Aufbau dieses Website-Testes mit den verschiedenen Kriterien, die zur Durchführung beachtet werden mussten sowie eine detaillierte Auswertung und Zusammenfassung der Ergebnisse werden in den kommenden Kapiteln beschrieben. Zunächst folgen aber die Hypothesen, der aktuelle Forschungsüberblick in diesem Bereich und eine Erläuterung der gewählten Methode.

4.1 Die Grundidee und die Hypothesen

Um die Toleranzgrenze für Duplicate Content bei Google zu identifizieren, wird ein Test an verschiedenen Websites durchgeführt. Dazu sollen Websites aufgebaut und auf ein bestimmtes Keyword optimiert werden. Wenn stabile Rankings erreicht sind, wird ein bestimmter Anteil an Duplicate Content auf der Seite eingebaut und die Auswirkungen auf das Ranking bei Google werden untersucht.

Der Test simuliert verschiedene Situationen. Dabei wird auf der einen Seite geprüft, welche Auswirkungen interner Duplicate Content hat. Der Duplicate Content besteht hier also innerhalb einer Domain. Dazu wird bei verschiedenen Domains jeweils ein festgelegter prozentualer Anteil an internem Duplicate Content eingebaut.

Auf der anderen Seite werden auch die Auswirkungen von externem Duplicate Content untersucht. Der Duplicate Content besteht hier also zwischen zwei Domains und nicht innerhalb der gleichen Domain wie oben. Dem entsprechend werden sowohl Seiten be-

nötigt, von denen der Content geklaut wird, als auch Seiten, bei denen der Content dann eingebaut wird. Der externe Duplicate Content wird in den gleichen prozentualen Stufen eingebaut wie der interne Duplicate Content.

In der SEO-Welt existieren verschiedene Mythen und Vorstellungen über die Auswirkungen von Duplicate Content und die Bestrafungen von Google. Um diese Mythen aufzuklären, sollen mit dem Website-Test folgende Hypothesen bestärkt oder widerlegt werden:

- Duplicate Content spielt in Bezug auf das Ranking in der Suchmaschine Google eine Rolle.
- Es gibt eine Bestrafung in Form einer Herabsetzung des Rankings, wenn sich zu viel Duplicate Content auf der Seite befindet.
- Dabei spielt es in Bezug auf das Erreichen der Toleranzgrenze keine Rolle, ob es sich um internen oder externe Duplicate Content handelt.
- Grundsätzlich wirkt sich aber interner Duplicate Content stärker auf das Ranking aus, als externer.
- Wenn erst einmal die Toleranzgrenze für Duplicate Content überschritten ist, macht es keinen Unterschied wie viel Duplicate Content auf der Seite ist, die Seiten werden dann immer um die gleiche Anzahl an Positionen im Ranking herabgesetzt.

Durch die Hypothesen werden auf der einen Seite allgemeine Fragen geklärt, wie etwa, ob sich Duplicate Content überhaupt auf das Ranking auswirkt. Auf der anderen Seite werden auch detailliertere Vermutungen aufgestellt, zum Beispiel dass sich interner Duplicate Content stärker auf das Ranking auswirkt als externer. Der Test wurde so konzipiert, dass sich die Hypothesen damit bestätigen oder widerlegen lassen.

4.2 Bisherige Tests und aktuelle Situation

Duplicate Content ist ein umstrittenes Thema unter den SEOs. Es wird viel darüber diskutiert, ob es Abstrafungen wegen zu viel Duplicate Content gibt und wie Google nun genau darauf reagiert. Auch verschiedene Mitarbeiter von Google äußern sich in Blog-

artikeln oder Videos immer wieder zu dem Thema, wie beispielsweise Matt Cutts, der Head of Webspam in seinen Webmaster-Help-Videos. In diesem Kapitel soll aufgezeigt werden, welche Meinungen aktuell über Duplicate Content in der SEO-Welt kursieren (2013) und welche Experimente und Tests es schon in diesem Bereich gab.

Bei den Recherchen sowohl in den großen bekannten SEO-Blogs als auch in kleinen privaten Blogs konnten insgesamt 23 Tests mit Bezug auf Duplicate Content gefunden werden, die im Zeitraum von Januar 2006 bis März 2013 durchgeführt wurden. Davon wurden aber nur 16 zu Ende gebracht und ausgewertet. Bei den anderen sieben Tests ist ausschließlich der Versuchsaufbau erläutert worden. Eine Mehrzahl der Tests wurde von kleinen Bloggern durchgeführt. Die großen bekannten SEO-Blogs haben kaum Tests in diesem Bereich veröffentlicht. Des Weiteren ist zu sagen, dass kein Experiment gefunden wurde, dass sich mit genau der gleichen Problematik beschäftigt wie der im Rahmen dieser Bachelorarbeit durchgeführte Website-Test. Es wurde zuvor noch nie mit verschiedenen Anteilen von internem und externem Duplicate Content und deren Auswirkungen experimentiert. Die Ergebnisse des Testes im Rahmen dieser Arbeit können also zu neuen, interessanten und bisher nicht bestätigten Ergebnissen führen. Viele der gefundenen Tests beschäftigen sich mit der richtigen Verwendung des Canonical-Tags (siehe z.B. MEYERS II 2011). Vor dem Panda-Update wurde außerdem viel mit Duplicate Content in Artikelverzeichnissen experimentiert (siehe z.B. LETSPROMOTEIT.DE 2007).

Im Folgenden werden fünf ausgewählte Experimente und Case Studies vorgestellt, die als Ausgangsbasis für diesen Website-Test, am interessantesten sind. Sie haben sich allgemein mit der Auswirkung von Duplicate Content auf das Ranking beschäftigt und verschiedene Faktoren getestet.

Experiment #1: Wie gut kommt Google mit Duplicate Content zurecht?

durchgeführt von Frank Troebel im Januar 2008 bei boeser-seo.blog.de

Ziel: Mit seinem etwas extremen Feldversuch wollte Frank Troebel herausfinden, wie gut Google mit Duplicate Content zurechtkommt und wie auf eine sehr große Masse an Duplicate Content reagiert wird. Er stellt aber keine besondere These auf, die er untersuchen möchte.

Durchführung: Troebel erstellt eine reine Affiliate-Seite mit 100 % Duplicate Content und optimiert diese auf ein starkes Keyword. Auf der Startseite verwendet er Unique Content. In drei Monaten baute er 500 Backlinks auf und ließ die Seite dann ruhen.

Ergebnisse: Zunächst wurden 20.000 Seiten indexiert und die Seite rankte für das Hauptkeyword auf Platz 10. Dann war sie etwa vier Wochen lang immer wieder mal für 2 bis 3 Tage nicht mehr in den SERPs zu finden und dann wieder auf Platz 10. Die nächsten vier Wochen wird die Seite nur noch selten in den Suchergebnissen bei Google angezeigt. Nach insgesamt 3 Monaten rankt sie teilweise wieder auf Platz 20.

Fazit: Zunächst konnte sich Google wohl nicht entscheiden, ob die Seite ranken sollte oder nicht, schließlich wurde sie als Duplicate Content identifiziert und aus dem Index entfernt, tauchte aber später trotzdem manchmal wieder auf. Wichtig ist aber die Erkenntnis, dass es sich beim Duplicate Content-Filter um einen Domainfilter handelt, da die Seite nicht nur für das Hauptkeyword, sondern auch für alle anderen Keywords ausgeschlossen wurde.

Quellen: TROEBEL I 2008, TROEBEL II 2008, TROEBEL III 2008

Experiment #2: Gibt es eine Duplicate Content Penalty?

durchgeführt von Daniel McGonagle im September 2009 bei linkvanareviews.com

Ziel: McGonagle möchte bestehende Gerüchte über die Duplicate Content Penalty aus dem Weg räumen und mit einem Experiment beweisen, dass es eine Penalty gibt.

Durchführung: Er hat dazu genau den gleichen Artikel auf drei Blogs mit unterschiedlichen Titeln gepostet. Die Blogs liegen alle auf dem gleichen Host.

Ergebnisse: Die Artikel wurden jeweils aus den SERPs entfernt, wenn eine neue Version online gestellt wurde. Es wurde also immer nur die aktuellste Version aufgeführt.

Fazit: Es gibt anscheinend eine Penalty, wenn exakt der gleiche Content unverändert mehrmals auf dem gleichen Host liegt. Dies ist aus der Sicht von Google der gleiche Ursprung.

Quellen: MCGONAGLE I 2009, MCGONAGLE II 2009

Experiment #3: Wann Duplicate Content auch gut für das Ranking sein kann

durchgeführt von John Hargrave im Juni 2010 bei searchenginejournal.com

Ziel: Nach einer Aussage von Dan Crow (Produktmanager Google Crawl Systems) kommt es im SEO nur auf zwei Dinge an: sehr guter Content und sehr gute Backlinks. Hargrave möchte in einer Case Study zeigen, dass dies auch für Duplicate Content gilt.

Durchführung: Eine Kundenseite wollte für „credit card concierge“ ranken. Also wurde dafür ein sehr guter Artikel geschrieben und Linkbuilding betrieben. Die Seite stieg auf Platz 6. Tim Ferris von der New York Times setzte auf seinem Blog aber nicht nur einen wertvollen Link, sondern kopierte und veröffentlichte auch den ganzen Artikel Wort für Wort. Dieser Post wurde dann wiederum von vielen anderen reposted, getwittert und geshared. Es entstand also viel Duplicate Content.

Ergebnisse: Die Seite des Kunden rankt weiterhin in den Top 10 und wurde nicht abgestraft.

Fazit: Wenn guter Content dupliziert wird und dadurch sehr gute Backlinks zustande kommen, gibt es keine Probleme oder Bestrafungen für die Seite, sondern eher Vorteile und gewinnbringende Situationen.

Quelle: HARGRAVE 2010

Experiment #4: Welche Auswirkungen hat das Panda-Update?

durchgeführt von Tommy Landry im Februar bis Juni 2011 bei returnnonnow.com

Ziel: Landry hörte davon, dass das Panda-Update geplant sei und dass hier Duplicate Content besonders stark abgestraft werden sollte. Mit dem Experiment wollte er herausfinden, wie hart das Update tatsächlich wird.

Durchführung: Er setzte für das Experiment zwei Seiten auf. Eine mit 100 % Duplicate Content und eine mit einigen Seiten mit Unique Content und einem bestimmten Anteil an Duplicate Content. Die Seiten stellte er bereits vor dem Rollout vom Panda-Update online und beobachtet die Entwicklung danach.

Ergebnisse: Die Seiten wurden hart bestraft als Panda im Februar 2011 online ging. Die Seite mit 100 % Duplicate Content war fast komplett aus den Suchergebnissen verschwunden, die andere hatte einen Einbruch von weniger als 50 %.

Fazit: Es spielte in Bezug auf die Bestrafung durch das Panda-Update also durchaus eine Rolle wie viel Duplicate Content sich auf der Seite befand.

Quelle: LANDRY 2011

Experiment #5: Mit geklautem Content besser ranken als das Original

durchgeführt von Dan Petrovic im November 2012 bei dejanseo.com.au

Ziel: Es gilt die Annahme, dass wenn es mehrere Domains mit dem gleichen Content gibt, diejenige ins Ranking aufgenommen wird, die den höchsten PageRank hat. Petrovic möchte genau dies mit einem Experiment beweisen.

Durchführung: Petrovic hat von vier anderen qualitativ hochwertigen und von Google anerkannten Seiten Content geklaut und auf jeweils einer brandneuen Seite veröffentlicht. Durch Backlinkaufbau hat er versucht, einen höheren PageRank zu erreichen als die Originalseiten und also höher zu ranken.

Ergebnisse: Bei zwei Fällen gelang es, dass die Originalseite aus den SERPs verschwand, bei einer rankte die Duplicate Content-Seite höher als das Original und bei der vierten Seite gelang es nicht, die Originalseite anzugreifen.

Fazit: Es ist tatsächlich für eine brandneue Seite möglich mit geklautem Content höher zu ranken als die etablierte Originalseite.

Quelle: PETROVIC 2012

Die meisten der hier aufgezeigten Experimente liegen schon einige Jahre zurück und wurden vor dem großen Panda-Update durchgeführt. Daraus folgt, dass die Schlussfolgerungen über die Arbeitsweise von Google vermutlich heute nicht mehr zutreffen. Dass die Experimente aber schon einige Zeit zurückliegen, zeigt auch, dass Duplicate Content ein ständig diskutiertes Thema ist und sich die SEOs schon lange damit beschäftigen. Auch von Google selbst kommen immer wieder Hinweise zur Behandlung

von Duplicate Content in den SERPs. Es kursieren auch aktuell verschiedene Theorien über die Duplicate Content Penalty.

Martin Mißfeldt gehört beispielsweise zu den SEOs, die glauben, dass es keine Duplicate Content Penalty gibt. Das Zustandekommen seiner Annahme beschreibt er in dem Artikel „Duplicate Content Penalty nur ein Seo-Gespenst?!“ vom 22.04.2013. Er führt hier als Beweis verschiedene Aussagen von Matt Cutts und John Müller zum Thema an und kommt zu folgendem Schluss: Zu Problemen kann es nur bei großen Sites kommen, wenn durch zu viel Duplicate Content Crawler-Kapazität verschenkt wird oder wenn man bei einer Seite die doppelten Inhalte herausfiltert und nur noch sehr wenig Unique Content übrig bleibt. Dann hat man aber ein Thin Content-Problem und wird deswegen bestraft (vgl. MISSFELDT 2013). Martin Mißfeldt hat zu seinen Theorien keinerlei Experimente oder Tests durchgeführt. Er bezieht sich hier nur auf die Aussagen von Google selbst.

Diesen Aussagen stand Julian Dziki von der SEOKRATIE schon 2010 kritisch gegenüber, als er sagte: „Was Google sagt und was Google tut sind zwei völlig unterschiedliche Dinge“ (DZIKI 2010). Dieser Meinung ist auch Barry Schwartz vom Search Engine Land. Er glaubt den neuesten Aussagen von Matt Cutts nicht. Cutts wird in einem seiner Webmaster-Help-Videos die Frage gestellt, ob Duplicate Content bei Disclaimern und Terms of Services ebenfalls bestraft wird. Er antwortet: “I wouldn't stress about this unless the content that you have duplicated is spammy or keyword stuffing” (CUTTS 2013). Cutts sagt dies zwar im Zusammenhang mit der gestellten Frage, dennoch sind viele SEOs über die Aussage verwundert. Denn Cutts formuliert die Antwort so, als ob nur Duplicate Content mit schlechter Qualität abgestraft wird. Laut Barry Schwarz sind SEOs und Webmaster hier anderer Meinung. Es gab schon oft ähnliche Stellungnahmen von Google, doch viele Webmaster haben auch Erfahrungen mit Abstrafungen wegen Duplicate Content gemacht, so Schwarz (vgl. SCHWARZT II 2013).

Die Diskussion ob es eine Duplicate Content Penalty gibt und wenn ja, in welchem Maß, ist also auch 2013 trotz scheinbar eindeutiger Aussagen von Seiten von Google nicht beendet. Es werden immer wieder neue Theorien aufgestellt und diskutiert. Einen allumfassenden Test mit wirklich endgültigen und glaubhaften Aussagen gab es bisher noch nicht. Einige Ideen, wie so ein Test aussehen könnte, werden in Kapitel 6 geliefert.

Eine sehr interessante Zusammenstellung von vielen verschiedenen SEO-Experimenten findet sich außerdem bei searchenginejournal.com (siehe TUKNOV 2013). Diese behandeln nicht nur Duplicate Content-Themen, sondern auch viele andere diskutierte Fragen aus der SEO-Szene.

4.3 Die gewählte Methode

Als Methode für den Test wurde das naturwissenschaftliche Experiment gewählt:

„Gemäß der lateinischen Bedeutung von *experimentum*, nämlich Versuch, Probe, Beweis, handelt es sich beim Experiment um eine Untersuchung, bei der gezielt eine bestimmte Annahme oder Vermutung geprüft und entweder bewiesen oder widerlegt werden soll“ (E-TEACHING.ORG 2008).

Die Annahme, die hier untersucht werden soll, ist also, dass Duplicate Content in Bezug auf das Ranking eine Rolle spielt und es eine bestimmte Toleranzgrenze von Duplicate Content bei Google gibt. Um dies zu überprüfen, ist ein bestimmter Versuchsaufbau nötig bei dem Daten erhoben werden. Wesentliches Merkmal hierbei ist, dass eine Variable während des Experiments bewusst verändert, also manipuliert wird. Dies ist die unabhängige Variable. Im Experiment sollen ihre Auswirkungen auf die unveränderte abhängige Variable untersucht werden (vgl. GORNIAK et al. 2010). Im vorliegenden Website-Test ist der Einbau von Duplicate Content die unabhängige Variable und die daraus möglicherweise resultierende Rankingveränderung die abhängige Variable. Genau dieser Zusammenhang soll im Test bewiesen oder widerlegt werden. Bevor die abhängige Variable aber untersucht werden kann, muss sie operationalisiert werden. Sie muss also messbar sein, damit Auswirkungen der unabhängigen Variablen festgestellt werden können (vgl. GORNIAK et al. 2010). Dies wurde im Website-Test so gelöst, dass täglich die aktuelle Rankingposition jeder URL gemessen wird und diese dann jeweils vor und nach dem Einbau des Duplicate Content verglichen werden. Neben der unabhängigen Variablen können sich außerdem Störvariablen auf die abhängige Variable auswirken und die Auswertung der Ergebnisse erschweren. Die interne Validität, also die Eindeutigkeit der Schlussfolgerungen, kann so behindert werden (vgl. GORNIAK et

al. 2013). Beim Website-Test stellen die anderen Rankingfaktoren die Störvariablen dar. Sie haben ebenfalls Einfluss auf das Ranking und erschweren die Aussage, dass eine mögliche Veränderung des Rankings auf dem Einbau von Duplicate Content beruht. Um dieses Problem zu vermeiden und die Störvariablen also zu kontrollieren, werden die Websites so aufgebaut, dass möglichst viele der anderen Rankingfaktoren ausgeschlossen werden oder sich neutralisieren (siehe Kapitel 4.4 zum Aufbau der Domains). So kann eindeutig festgestellt werden, dass mögliche Veränderungen des Rankings unmittelbar nach dem Einbau von Duplicate Content auch wirklich darauf beruhen. Die interne Validität ist also wieder gegeben.

Um das Experiment, also in diesem Falle den Website-Test, durchführen zu können und aussagekräftige Ergebnisse zu erhalten, müssen einige Gütekriterien erfüllt werden. Dabei wird zwischen Haupt- und Nebengütekriterien unterschieden. Die Hauptgütekriterien besagen, dass ein Experiment objektiv, reliabel und valide sein muss (vgl. LIERNERT / RAATZ 1998, S. 7). Objektivität bedeutet, dass die Ergebnisse unabhängig vom Versuchsleiter zustande kommen müssen. Dies ist hier gegeben, weil der Versuchsleiter nur über den Aufbau der Seiten einen indirekten Einfluss auf das Ranking hat. Da die Seiten aber alle sehr ähnlich aufgebaut werden (zum Ausschluss der Störvariablen), ist dieser Einfluss bei allen Seiten gleich und normalisiert sich so. Die Reliabilität beinhaltet die Zuverlässigkeit der Methode und gibt an, ob der gefundene Wert auch bei wiederholten Messungen wieder erfasst werden kann (vgl. E-TEACHING.ORG 2008). Wenn die Webseiten bei einem wiederholten Website-Test genauso aufgebaut werden und in der Zwischenzeit kein Update bei Google ausgespielt wurde, sollten zumindest ähnliche Ergebnisse zustande kommen. Wenn ein Update ausgespielt wurde, besonders eines, das sich auf den Content bezieht, oder etwas an den Webseiten verändert wird, dann werden vermutlich andere Ergebnisse mit dem Test erreicht. In dieser Situation wäre es dann aber besonders interessant herauszufinden, wie Google nun auf den Duplicate Content reagiert. Die Validität als drittes Hauptgütekriterium besagt schließlich, wie genau wirklich das gemessen wurde, was auch gemessen werden sollte. Hier kann beim Website-Test davon ausgegangen werden, dass dies mit der Messung der Rankingveränderung voll erfüllt wird. Zu Problemen kann es hier zum Beispiel bei psychologischen Experimenten kommen, wenn komplexe Faktoren wie etwa die Kreativität gemessen werden sollen (vgl. E-TEACHING.ORG 2008).

Darüber hinaus sollten, wenn möglich, auch die vier Nebengütekriterien erfüllt werden. Das Experiment soll hier normiert, vergleichbar, ökonomisch und nützlich sein (vgl. LIENERT / RAATZ 1998, S. 7). Die Normierung und Vergleichbarkeit, also dass die Ergebnisse mit denen von anderen Experimenten vergleichbar sind, spielen in Bezug auf den Website-Test keine große Rolle, da es vorher noch keine umfangreichen Tests im Bereich Duplicate Content bei Google gab (vgl. E-TEACHING.ORG 2008). Ebenfalls vernachlässigt werden kann die Ökonomie. Das Experiment findet im Rahmen einer Bachelorarbeit und nicht in der freien Wirtschaft statt. Die Nützlichkeit wiederum ist gegeben. Die gemessenen Daten des Experiments haben zumindest für die SEO-Szene einen praktischen und wissenschaftlichen Nutzen. Wie groß dieser ist, hängt schließlich von den letztendlichen Ergebnissen ab.

Beim hier durchgeführten Website-Test werden also alle drei Hauptgütekriterien eines wissenschaftlichen Experimentes plus das Nebengütekriterium Nützlichkeit erfüllt. Die anderen Nebengütekriterien spielen hier keine Rolle und können beim Design des Testes vernachlässigt werden.

4.4 Aufbau und Rahmenbedingungen

Ziel des Testes ist es, herauszufinden, wie Google auf den Einbau von Duplicate Content reagiert und eine mögliche Toleranzgrenze von Duplicate Content zu identifizieren. Die Seiten für den Test müssen also so aufgebaut werden, dass bei einer möglichen Rankingveränderung mit Sicherheit gesagt werden kann, dass diese wegen des Einbaus von Duplicate Content entstand. Dazu sollen andere Rankingfaktoren, wie etwa die Keyworddichte, Keyword-Domains oder die Anzahl von Backlinks, so gut es geht ausgeschlossen werden. Damit dies gelingt, wurden verschiedene Kriterien für die Keywords, die Domainnamen und die Texte aufgestellt.

Um den Test durchführen zu können, müssen vier verschiedene Typen von Domains erstellt werden:

- Typ A: auf das Keyword optimierte Seiten; an dieser Domain werden keine Veränderungen bezüglich Duplicate Content vorgenommen, sie dient zu Vergleichszwecken
- Typ B: hier wird Duplicate Content eingebaut und zwar ein festgesetzter prozentualer Anteil am Text; der Duplicate Content bezieht sich hier auf die Domain selbst (interner Duplicate Content)
- Typ C: hier wird ebenfalls Duplicate Content eingebaut, aber von einer anderen Domain (externer Duplicate Content)
- Typ D: dies ist die Domain von der Duplicate Content bei Domaintyp C eingebaut wird

Der Duplicate Content soll in den folgenden fünf festen prozentualen Anteilen eingebaut werden:

10 %, 30 %, 50 %, 70 %, 90 %

Diese gelten sowohl für den Einbau von externem als auch von internem Duplicate Content. Außerdem soll es eine Domain geben, die eine andere Domain komplett kopiert, also 100 % externen Duplicate Content aufweist. Die Anteile beziehen sich jeweils auf den Text einer Seite. Es wird also ein Textabschnitt, der dem gewünschten Anteil entspricht, kopiert und ein gleich großer Textabschnitt auf einer anderen Seite wird dadurch ersetzt (siehe Abbildung 09). So ändert sich die Größe der Domain nicht. Es wird jeweils nur der Text der Seite mit Duplicate Content versehen. Title-Tags und Meta-Descriptions bleiben für jede URL individuell erhalten.

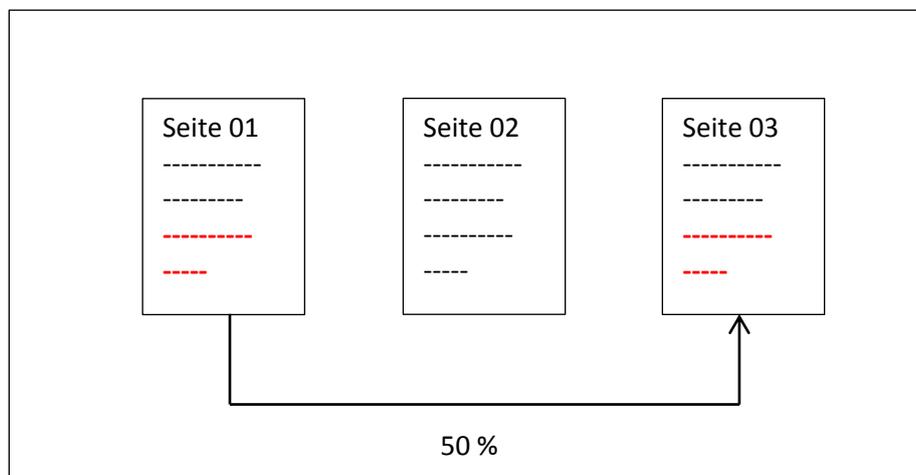


Abbildung 09: Der Einbau von internem Duplicate Content

Es ist nicht vorherzusehen, wie lange Google braucht, um den Duplicate Content zu erkennen und darauf zu reagieren. Durch den einmaligen Einbau von festen Anteilen von Duplicate Content pro Seite wird Google dafür genug Zeit gegeben, sodass es am Ende des Testes zu guten Ergebnissen kommen sollte. Dem gegenüber stünde die Methode, den Anteil bei den Domains schrittweise zu erhöhen und zu überprüfen, wann Google auf die Veränderung reagiert. Hierbei wäre aber das Problem, dass man nicht genau weiß, ob Google den eingebauten Duplicate Content schon erkannt und möglicherweise darauf reagiert hat, bevor die nächste Stufe einbaut wird. Aus diesem Grund wurde hier die Methode mit festen prozentualen Anteilen gewählt.

Für den Einbau des Duplicate Contents gibt es verschiedene Möglichkeiten. So kann er von der Startseite oder von einer Unterseite geklaut werden. Ebenso kann er natürlich auch auf der Startseite oder einer Unterseite eingebaut werden. Daraus ergeben sich verschiedene Modelle zum Verschieben des Contents: von der Startseite auf eine Unterseite, von einer Unterseite auf die Startseite, von einer Unterseite auf eine Unterseite und bei externem Duplicate Content noch von einer Startseite auf eine Startseite. Bei den Seiten eines Typs mit entsprechender prozentualer Stufe soll immer die gleiche Variante gewählt werden, also beispielsweise bei Typ B mit 10 % immer von der Startseite auf eine Unterseite. Außerdem kann der Content von unten oder von oben beginnend geklaut werden. So wird einmal die `<h1>` mitgenommen und einmal nicht. Dies soll so variiert werden, dass es jeweils einige Seiten mit geklauter `<h1>` gibt und einige ohne.

Um gute Ergebnisse erzielen zu können, werden insgesamt 30 Domains aufgebaut. Diese Anzahl setzt sich wie folgt zusammen:

- Insgesamt gibt es zehn Domains von Typ B mit internem Duplicate Content. Von jeder prozentualen Stufe sollen immer zwei Domains erstellt werden, damit die möglichen Rankingveränderungen verglichen und von der zweiten Domain bestätigt oder widerlegt werden können.
- Von den Typ-C-Domains wurden elf Domains erstellt. Hier werden ebenfalls jeweils zwei Domains mit entsprechender prozentualer Stufe aufgebaut, also 10, und noch die eine weitere, bei der eine Domain komplett kopiert wird.
- Hinzu kommen insgesamt sechs Domains von Typ D, von denen der Duplicate Content geklaut wird.

- Zum Vergleich gibt es insgesamt drei Typ-A-Domains, bei denen kein Duplicate Content eingebaut wird.

Die Domains sollen auf drei verschiedene Keywords optimiert werden, also 10 Domains pro Keyword. Jede Domain soll drei Texte (URLs) enthalten. Somit gibt es also bei jedem Keyword 30 URLs, die darauf ranken. Durch die Verwendung von drei verschiedenen Keywords besteht keine Gefahr wegen Spam-Verdacht von Google ausgeschlossen zu werden. Außerdem können die Seiten schneller online gestellt werden und stabile Rankings erreichen, sodass dann der Duplicate Content eingebaut werden kann. Denn dies kann bei der Optimierung auf drei verschiedene Keywords parallel geschehen, im Gegensatz dazu müssten bei der Optimierung von 30 Domains auf ein Keyword die Seiten nacheinander bearbeitet werden. Da nur 12 Wochen zur Bearbeitung zur Verfügung stehen, ist es zeitlich gesehen effizienter und leichter die Domains auf drei Keywords zu verteilen. Somit bleibt am Ende mehr Zeit zur Beobachtung der Entwicklung der Rankings.

Um fremde Einflüsse von anderen Domains auszuschließen, soll auf fiktive Keywords optimiert werden. Dadurch findet der Test in einer isolierten Testumgebung statt und Veränderungen im Ranking können leicht untersucht werden. Allerdings fehlt hier dann die typische Wettbewerbssituation, die die SERPs normalerweise prägt. Diese mit aufzunehmen, also auf ein reales Keyword zu optimieren und den Test hier durchzuführen, ist im Rahmen der Bachelorarbeit nicht möglich. Die erstellten Seiten müssten sich erst gegen andere im Ranking behaupten, wodurch eine lange Vorlaufzeit bis zum Beginn des eigentlichen Testes entstehen würde. Um einen idealen Test durchzuführen, sollte auf eine zeitliche Beschränkung verzichtet und der Test anhand von realen Keywords durchgeführt werden (mehr dazu in Kapitel 6).

Folgende fiktive Keywords wurden zur Optimierung gewählt:

- Elbwelpe
- Randballsport
- Jahreszeitentechnologie

Diese Keywords hatten zuvor keine Suchergebnisse bei Google, eignen sich also gut für die künstliche Testumgebung ohne fremde Einflüsse. Außerdem lassen sich die gewählten Keywords in verschiedene Richtungen interpretieren, sodass sich gut jeweils 30 ver-

schiedene Texte dazu schreiben lassen können. Auch dies musste bei der Wahl beachtet werden. So kann „Elbwelppe“ etwa eine kleine Hunderasse, eine Barkasse für Hafensrundfahrten oder auch ein Modelabel sein.

Die verschiedenen Domaintypen wurden nach folgendem Schema auf die einzelnen Keywords verteilt (siehe Abbildung 10). Bei jedem Keyword gibt es eine Domain vom Typ A zu Vergleichszwecken und zwei Typ-D-Domains. Die anderen Domains der Typen B und C wurden entsprechend aufgeteilt. Der Buchstabe steht dabei jeweils für den Typ der Domain, die erste Zahl für den prozentualen Anteil an Duplicate Content, der eingebaut werden soll und die zweite Zahl ist die Nummer der Domain. Bei den Typ-C-Domains findet sich noch eine Angabe zu jeweils einer D-Domain. Dies ist die Domain, von der dann der Content geklaut wird.

Elbwelppe	Randballsport	Jahreszeitentechnologie			
A - 1	A - 2	A - 3			
D - 1	D - 3	D - 5			
D - 2	D - 4	D - 6			
B - 10 - 1	B - 10 - 2	B - 30 - 1			
B - 50 - 1	B - 30 - 2	B - 50 - 2			
B - 70 - 2	B - 70 - 1	B - 90 - 1			
B - 90 - 2	C - 30 - 1	D3	C - 10 - 2	D5	
C - 50 - 1	D1	C - 50 - 2	D4	C - 70 - 1	D6
C - 30 - 2	D1	C - 70 - 2	D3	C - 90 - 2	D6
C - 10 - 1	D2	C - 90 - 1	D4	C - 100 - 1	D5

Legende:

A -	Vergleichsdomain ohne DC
B -	Einbau interner DC
C -	Einbau externer DC
D -	beklaute Domain externer DC
- x -	Prozentangabe für Einbau DC
- y	Nummer der Domain

Abbildung 10: Verteilungsplan bei der Optimierung auf drei Keywords

Die Domains, die für den Test registriert wurden, können Tabelle 01 entnommen werden. Sie wurden alle beim gleichen Anbieter gehostet, sodass hier Unterschiede ausgeschlossen werden können. Die Domains für ein Keyword wurden außerdem immer am gleichen Tag registriert. Sie haben also immer alle das gleiche Domainalter, wodurch auch dieser Rankingfaktor ausgeschlossen wurde. Die Domains wurden nur für die Dauer des Testes registriert und danach wieder entfernt.

Elbwelppe	Randballsport	Jahreszeitentechnologie
hund-vom-fluss.de	mein-mannschaftssport.de	jahreszeitenkraftwerk.de
mode-in-meinen-augen.de	sieben-meter.de	jahres-magazin.de
hund-fritz-und-ich.de	student-macht-sport.de	draussen-schlafen.de
schwimm-mit-mir.de	mannschaftssportturnier.de	wetterbeeinflussung.de
fraulein-chuckles-geschichten.de	internationaler-mannschaftssport.de	mein-voller-kleiderschrank.de
barkasse-im-hafen.de	weil-wir-sportler-sind.de	technik-im-auto.de
weil-wir-kinder-sind.de	handball-in-den-medien.de	temperatur-im-auto.de
hundetatzentanz.de	sport-denken.de	terrarium-technik.de
campus-zeitschrift.de	schueler-machen-sport.de	jahresbauer.de
meine-bands.de	ballsport-verein.de	handballspray.de

Tabelle 01: Domains, die für den Test registriert wurden

Um die Grundbedingung der Randomisierung eines experimentellen Designs zu erfüllen (also der zufälligen Aufteilung auf die Versuchsgruppen), wurden die Domains nach der Registrierung per Zufall den einzelnen Domaintypen A bis D zugeteilt. Es stand also beispielsweise vor der Registrierung nicht fest, dass die Domain fraulein-chuckles-geschichten.de die Vergleichsdomain beim Keyword „Elbwelppe“ wird und hier also kein Duplicate Content eingebaut wird. Die zufällige Zuteilung der Typen auf die Seiten noch bevor die Seiten online gestellt werden, führt dazu, dass jeweils die beiden Seiten, bei denen der gleiche Anteil an internem oder externem Duplicate Content eingebaut wird, vor dem Einbau auf völlig unterschiedlichen Positionen stehen. Es kann also passieren, dass einmal 90 % interner Duplicate Content bei einer URL auf Position 1 eingebaut wird und einmal bei einer URL auf Position 25. So kann zugleich überprüft werden, ob die ursprüngliche Position im Ranking auch eine Rolle in Bezug auf die Reaktion von Google spielt. Eine andere Variante wäre, erst kurz vor dem Einbau des Duplicate Contents festzulegen, bei welcher Domain welcher Anteil an Duplicate Content eingebaut wird. Dann könnte etwa immer bei zwei Domains der gleichen Position bei verschiedenen Keywords der gleiche Anteil an Duplicate Content eingebaut werden. Dann würde aber die Grundbedingung der Randomisierung eines experimentellen Designs nicht erfüllt werden und somit ist diese Variante hier nicht zulässig.

Um auch weitere Rankingfaktoren so gut es geht auszuschließen, sollen die Domains recht ähnlich und alle nach den gleichen Kriterien aufgebaut werden. So soll jede Domain aus drei URLs mit jeweils einem Text von 300 Wörtern Länge bestehen. Es sind also alle Domains gleich groß und es gibt keine Unterscheidung zwischen großen und

kleinen Domains. Die Seiten werden alle zunächst mit drei Texten mit Unique Content bestückt. Diese werden dann später durch Duplicate Content ersetzt. Auf den Seiten befindet sich nur Text. Es werden keine Bilder oder andere Elemente verwendet, die das Ranking zusätzlich beeinflussen könnten. Außerdem wird jeder Domain ein Impressum mit Haftungs- und Urheberrechtshinweisen hinzugefügt. Es enthält außerdem den Hinweis, dass es sich um Domains mit fiktivem Inhalt zu Testzwecken handelt. Die Impressumsseiten werden nicht auf die Keywords optimiert, somit spielen sie auch für das Ranking keine Rolle.

Die Domains werden mit Wordpress aufgebaut und haben alle das gleiche Theme. Bei allen Seiten wird das SEO-Plugin von Yoast installiert, aber keine weiteren Plugins. Der Blogtitel entspricht dem Domainnamen, der Untertitel ist der Domaintyp (z.B. B-90-1). Die Konfiguration und Grundeinstellungen sind bei allen Domains gleich. So wird beispielsweise überall die Kommentar-Funktion deaktiviert.

Es gibt keine Keyword-Domains, da diese noch immer einen Einfluss auf das Ranking haben, wenn auch einen geringeren im Vergleich zu den letzten Jahren (vgl. SEARCHMETRICS 2013). Die Domain hat einen neutralen Namen, der aber thematisch zum Inhalt der Seite passt. Der erste Text befindet sich auf der Startseite, die beiden anderen sind dann jeweils unter folgender URL erreichbar: <http://domain.de/hauptüberschrift.html>. In der URL werden Wörter durch Bindestriche getrennt und die Umlaute werden wie folgt ersetzt: ä = a, ö = o, ü = u, ß = ss.

Alle Seiten sollen mit individuellen Title-Tags und Meta-Descriptions versehen werden. Bei allen Title-Tags der Seiten steht das Keyword an erster Position. Sie sind alle 60 bis 65 Zeichen lang und nach folgendem Schema aufgebaut:

<Keyword> <Füllwörter> | domain.de

In den Meta-Descriptions kommt das Keyword zwei Mal vor. Sie haben jeweils eine Länge von 150 bis 155 Zeichen und bestehen aus zwei Sätzen. Der erste Satz gibt eine kurze Zusammenfassung des Inhalts, der zweite Satz enthält den Call-to-Action. Im Menüpunkt wird das Keyword nicht genannt.

Für die Domains wird kein Backlinkaufbau betrieben. Backlinks sind nach wie vor einer der wichtigsten Rankingfaktoren (vgl. SEARCHMETRICS 2013). Durch verschiedene Anzahlen von Verlinkungen und verschiedene Qualitäten der Backlinks wird das Ran-

king stark beeinflusst. Es wäre sehr schwer und nahezu unmöglich, für alle Testseiten die gleiche Anzahl und vor allem qualitativ gleichwertige Backlinks aufzubauen. Somit würde es also einen Faktor geben, der das Ranking stark verzerren und sich möglicherweise auch unterschiedlich auf den Einfluss von Duplicate Content auswirken könnte. Um dies zu vermeiden, werden die Seiten gar nicht extern verlinkt. Guter Backlinkaufbau ist außerdem sehr zeitaufwändig und kann recht lange dauern. Dies für alle Seiten gleichermaßen zu betreiben, wäre im Rahmen der Bachelorarbeit nicht möglich gewesen. An dieser Stelle sei auch wieder auf den idealen Test aus Kapitel 6 hingewiesen. Bei entsprechenden Kapazitäten würde es sich durchaus lohnen, die Auswirkungen von Duplicate Content bei Seiten mit und ohne Backlinks zu vergleichen.

Wie bereits in Kapitel 2 beschrieben, wirken sich auch verschiedene Content-Faktoren unterschiedlich auf das Ranking aus. So spielt etwa die Gesamtlänge oder auch die Position der Keywords eine Rolle. Damit keine Texte aufgrund von der Struktur im Ranking bevorzugt werden, sollen sie ebenfalls recht ähnlich und nach folgenden Kriterien aufgebaut werden:

- Gesamtlänge: 300 Worte (ohne Überschriften)
- Keywordnennung: insgesamt 7 Mal (ohne Überschriften)
- Unterteilung in drei Abschnitte mit einer Hauptüberschrift und zwei Zwischenüberschriften
- Hauptüberschrift: H1, Position Keyword am Anfang
Zwischenüberschrift 1: H2, Position Keyword egal
Zwischenüberschrift 2: H2, Position Keyword egal
- Abschnitt 1: Keyword 3 Mal, Keyword an erster bis dritter Position
Abschnitt 2: Keyword 2 Mal
Abschnitt 3: Keyword 2 Mal
- kein , oder <i>
- keine externen Links
- 1 interner Link je Text

Durch die gleich häufige Nennung des Keywords in den Texten, wird der Rankingfaktor Keyworddichte ausgeschlossen. Der Inhalt der <h>-Überschrift und die Nennung des Keywords hier hat außerdem auch eine große Bedeutung für das Ranking (vgl. SEARCHMETRICS 2013). Durch die gleiche <h>-Tag-Struktur der Seiten wird auch die-

ser Rankingfaktor neutralisiert. Außerdem gibt es keine weiteren Hervorhebungen von Wörtern (durch , oder <i>), die möglicherweise das Ranking beeinflussen könnten.

Aus den Kriterien ergibt sich folgende Struktur für die Texte (Abbildung 11):

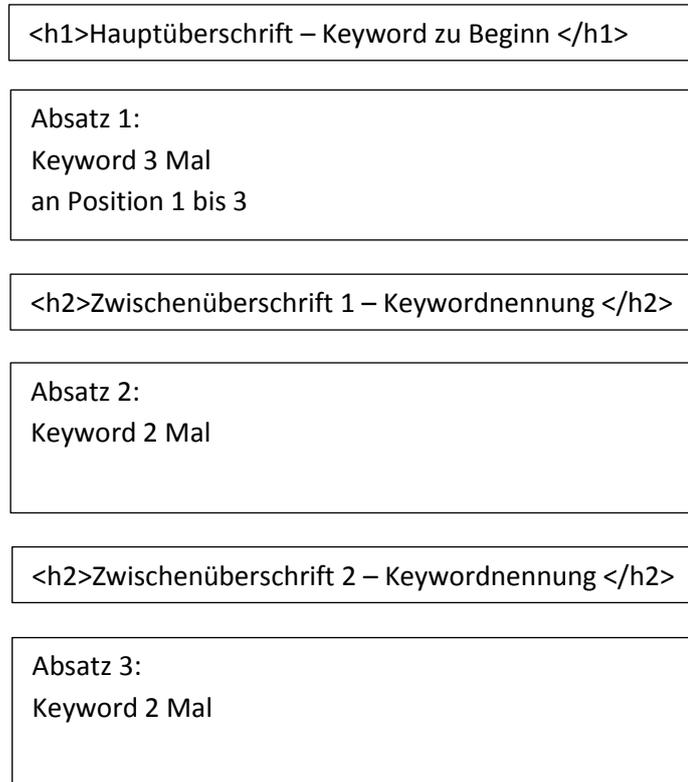


Abbildung 11: Struktur für die Texte

Alle Kriterien wurden ähnlich folgendem Beispiel bei allen Domains umgesetzt (Abbildung 12):

Domainname: BARKASSE IM HAFEN

Domaintyp: B-70-2

Menüpunkte ohne Keyword-nennung: DIE TRADITIONSBARKASSE, DIE CREW, DIE FAHRTEN, IMPRESSUM

Die Fahrten

16. JUNI 2013

Elbwelpe auf großer Fahrt

Der Elbwelpe läßt in den Sommermonaten von April bis September sechs Mal täglich zur großen Hafensrundfahrt ein. Die Tour dauert circa eine Stunde und führt vorbei an großen Pötten, Containern und neuen Häusern. Die Elbwelpe startet an den Landungsbrücken. Dann geht es an der Blohm + Voss Werft vorbei zum Containerterminal. Von hier aus geht die Fahrt durch die Ellerholz Schleuse zur Hafen-City und dann durch die Speicherstadt wieder zurück zu den Landungsbrücken. Während der gesamten Fahrt gibt Elbwelpe Kapitän Peter Petersen interessante Informationen und erzählt Geschichten aus vergangenen Tagen über den Hafen, die Elbe und das Schiff. Es können an Board Getränke und kleine Snacks erworben werden. Auch Kinderwagen und Hunde können mit auf die Fahrt genommen werden. Zudem ist die Barkasse barrierefrei.

interne Link zum Text „Die Crew“

Mit der Elbwelpe durch die Fleets

Zwei Mal täglich startet die Elbwelpe zusätzlich zur großen Fleetfahrt von den Landungsbrücken. Hier kann die Innenstadt vom Wasser aus erkundet werden. Es geht durch die Kanäle bis zum Rathausmarkt, dann zurück durch die Altstadt bis zur Hafen-City. Nach einem Abstecher durch die Speicherstadt kommt die Elbwelpe dann wieder sicher an den Landungsbrücken an. Die große Fleetfahrt dauert circa zwei Stunden und ist besonders für Architekturliebhaber, Geschichtsinteressierte und Leute geeignet, die die Stadt aus einem anderen Blickwinkel kennen lernen wollen. Genauso wie bei der großen Hafensrundfahrt erzählt der Kapitän viele spannende Geschichten und für das leibliche Wohl ist gesorgt. Kommen Sie an Board und tauchen Sie in die Geschichte Hamburgs ein!

Tagesausflüge mit der Elbwelpe

Die Elbwelpe kann außerdem für Tagesausflüge oder Gruppenfahrten gebucht werden. Bis zu 50 Personen haben auf dem Schiff Platz. Es gibt verschiedene Angebote etwa für Klassenfahrten, Seniorengruppen oder auch Junggesellenabschiede. Die Elbwelpe ist für (fast) jeden Spaß zu haben. Kommen Sie einfach vorbei und informieren sich. Je nach Wunsch können auch ein Fischerbuffet oder besondere Getränke geordert werden. Wir freuen uns sehr auf Ihren Besuch!

erster Absatz, 3 Mal Keyword

zweiter Absatz, 2 Mal Keyword

dritter Absatz, 2 Mal Keyword

Zwischenüberschrift <h2> mit Keyword

Zwischenüberschrift <h2> mit Keyword

Abbildung 12: Screenshot der Seite barkasse-im-hafen.de/elbwelpe-auf-grosser-fahrt/ als Beispiel für die Umsetzung der Kriterien, optimiert wurde auf das Keyword „Elbwelpe“ (aufgenommen am 05.07.2013)

Im Footer der Seiten wurde außerdem noch zusätzlich intern auf alle anderen URLs der Domain verlinkt (siehe Abbildung 13). So soll gewährleistet sein, dass der Crawler auch immer alle URLs der Domain besucht und die Seiten schnell indexiert werden.



Abbildung 13: Footer der Seite barkasse-im-hafen.de/elbwelpe-auf-grosser-fahrt/ als Beispiel für interne Verlinkung auf alle anderen URLs der Domain (aufgenommen am 05.07.2013)

4.5 Durchführung des Testes

Der Test soll auf folgende Art und Weise durchgeführt werden: Zunächst sollen die Domains mit jeweils drei unigen auf das Keyword optimierten Texten online gestellt werden. Die Domains eines Keywords werden jeweils innerhalb von höchstens zwei Tagen online gestellt. So werden die Seiten alle etwa zum gleichen Zeitpunkt indexiert und es gibt keine Vorteile für Seiten, die vielleicht früher indexiert werden. Wenn die Seiten indexiert sind, bekommen sie etwa zwei Wochen Zeit, um stabile Rankings aufzubauen. Schon ab diesem Zeitpunkt werden die Rankings permanent beobachtet. Wenn es dann keine großen Veränderungen in den Rankings mehr gibt und sie also stabil sind, wird der Duplicate Content eingebaut. Dies geschieht bei allen betroffenen Seiten eines Keywords am gleichen Tag. Danach wird die Entwicklung des Rankings weiterhin beobachtet. Während des gesamten Testes werden die Google Webmaster Tools zur Hilfe genommen um beispielsweise zu prüfen, ob die Seiten schon indexiert sind und wie oft sie gecrawlt werden.

4.5.1 Die Domains vom Keyword „Elbwelpe“

Im ersten Schritt wurden die Seiten zum Keyword „Elbwelpe“ online gestellt. Am 16.06.2013 zunächst hund-fritz-und-ich.de, schwimm-mit-mir.de, fraulein-chuckles-geschichten.de, barkasse-im-hafen.de, weil-wir-kinder-sind.de, hundetatzentanz.de und

campus-zeitschrift.de. Einen Tag später am 17.06.2013 folgten dann meine-bands.de, hund-vom-fluss.de und mode-in-meinen-auge.de.

Um die Indexierung der Seiten voranzutreiben, wurde am 15.06.2013 der Blogspot-Blog fraulein-chuckle.blogspot.de erstellt. Hier wurden am 19.06.2013 alle Domains von der gleichen URL verlinkt. So neutralisiert sich der Rankingfaktor Backlinks, da auf alle URLs die gleiche Anzahl und qualitativ gleichwertige Backlinks zeigen. Doch der Blog war ab diesem Tag nicht mehr im Index von Google zu finden, somit dürften die Links keine großen Auswirkungen auf den Indexierungsstatus und das Ranking haben. Die Seiten wurden dennoch stetig indexiert. Sie gelangten über eine vom SEO-Plugin in Wordpress automatisch generierte Sitemap in den Index. Die ersten Seiten waren bereits am Tag der Online-Stellung indexiert. Am 22.06.2013 wurden die Seiten untereinander verlinkt, um auf die letzten nicht indexierten Seiten aufmerksam zu machen. Am 23.06.2013 wurden für alle Domains schließlich die Google Webmaster Tools eingerichtet. Bis zu diesem Zeitpunkt noch nicht indexierte Seiten, wurden über die Google Webmaster Tools am gleichen Tag an den Index gesendet. Dies waren:

- <http://hund-vom-fluss.de/elbwelpe-in-der-familie-halten/>
- <http://mode-in-meinen-auge.de/elbwelpe-grunder-max-muller/>
- <http://mode-in-meinen-auge.de/elbwelpe-designer-auf-der-suche-nach-inspirationen/>
- <http://mode-in-meinen-auge.de/impressum/> (am 25.06.2013)
- <http://hund-fritz-und-ich.de/der-elbwelpe-und-die-zeitung/>
- <http://barkasse-im-hafen.de/elbwelpe-auf-grosser-fahrt/>

Außerdem befanden sich zwei 404-Seiten im Index von Google. Diese wurden mit einem Antrag über die Google Webmaster Tools entfernt. Dabei handelte es sich um folgende Seiten:

- <http://mode-in-meinen-auge.de/elbwelpe02/> (Entfernung am 28.06.2013)
- <http://hund-fritz-und-ich.de/erste-probleme/> (Entfernung am 25.06.2013)

Somit waren schließlich am 29.06.2013 alle URLs der Domains für das Keyword „Elbwelpe“ fehlerfrei indexiert, also 12 Tage nachdem die letzten Seiten online gestellt wurden. Die Webmaster-Tools wurden erst eine Woche nach dem Launch der Seiten eingesetzt und zu diesem Zeitpunkt waren bereits fast alle URLs nur über die Wordpress-

Sitemap und interne Verlinkung indexiert. Da durch den Einsatz der Webmaster-Tools die restlichen Seiten aber innerhalb eines Tages indexiert wurden (abgesehen von den 404Seiten), wurde entschieden, dass die restlichen Seiten sofort über die Google Webmaster-Tools an den Index gesendet werden sollen. Durch dieses Vorgehen kommt es zu einer deutlichen zeitlichen Ersparnis. Bei den Seiten von „Elbwelpé“ konnte aber gezeigt werden, wie schnell die Seiten auch ohne Backlinkaufbau nur über die Wordpress-Sitemap indexiert wurden. Der Indexierungsstatus im zeitlichen Verlauf wird noch einmal in Tabelle 02 dargestellt. Hier wurde für jede Domain an jedem Tag nach dem Launch eingetragen, wie viele Seiten indexiert sind. Hierbei sind 4 Seiten der gewünschte Maximalwert, da das Impressum mitgezählt wird.

Domain	Tage nach Launch											
	1	2	3	4	5	6	7*	8	9	10	11	12
hund-vom-fluss.de	0	3	4	3	3	3	4	4	4	4	4	4
mode-in-meinen- augen.de	1	1	0	1	1	3	3	4 + 404	4 + 404	4 + 404	4 + 404	4
hund-fitz-und- ich.de	404	1+2 404	3 + 404	3 + 404	3	3 + 404	4 + 404	4 + 404	4	4	4	4
schwimm-mit- mir.de	0	1	2	4	4	4	4	4	4	4	4	4
fraulein-chuckles- geschichten.de	0	2	2	4	4	4	4	4	4	4	4	4
barkasse-im- hafen.de	1	2	3	2	2	2	4	4	4	4	4	4
weil-wir-kinder- sind.de	0	1	1	2	2	2	4	4	4	4	4	4
hundetatzentanz .de	3	4	4	4	4	4	4	4	4	4	4	4
campus- zeitschrift.de	2	4	4	4	4	4	4	4	4	4	4	4
meine-bands.de	0	2	4	4	4	4	4	4	4	4	4	4

Tabelle 02: Indexierungsstatus der Domains beim Keyword „Elbwelpé“ im zeitlichen Verlauf

(*an Tag 7 wurden die Google Webmaster Tools installiert)

Etwa bis zum 26.06.2013 gab es täglich große Schwankungen im Ranking. Bis zum 29.06.2013 veränderte es sich nur noch wenig und ab diesem Zeitpunkt blieb das Ranking über die nächsten fünf Tage stabil (siehe Abbildung 14). Somit wurde der Duplicate Content am 04.07.2013 eingebaut und online gestellt. Beim Ranking ist auffällig, dass die URLs einer Domain oft an aufeinanderfolgenden Positionen ranken. Es kommt

nur selten vor, dass eine einzelne URL allein zwischen den URLs anderer Domains steht. Es wird beobachtet, ob sich dies auch bei den anderen Keywords so entwickelt.

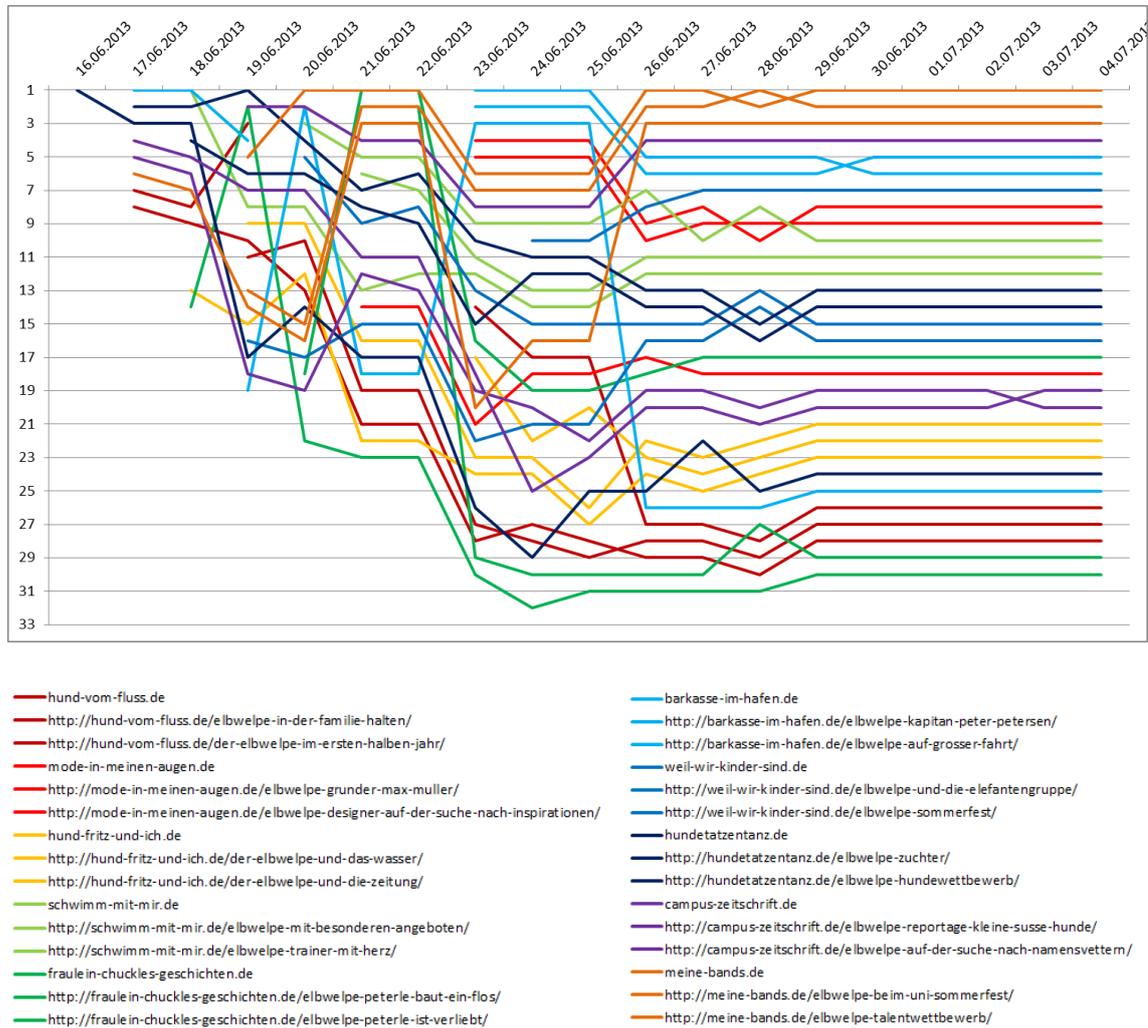


Abbildung 14: Entwicklung der Rankings beim Keyword „Elbwelpe“ von der Onlinestellung bis zum Einbau von Duplicate Content am 04.07.2013

Die Domain fraulein-chuckles-geschichten.de ist beim Keyword „Elbwelpe“ die Vergleichsdomain. Hier wurde also kein Duplicate Content eingebaut. Vom Typ B (Einbau interner Duplicate Content) gibt es bei „Elbwelpe“ vier Domains. So wurden 10 %, 50 %, 70 % und 90 % interner Duplicate Content bei unterschiedlichen Domains eingebaut. Hinzu kommen drei Domains von Typ C mit jeweils 10 %, 30 % und 50 % externem Duplicate Content. Der Content hierfür wurde von zwei Typ-D-Domains geklaut. Bei welchen URLs der Duplicate Content eingebaut wurde und von welchen er geklaut wurde, kann Tabelle 03 entnommen werden.

Content eingebaut

intern		extern	
B-10-1	http://weil-wir-kinder-sind.de/elbwelpe-und-die-elefantengruppe/	C-10-1	http://campus-zeitschrift.de
B-50-1	http://schwimm-mit-mir.de/elbwelpe-trainer-mit-herz/	C-30-2	http://hund-fritz-und-ich.de/der-elbwelpe-und-das-wasser/
B-70-2	http://barkasse-im-hafen.de	C-50-1	http://hund-vom-fluss.de/der-elbwelpe-im-ersten-halben-jahr/
B-902	http://mode-in-meinen-agen.de/elbwelpe-designer-auf-der-suche-nach-inspirationen/		

Content geklaut

intern		extern	
10 %	http://weil-wir-kinder-sind.de	10 %	http://meine-bands.de/elbwelpe-beim-uni-sommerfest/
50 %	http://schwimm-mit-mir.de/elbwelpe-mit-besonderen-angeboten/	30 %	http://hundetatzentanz.de/elbwelpe-hundewettbewerb/
70 %	http://barkasse-im-hafen.de/elbwelpe-auf-grosser-fahrt/	50%	http://hundetatzentanz.de
90 %	http://mode-in-meinen-agen.de		

Tabelle 03: Übersicht der URLs mit verschiedenen Anteilen an internem und externem Duplicate Content beim Keyword „Elbwelpe“

Für den Einbau des Duplicate Contents wurden folgende Varianten gewählt:

- B-10-1: von Startseite auf Unterseite, von oben (<h1> wurde kopiert)
- B-50-1: von Unterseite auf Unterseite, von unten (untere <h2> wurde kopiert)
- B-70-2: von Unterseite auf Startseite, von oben (<h1> und obere <h2> wurden kopiert)
- B-90-2: von Startseite auf Unterseite, von unten (beide <h2> wurden kopiert)
- C-10-1: von Unterseite auf Startseite, von unten (keine <h> wurde kopiert)
- C-30-2: von Unterseite auf Unterseite, von oben (<h1> wurde kopiert)
- C-50-1: von Startseite auf Unterseite, von unten (beide <h2> wurden kopiert)

Ein Beispiel: Von der Startseite der Domain mode-in-meinen-augen.de sollen 90 % Content geklaut und intern auf der URL <http://mode-in-meinen-augen.de/elbwelpedesigner-auf-der-suche-nach-inspirationen/> eingebaut werden. Dazu wird wie folgt vorgegangen:

- Der Text der Startseite enthält mit Überschriften 313 Wörter.
- 90 % von 313 sind 281,7, also 282 Wörter (hier muss immer aufgerundet werden, da sich sonst zu wenig Duplicate Content auf den Seiten befindet).
- Von dem Text der URL, bei dem der Duplicate Content eingebaut werden soll, werden dann 282 entfernt und durch 282 aus dem Text der Startseite ersetzt.

Das Keyword soll dabei gleich oft in dem entfernten und kopierten Teil vorkommen. Die Überschriften werden hier nun mit eingerechnet, es kann also sein, dass der Text auf der neuen Seite dann länger als 300 Wörter ist (z.B. wenn hier vorher lange Überschriften waren und nun kurze eingebaut werden, bleibt mehr Platz für Wörter aus dem Text). Der interne Link soll, wenn möglich, genauso erhalten bleiben. Wenn der Link durch den Einbau des Duplicate Content aber verloren gehen würde, soll er an etwa gleicher Stelle mit ähnlichem Linktext wieder eingebaut werden.

In der folgenden Tabelle 04 sind noch einmal alle auf das Keyword „Elbwelpe“ optimierten Domains mit ihren Typen und Veränderungen dargestellt:

Domain	Typ	Veränderung Content
fraulein-chuckles-geschichten.de	A-1	keine Veränderung
weil-wir-kinder-sind.de	B-10-1	Einbau interner Duplicate Content
schwimm-mit-mir.de	B-50-1	
barkasse-im-hafen.de	B-70-2	
mode-in-meinen-augen.de	B-90-2	
campus-zeitschrift.de	C-10-1-D2	Einbau externer Duplicate Content
hund-fritz-und-ich.de	C-30-2-D1	
hund-vom-fluss.de	C-50-1-D1	
hundetatzentanz.de	D-1	keine Veränderung
meine-bands.de	D-2	

Tabelle 04: Übersicht der Domains mit ihren Typen zum Keyword „Elbwelpe“

Die Entwicklung des Rankings nach dem Einbau des Duplicate Content und die daraus resultierenden Beobachtungen und Ergebnisse werden in Kapitel 4.6 ausführlich erläutert.

4.5.2 Die Domains vom Keyword „Randballsport“

Als nächstes wurden am 24.06.2013 alle auf das Keyword „Randballsport“ optimierten Domains online gestellt. Am folgenden Tag wurden die Google Webmaster Tools für alle Domains eingerichtet und jeweils alle Seiten an den Index gesendet, damit sie schnell indexiert werden. Dies gelang sehr gut und am 01.07.2013 waren schließlich alle Seiten fehlerfrei indexiert. Bei den Domains mein-mannschaftssport.de und sport-denken.de wurde das Impressum jeweils am 01.07.2013 noch einmal über die Google Webmaster Tools an den Index gesendet, da dieses bis dahin noch nicht indexiert war. Der Verlauf des Indexierungsstatus kann der folgenden Tabelle 05 entnommen werden.

Domain	Tage nach Launch							
	1*	2	3	4	5	6	7	8
mein-mannschaftssport.de	4	3	3	3	3	3	4**	4**
sieben-meter.de	0	0	0	3	3	4	4**	4**
student-macht-sport.de	4	4	4	4	4	4	4**	4**
mannschaftssport.de	4	4	4	4	4	4	4**	4**
ballsport-verein.de	4	4	4	4	4	4	4**	4**
weil-wir-sportler-sind.de	4	3	3	3	3	4	4**	4**
handball-in-den-medien.de	4	4	4	4	4	4	4**	4**
sport-denken.de	0	0	3	3	3	3	4**	4**
schueler-machen-sport.de	3	3	4	4	4	4	4**	4**
internationaler-mannschaftssport.de	4	4	4	4	4	4	4	4**

Tabelle 05: Indexierungsstatus der Domains beim Keyword „Randballsport“ im zeitlichen Verlauf (*an Tag 1 wurden die Google Webmaster Tools installiert, **an diesem Tag tauchten die Author-Seiten auf)

Am 01.07.2013 tauchten bei den Domains des Keywords „Randballsport“ URLs nach folgendem Schema im Index auf: domain.de/author/admin. Diese wurden automatisch von Wordpress erstellt und gelangten über eine vom WordPress SEO-Plugin generierte Author-/User-Sitemap in den Index. Nach der Entdeckung wurde diese automatische Sitemap bei allen Domains von „Randballsport“ deaktiviert. Bei den Domains von „Elbwelpe“ waren diese Seiten nicht aufgetaucht, weil hier noch eine ältere Version des SEO-Plugins verwendet wurde, bei dem es diese Funktion noch nicht gab (Version

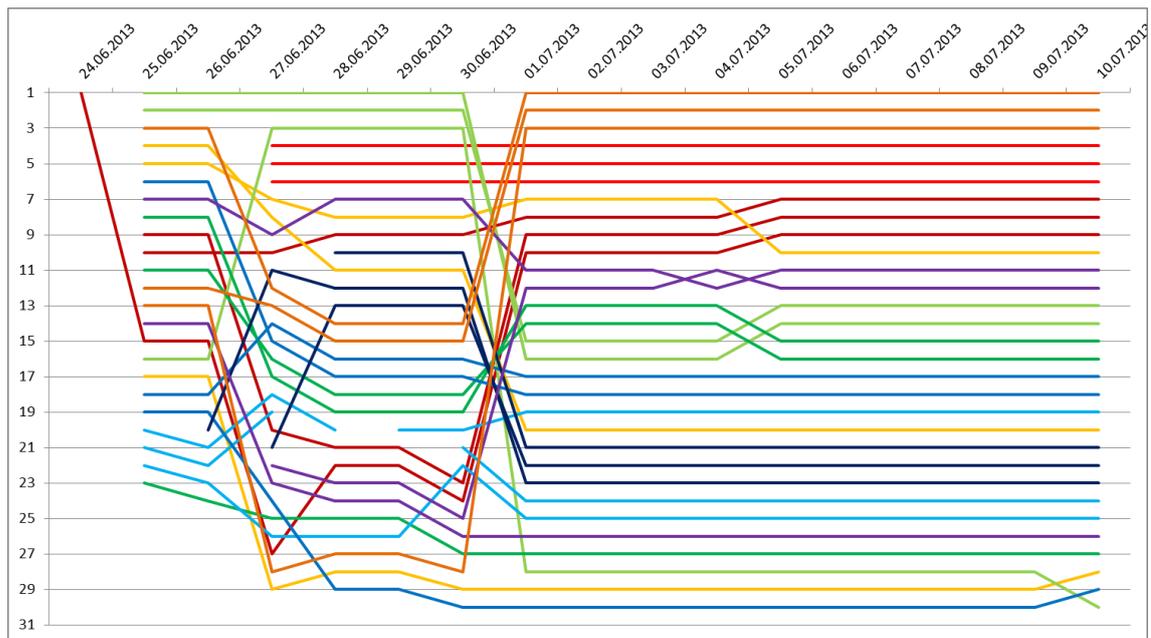
1.4.7). Bei den „Randballsport“-Seiten wurden Version 1.4.10 verwendet. Auch nach der Deaktivierung der Funktion blieben die Seiten trotzdem im Index und wurden bei einer Site-Abfrage der Domains weiterhin angezeigt. Da sich aber kein Content auf den Seiten befindet (siehe Abbildung 15) und sie also nicht auf das jeweilige Keyword ranken, wurde entschieden, dass sie nicht aus dem Index entfernt werden müssen. Sie beeinflussen den Test nicht und können also unbeachtet bleiben.

MEIN MANNSCHAFTSSPORT - C-90-1-D4				
ALLGEMEINE INFORMATIONEN	DIE GRUNDREGELN	WEITERE REGELN	IMPRESSUM	
Allgemeine Informationen	Die Grundregeln	Weitere Regeln	Impressum	
THEME: SIMPLE STYLE BY FIMPLY				

Abbildung 15: Screenshot der Seite <http://mein-mannschaftssport.de/author/admin/> (aufgenommen am 01.07.2013)

Die interne Verlinkung wurde für die Domains mein-mannschaftssport.de, studentmacht-sport.de, mannschaftssportturnier.de, ballsport-verein.de und weil-wir-sportler-sind.de am 30.06.2013 gesetzt, für die restlichen Domains am 01.07.2013. Hier wurde die interne Verlinkung nicht dazu gebraucht, damit die restlichen Seiten indexiert werden, wie bei „Elbwelpe“, sondern damit sich nochmal etwas an den Seiten verändert und der Crawler weiterhin oft die Seiten besucht.

Da die Seiten alle sehr schnell indexiert waren und rankten, entwickelte sich das Ranking schnell konstant. Nach dem Einbau der internen Verlinkung am 30.06./01.07.2013 wurde das Ranking noch einmal durcheinander geworfen (siehe Abbildung 16). Dies zeigt, welche Auswirkungen die internen Links auf das Ranking haben. Danach blieb das Ranking bis auf kleine Veränderungen konstant. Wie bei den Seiten von „Elbwelpe“ auch, ranken hier meist alle drei URLs einer Domain an aufeinanderfolgenden Positionen. Am 10.07.2013, also zwei Wochen und zwei Tage nachdem die Seiten online gingen, wurde schließlich der Duplicate Content eingebaut.



- mein-mannschaftssport.de
- http://mein-mannschaftssport.de/randballsport-allgemeine-grundregeln/
- http://mein-mannschaftssport.de/randballsport-weitere-regeln/
- sieben-meter.de
- http://sieben-meter.de/randballsport-kindergarten/
- http://sieben-meter.de/randballsport-herrenmannschaft-wird-kreismeister/
- student-macht-sport.de
- http://student-macht-sport.de/das-randballsport-team-hsd-trainiert/
- http://student-macht-sport.de/das-randballsport-turnier-startet/
- mannschaftssportturnier.de
- http://mannschaftssportturnier.de/randballsport-turnier-die-spieleweise/
- http://mannschaftssportturnier.de/randballsport-turnier-an-und-abreise/
- ballsport-verein.de
- http://ballsport-verein.de/die-randballsport-mannschaften/
- http://ballsport-verein.de/die-randballsport-spieberichte-der-letzten-saison/
- weil-wir-sportler-sind.de
- http://weil-wir-sportler-sind.de/randballsport-deckungsvarianten/
- http://weil-wir-sportler-sind.de/randballsport-angriffsvarianten/
- handball-in-den-medien.de
- http://handball-in-den-medien.de/randballsport-mehr-verbände-bitte/
- http://handball-in-den-medien.de/randballsport-mehr-aufmerksamkeit-bitte/
- sport-denken.de
- http://sport-denken.de/randballsport-ost-der-neue-bezirksmeister/
- http://sport-denken.de/die-randballsport-bezirksliga-absteiger/
- schueler-machen-sport.de
- http://schueler-machen-sport.de/die-randballsport-ag-beim-schulturnier/
- http://schueler-machen-sport.de/randballsport-im-unterricht/
- internationaler-mannschaftssport.de
- http://internationaler-mannschaftssport.de/randballsport-turnier-international/
- http://internationaler-mannschaftssport.de/randballsport-und-fair-play/

Abbildung 16: Entwicklung der Rankings beim Keyword „Randsport“ von der Onlinestellung bis zum Einbau von Duplicate Content am 10.07.2013

Von den Domains beim Keyword „Randsport“ dient sieben-meter.de als Vergleichsdomain. Hier wurde kein Duplicate Content eingebaut. Es gibt drei Domains von Typ B, bei denen interner Duplicate Content eingebaut wird und zwar jeweils einmal 10 %, 30 % und 70 %. Hinzu kommen vier Domains vom Typ C mit Einbau von externem Duplicate Content. Hier wurden die Anteile 30 %, 50 %, 70 % und 90 % verwendet. Der Content wurde hierfür von zwei Typ-D-Domains geklaut. Tabelle 06 zeigt, bei welchen URLs Duplicate Content eingebaut wurde und von welchen er geklaut wurde.

Content eingebaut

intern		extern	
B-10-2	http://sport-denken.de/die-randballsport-bezirksliga-absteiger/	C-30-1	http://student-macht-sport.de/das-randballsport-turnier-startet/
B-30-2	http://handball-in-den-medien.de/	C-50-2	http://schueler-machen-sport.de/randballsport-im-unterricht/
B-70-1	http://internationaler-mannschaftssport.de/	C-70-2	http://mannschaftssportturnier.de/
		C-90-1	http://mein-mannschaftssport.de/randballsport-weitere-regeln/

Content geklaut

intern		extern	
10 %	http://sport-denken.de/	30 %	http://ballsport-verein.de/die-randballsport-spielberichte-der-letzten-saison/
30 %	http://handball-in-den-medien.de/randballsport-mehr-verbände-bitte/	50 %	http://weil-wir-sportler-sind.de/
70 %	http://internationaler-mannschaftssport.de/randballsport-turnier-international/	70 %	http://ballsport-verein.de/
		90 %	http://weil-wir-sportler-sind.de/randballsport-deckungsvarianten/

Tabelle 06: Übersicht der URLs mit verschiedenen Anteilen an internem und externem Duplicate Content beim Keyword „Randballsport“

Für den Einbau des Duplicate Content beim Keyword „Randballsport“ wurden folgende Varianten gewählt:

- B-10-2: von Startseite auf Unterseite, von oben (<h1> wurde kopiert)
- B-30-2: von Unterseite auf Unterseite, von unten (untere <h2> wurde kopiert)
- B-70-1: von Unterseite auf Startseite, von unten (beide <h2> wurden kopiert)
- C-30-1: von Unterseite auf Unterseite, von oben (<h1> wurde kopiert)
- C-50-2: von Startseite auf Unterseite, von oben (<h1> und obere <h2> wurden kopiert)
- C-70-2: von Startseite auf Startseite, von unten (beide <h2> wurden kopiert)
- C-90-1: von Unterseite auf Unterseite, von unten (beide <h2> wurden kopiert)

Der folgenden Tabelle 07 können noch einmal alle auf das Keyword „Randballsport“ optimierten Domains mit ihren Typen und Veränderungen in der Übersicht entnommen werden:

Domain	Typ	Veränderung Content
sieben-meter.de	A-2	keine Veränderung
sport-denken.de	B-10-2	Einbau interner Duplicate Content
handball-in-den-medien.de	B-30-2	
internationaler-mannschaftssport.de	B-70-1	
student-macht-sport.de	C-30-1-D3	Einbau externer Duplicate Content
schueler-machen-sport.de	C-50-2-D4	
mannschaftssportturnier.de	C-70-2-D3	
mein-mannschaftssport.de	C-90-1-D4	
ballsport-verein.de	D-3	keine Veränderung
weil-wir-sportler-sind.de	D-4	

Tabelle 07: Übersicht der Domains mit ihren Typen zum Keyword „Randballsport“

4.5.3 Die Domains vom Keyword „Jahreszeitentechnologie“

Am 29.06.2013 wurden schließlich die Domains vom Keyword „Jahreszeitentechnologie“ online gestellt. Auch hier wurden am folgenden Tag die Google Webmaster Tools eingerichtet und alle URLs an den Index gesendet. Die Indexierung erfolgte schnell. Es gab nur ein Problem mit einer URL der Domain jahres-magazin.de. Hier lag in der URL <http://jahres-magazin.de/jahreszeitentechnologie-aktiv-ser-neue-sleeping-bag-20> ein Tippfehler vor. Also wurde die URL über die Google Webmaster Tools aus dem Index entfernt, in Wordpress korrigiert und erneut an den Index gesendet. Es dauert dann noch bis zum 06.07.2013 bis diese wieder indexiert war und alle URLs fehlerfrei im Index waren. Die Indexierung kann im Verlauf der folgenden Tabelle 08 entnommen werden.

Domain	Tage nach Launch						
	1*	2	3	4	5	6	7
jahreszeitenkraftwerk.de	4	4	4	4	4	4**	4**
mein-voller-kleiderschrank.de	4	4	4	4	4	4	4**
draussen-schlafen.de	4	4	4	4	4	4**	4**
wetterbeeinflussung.de	4	4	4	4	4	4	4**
temperatur-im-auto.de	4	4	4	4	4	4	4**
technik-im-auto.de	4	4	4	4	4	3**	4**
jahres-magazin.de	3	3	3	3	3	3**	4**

terrarium-technik.de	4	4	4	4	4	4	4**
jahresbauer.de	4	4	4	4	4	4	4**
handballspray.de	4	4	4	4	4	4**	4**

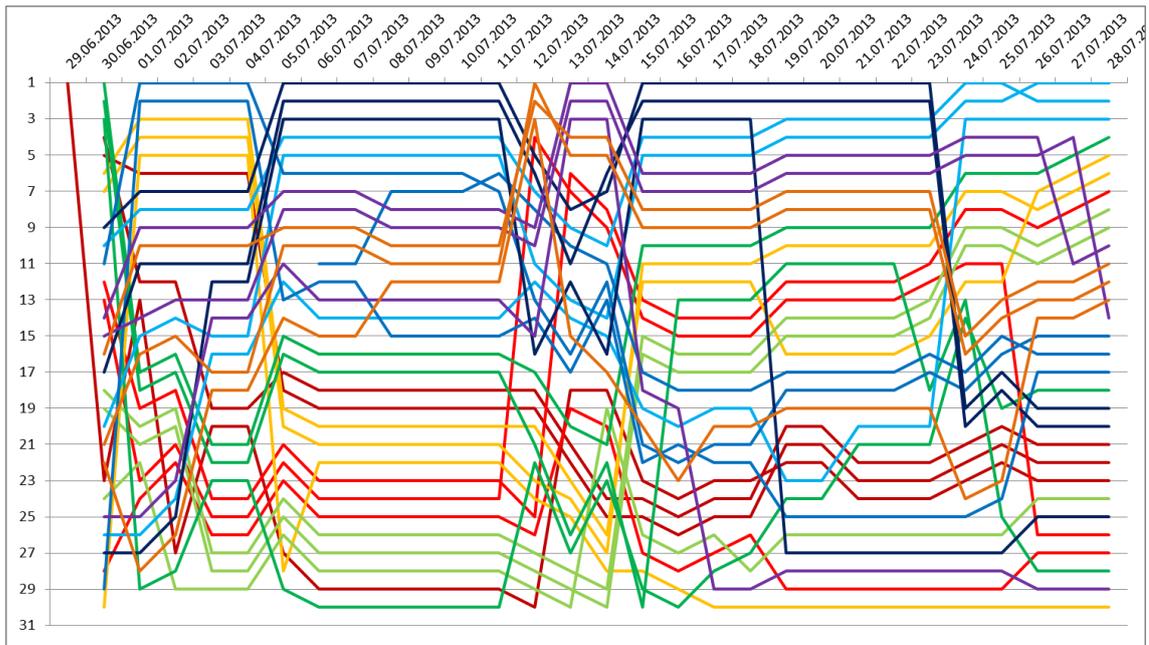
Tabelle 08: Indexierungsstatus der Domains beim Keyword „Jahreszeitentechnologie“ im zeitlichen Verlauf (*an Tag 1 wurden die Google Webmaster Tools installiert, **an diesem Tag tauchten die Author-Seiten auf)

Auch bei den Seiten von „Jahreszeitentechnologie“ wurden über das WordPress SEO-Plugin eine Author-/User-Sitemap generiert und die entsprechenden Seiten automatisch indexiert. Die Funktion wurde wie bei den Domains von „Randballsport“ deaktiviert. Die Seiten wurden aber nicht aus dem Index entfernt, da sie auch hier keinen Content aufweisen und den Test also nicht beeinflussen.

Die internen Links wurden hier alle am 07.07.2013 gesetzt. Dies diente dazu, um den Crawler weiterhin dazu zu animieren, die Seiten zu besuchen, damit der Duplicate Content dann ebenfalls schnell gefunden und darauf reagiert wird.

Die Seiten waren bis auf die eine zunächst fehlerhafte URL alle schnell indexiert und rankten. Es gab hier aber immer wieder großes Durcheinander (siehe Abbildung 17). Auf die intern gesetzten Links wurde scheinbar erst einige Tage später am 11.07.2013 reagiert. Danach beruhigte sich das Ranking aber nicht und blieb weiterhin instabil. Es gab immer wieder größere Veränderungen und Neupositionierungen von mehreren Seiten. Obwohl die Rankings nicht so stabil waren wie bei den Domains von „Elbwelpe“ und „Randballsport“, musste der Duplicate Content schließlich am 28.07.2013 online gestellt werden. Dieser Zeitpunkt wurde ausgewählt, damit die Entwicklung der Seiten noch über einen Zeitraum von drei Wochen beobachtet werden konnten. Wäre der Duplicate Content erst später eingebaut worden, wäre der Beobachtungszeitraum bis zur Abgabe der Bachelorarbeit zu knapp geworden.

Interessant ist hier, dass über einen langen Zeitraum die URLs von genau fünf Domains die Plätze 1 bis 15 belegten und die URLs der anderen fünf Domains die Positionen 16 bis 30. Es gab hier lange Zeit keine Vermischung. Ebenso ist auch hier die Gruppierung der URLs nach Domains wie bei den anderen Keywords auch zu beobachten.



- | | |
|---|---|
| — jahreszeitenkraftwerk.de | — technik-im-auto.de |
| — http://jahreszeitenkraftwerk.de/jahreszeitentechnologie-im-herbst/ | — http://technik-im-auto.de/die-jahreszeitentechnologie-in-der-verwendung/ |
| — http://jahreszeitenkraftwerk.de/jahreszeitentechnologie-im-fruehling/ | — http://technik-im-auto.de/jahreszeitentechnologie-expertenmeinung/ |
| — mein-voller-kleiderschrank.de | — jahres-magazin.de |
| — http://mein-voller-kleiderschrank.de/jahreszeitentechnologie-die-erfinder/ | — http://jahres-magazin.de/jahreszeitentechnologie-unterwegs-zu-besuch-im-jahreszeitenkraftwerk/ |
| — http://mein-voller-kleiderschrank.de/jahreszeitentechnologie-die-vorteile/ | — http://jahres-magazin.de/jahreszeitentechnologie-aktiv-der-neue-sleeping-bag-20/ |
| — draussen-schlafen.de | — terrarium-technik.de |
| — http://draussen-schlafen.de/jahreszeitentechnologie-im-test/ | — http://terrarium-technik.de/jahreszeitentechnologie-einbau-und-moeglichkeiten/ |
| — http://draussen-schlafen.de/jahreszeitentechnologie-im-schlafsack-sehr-zu-empfehlen/ | — http://terrarium-technik.de/jahreszeitentechnologie-fur-schildkroten/ |
| — wetterbeeinflussung.de | — jahresbauer.de |
| — http://wetterbeeinflussung.de/jahreszeitentechnologie-im-einsatz/ | — http://jahresbauer.de/jahreszeitentechnologie-und-die-vorteile/ |
| — http://wetterbeeinflussung.de/jahreszeitentechnologie-mit-problemen/ | — http://jahresbauer.de/jahreszeitentechnologie-bei-biobauer-hans/ |
| — temperatur-im-auto.de | — handballspray.de |
| — http://temperatur-im-auto.de/jahreszeitentechnologie-nicht-nur-fur-autos/ | — http://handballspray.de/jahreszeitentechnologie-im-gebrauch/ |
| — http://temperatur-im-auto.de/jahreszeitentechnologie-im-test/ | — http://handballspray.de/jahreszeitentechnologie-bei-den-stars/ |

Abbildung 17: Entwicklung der Rankings beim Keyword „Jahreszeitentechnologie“ von der Onlinestellung bis zum Einbau von Duplicate Content am 28.07.2013

Von den Domains beim Keyword „Jahreszeitentechnologie“ wurde handballspray.de als Vergleichsdomain gewählt. Außerdem gibt es drei Domains von Typ B mit 30 %, 50 % und 90 % internem Duplicate Content. Vom Typ C kommen vier Domains hinzu und zwar mit den Anteilen 10 %, 70 %, 90 % und 100 %, also einer komplett kopierten Domain. Entsprechend gibt es noch zwei Domains vom Typ D, von denen der Content geklaut wird. In Tabelle 09 wird gezeigt, von welchen URLs Content geklaut und bei welchen er eingebaut wurde.

Content eingebaut

intern		extern	
B-30-1	http://wetterbeeinflussung.de/	C-10-2	http://terrarium-technik.de/
B-50-2	http://mein-voller-kleiderschrank.de/jahreszeitentechnologie-die-erfinder/	C-70-1	http://jahreszeitenkraftwerk.de/
B-90-1	http://jahresbauer.de/jahreszeitentechnologie-und-die-vorteile/	C-90-2	http://draussen-schlafen.de/jahreszeitentechnologie-im-test/

Content geklaut

intern		extern	
30 %	http://wetterbeeinflussung.de/jahreszeitentechnologie-mit-problemen/	10 %	http://technik-im-auto.de/jahreszeitentechnologie-expertenmeinung/
50 %	http://mein-voller-kleiderschrank.de/jahreszeitentechnologie-die-vorteile/	70 %	http://jahres-magazin.de/
90 %	http://jahresbauer.de/	90 %	http://jahres-magazin.de/jahreszeitentechnologie-aktiv-der-neue-sleeping-bag-20/

100 % kopiert

Content geklaut	Content eingebaut
http://technik-im-auto.de/	http://temperatur-im-auto.de/
http://technik-im-auto.de/die-jahreszeitentechnologie-in-der-verwendung/	http://temperatur-im-auto.de/jahreszeitentechnologie-nicht-nur-fur-autos/
http://technik-im-auto.de/jahreszeitentechnologie-expertenmeinung/	http://temperatur-im-auto.de/jahreszeitentechnologie-im-test/

Tabelle 09: Übersicht der URLs mit verschiedenen Anteilen an internem und externem Duplicate Content beim Keyword „Jahreszeitentechnologie“

Folgende Varianten wurden beim Einbau des Duplicate Contents beim Keyword „Jahreszeitentechnologie“ gewählt:

- B-30-1: von Unterseite auf Startseite, von unten (untere <h2> wurde kopiert)
- B-50-2: von Unterseite auf Unterseite, von oben (<h1> und obere <h2> wurde kopiert)
- B-90-1: von Startseite auf Unterseite, von oben (alle Überschriften wurden kopiert)

- C-10-2: von Unterseite auf Startseite, von oben (<h1> wurde kopiert)
- C-70-1: von Startseite auf Startseite, von unten (beide <h2> wurden kopiert)
- C-90-2: von Unterseite auf Unterseite, von oben (alle Überschriften wurden kopiert)
- C-100-1: alle Seiten mit allen Überschriften wurden kopiert

Tabelle 10 zeigt noch einmal alle Domains, die auf das Keyword „Jahreszeitentechnologie“ optimiert wurden, mit den entsprechenden Typen und Veränderungen.

Domain	Typ	Veränderung Content
handballspray.de	A-3	keine Veränderung
wetterbeeinflussung.de	B-30-1	Einbau interner Duplicate Content
mein-voller-kleiderschrank.de	B-50-2	
jahresbauer.de	B-90-1	
terrarium-technik.de	C-10-2-D5	Einbau externer Duplicate Content
jahreszeitenkraftwerk.de	C-70-1-D6	
draussen-schlafen.de	C-90-2-D6	
temperatur-im-auto.de	C-100-1-D5	
technik-im-auto.de	D-5	keine Veränderung
jahres-magazin.de	D-6	

Tabelle 10: Übersicht der Domains mit ihren Typen zum Keyword „Jahreszeitentechnologie“

4.6 Beobachtung und Auswertung der Rankingentwicklung

Der Beobachtungszeitraum der URLs nach dem Einbau des Duplicate Contents beträgt 21 Tage. Die lückenlose Beobachtung erfolgt durch ein tägliches Abfragen der Rankings. Außerdem wurde ständig geprüft, ob noch alle Seiten indexiert sind und ob sie noch regelmäßig gecrawlt werden. Die Beobachtungen und Ergebnisse werden im Folgenden dargestellt und erläutert. Dabei werden zunächst allgemeine Entwicklungen gezeigt und schließlich jeweils die Entwicklung der URLs der einzelnen Typen genau erklärt. Danach folgt die Prüfung der zuvor aufgestellten Hypothesen.

Alle Tabellen und Diagramme mit Rankingentwicklungen, Screenshots der Rankings sowie alle Texte der URLs sind auf der dieser Bachelorarbeit beigefügten CD enthalten.

Allgemeine Entwicklung

Zunächst wird ein Blick auf die allgemeine Entwicklung der Rankings nach dem Einbau vom Duplicate Content geworfen. Am Beispiel der URLs vom Keyword „Elbwelpe“ in Abbildung 18 zeigt sich deutlich, dass der Duplicate Content durchaus Auswirkungen auf das Ranking hat und sich die Rankings der URLs zum Teil um sehr viele Positionen verändert haben. Dabei gibt es aber auch immer wieder Phasen, in denen es wenig Veränderungen gab und die Rankings nahezu stabil blieben.

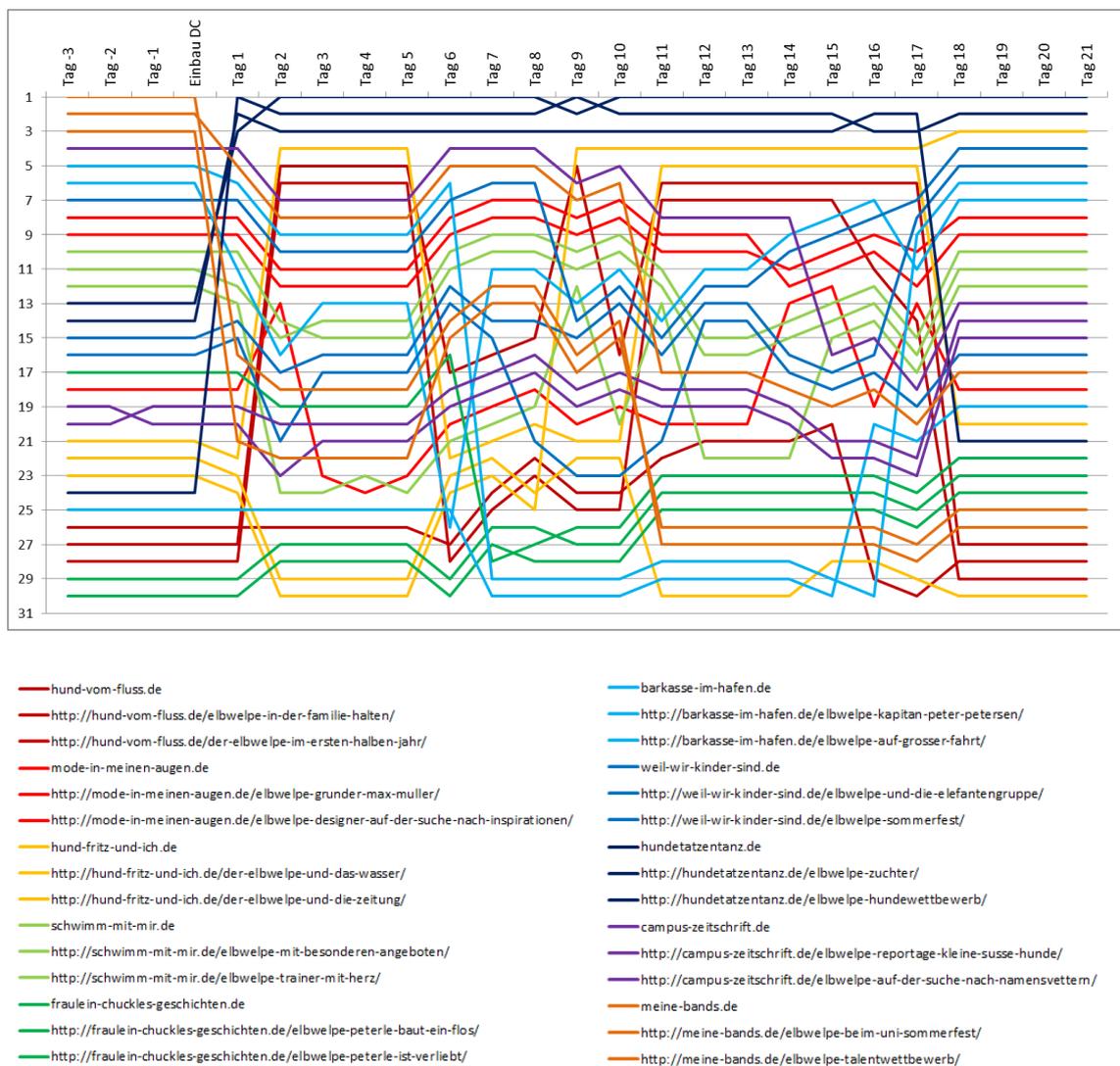


Abbildung 18: Entwicklung der Rankings im Zeitraum von 21 Tagen nach dem Einbau des Duplicate Contents beim Keyword „Elbwelpe“

Bei den anderen beiden Keywords zeichnet sich ein ähnliches Bild. Die entsprechenden Diagramme können im Anhang 05 und 06 dieser Arbeit eingesehen werden.

Somit ist die erste Hypothese, dass Duplicate Content in Bezug auf das Ranking in der Suchmaschine Google eine Rolle spielt, bestätigt.

Dies zeigt sich auch in folgender Abbildung 19. Hier ist dargestellt, wie viele URLs bei denen Content eingebaut oder von denen Content geklaut wurde, nach Einbau des Duplicate Contents jeweils im Vergleich zum Vortag ihre Position verändert haben. Es geht also noch nicht um die Höhe der Veränderung, sondern lediglich darum, ob sie sich verändert haben. Dabei spielt es keine Rolle, ob sie im Ranking gestiegen oder gefallen sind, sondern es geht nur darum, dass sie sich verändert haben. Betrachtet werden die URLs, bei denen interner Duplicate Content eingebaut oder geklaut wurde, externer Duplicate Content eingebaut oder geklaut wurde und die URLs der Domain, die komplett kopiert wurde. Insgesamt sind dies 46 URLs.

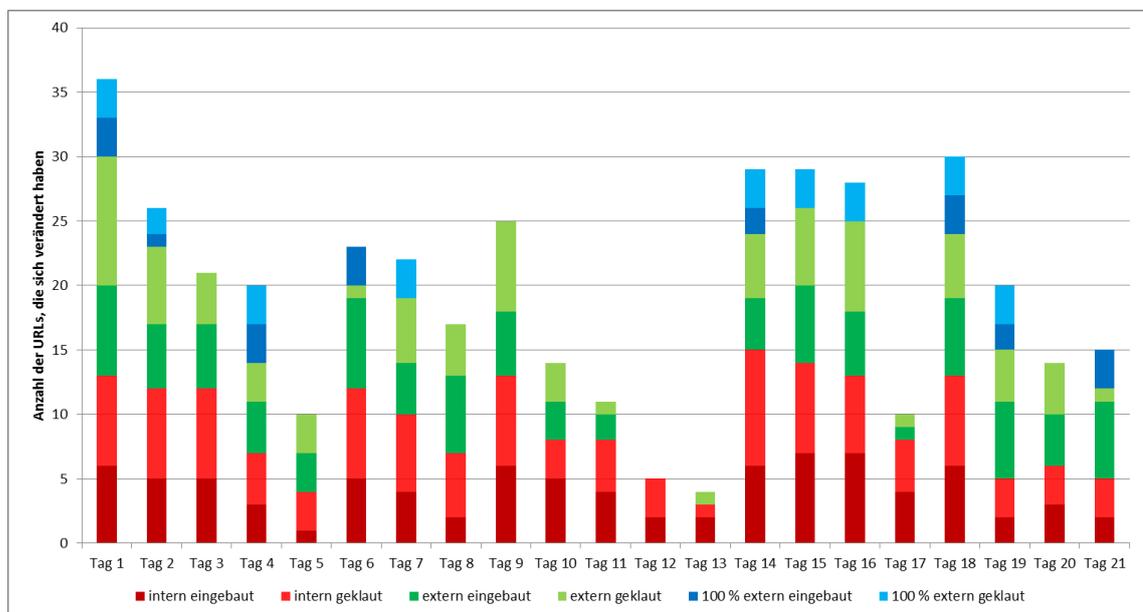


Abbildung 19: Anzahlen der URLs, die jeweils im Vergleich zum Vortag ihre Position verändert haben

Von allen URLs, bei denen Duplicate Content eingebaut oder geklaut wurde (46) haben pro Tag im Durchschnitt 19,4 URLs ihre Position im Ranking verändert. Das sind etwa 42 % der URLs, bei denen etwas verändert wurde.

Von den URLs, bei denen interner Duplicate Content eingebaut oder geklaut wurde (insgesamt 20) haben pro Tag im Durchschnitt 9,2 URLs ihr Ranking verändert. Das sind 46 %. Bei allen URLs, bei denen externer Duplicate Content geklaut oder eingebaut wurde (insgesamt 26), haben sich durchschnittlich 10,2 URLs pro Tag im Ranking

verändert. Das sind etwa 39.2 %. Also rund 7 % weniger als bei den URLs vom internen Duplicate Content.

Daraus lässt sich schließen, dass interner Duplicate Content eine größere Auswirkung auf die tägliche Veränderung der URLs im Ranking hat als externer. Somit ist auch Hypothese 3, dass sich interner Duplicate Content grundsätzlich stärker auf das Ranking auswirkt als externer, bestätigt.

Insgesamt haben die meisten URLs direkt an Tag 1 nach dem Einbau vom Duplicate Content ihre Position verändert. 36 URLs, bei denen interner oder externer Duplicate Content eingebaut oder geklaut wurde, veränderten am ersten Tag ihre Position. Das sind 78 % aller URLs, bei denen etwas verändert wurde, also ein sehr großer Anteil. In den meisten Fällen hat Google also sofort auf die Veränderungen reagiert. Über den Zeitraum von 21 Tagen gesehen haben die wenigsten URLs an Tag 13 ihre Position verändert. Hier waren es insgesamt nur 4 URLs mit einer veränderten Position im Vergleich zum Vortag.

Auch hier zeigt sich, dass es über den gesamten Zeitraum verteilt, immer wieder Phasen gab, in denen viele URLs ihre Position veränderten und Phasen, in denen insgesamt wenig URLs ihre Position veränderten. Diese Phasen stimmen bei externem und internem Duplicate Content nahezu überein. Meist folgt auf eine Phase mit wenigen URLs, die sich verändert haben, sofort eine Phase mit vielen URLs, die ihre Position im Vergleich zum Vortag verändert haben.

Zwischenfazit allgemeine Entwicklung

Der Einbau des Duplicate Contents zeigt deutliche Auswirkungen auf das Ranking. Google hat in den meisten Fällen sofort auf den Einbau reagiert und die URLs veränderten ihre Position. Es gab aber immer wieder auch Phasen, in denen sich das Ranking nicht oder nur kaum verändert hat. Danach gab es dann wieder große Veränderungen, sowohl positiv als auch negativ. Es scheint so, als ob Google die Seiten jeweils von Zeit zu Zeit neu bewertet und dann neu eingestuft hat.

Nach diesen allgemeinen Informationen über die grundsätzliche Veränderung der Rankings und den Einfluss des Duplicate Content, soll nun ein genauer Blick auf die Entwicklung der Seiten der einzelnen Typen A bis D geworfen werden. Dazu wurde zum einen jeweils die Durchschnittsposition vor und nach dem Einbau des Duplicate Contents berechnet und verglichen. Für die Durchschnittsposition nach dem Einbau des Duplicate Contents wurden die Positionen jeder URL addiert und durch die 21 Tage des Beobachtungszeitraums geteilt. Für die Durchschnittsposition vor dem Einbau des Duplicate Contents wurden jeweils die Positionen der letzten fünf Tage addiert und entsprechend durch fünf geteilt. Zum anderen wurden Diagramme erstellt, die die jeweilige konkrete Rankingentwicklung zeigen. Hier wurden außerdem lineare Trendlinien eingefügt, um leicht ablesen zu können, ob es eher eine positive oder negative Entwicklung ist.

Typ B: Duplicate Content intern eingebaut

Bei den Seiten, bei denen interner Duplicate Content eingebaut wurde, wird vermutet, dass die Rankingentwicklung ab einem bestimmten Anteil negativ ist. Es wird also geschaut, ob die Differenz der Durchschnittspositionen vor und nach dem Einbau von Duplicate Content negativ ist. Außerdem wird angenommen, dass die Rankingentwicklung, wenn der Duplicate Content von Google erkannt und bestraft wurde, einen ähnlichen Verlauf zeigt und die Seiten also nach dem gleichen Muster behandelt wurden. Dies wird ebenfalls bei den Seiten überprüft.

Abbildung 20 zeigt die Durchschnittspositionen der URLs, bei denen die jeweiligen Stufen Duplicate Content intern eingebaut wurden:

interner DC	vor Einbau DC	nach Einbau DC	Differenz
B-10-1	15	12,7	2,3
B-10-2	22	18,3	3,7
B-30-1	24,8	23,8	1,0
B-30-2	29	24,9	4,1
B-50-1	10	12,7	-2,7
B-50-2	20	17,6	2,4
B-70-1	3	19,9	-16,9
B-70-2	25	25,4	-0,4
B-90-1	6,6	2,8	3,8
B-90-2	9	9,6	-0,6

Abbildung 20: Durchschnittspositionen der Typ-B-URLs

10 %: Beide URLs stiegen im Durchschnitt um etwa den gleichen Wert im Ranking. (2,3 und 3,7 Positionen), bei der Rankingentwicklung in Abbildung 21 zeigen sich aber unterschiedliche Verläufe. Einmal liegt ein positiver Trend vor und einmal ein negativer. Die Behandlung der Seiten ist also nicht einheitlich. Dies lässt darauf schließen, dass sich so ein geringer Anteil an Duplicate Content noch nicht auf das Ranking auswirkt. Vielleicht wurde dieser gar nicht von Google erkannt, zumindest nicht gewertet.

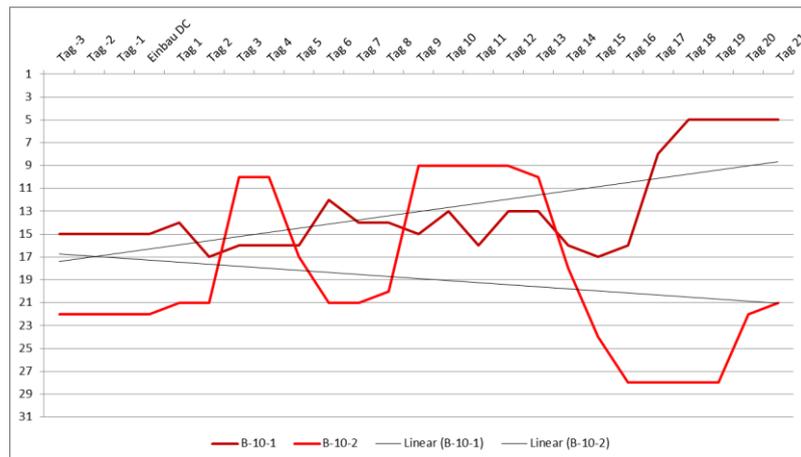


Abbildung 21: Rankingentwicklung der URLs mit 10 % internem Duplicate Content

30 %: Hier gestaltet sich das Bild ähnlich wie bei den 10 %-URLs. Auch hier haben sich beide URLs durchschnittlich im Ranking verbessert. Je weiter unten die URL vorher im Ranking stand, umso mehr hat sie sich verbessert. Dies zeigt auch die Rankingentwicklung der URLs (siehe Abbildung 22). Insgesamt gab es aber kaum Veränderungen, was auch die Durchschnittspositionen zeigen. Auch ein Anteil von 30 % internem Duplicate Content scheint also noch zu gering zu sein, um sich negativ auszuwirken.

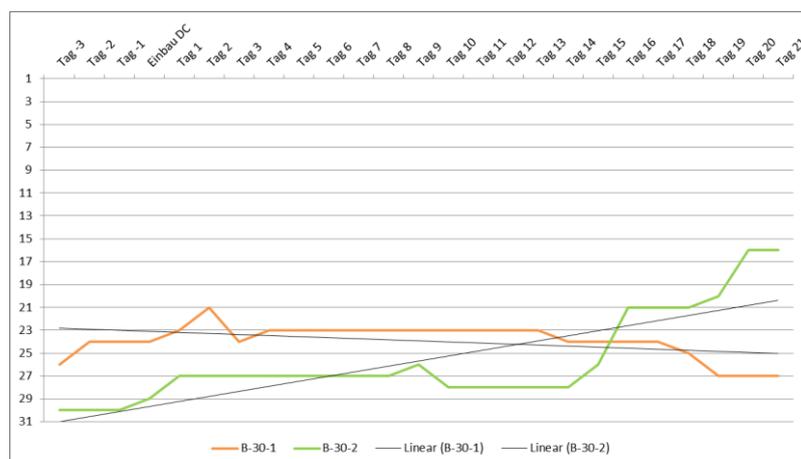


Abbildung 22: Rankingentwicklung der URLs mit 30 % internem Duplicate Content

50 %: Die URLs haben sich hier unterschiedlich entwickelt. Die erste URL (grüne Linie) stand vor dem Einbau vom Duplicate Content auf Position 10 und fiel dann durchschnittlich um 2,7 Positionen. Die zweite URL stand vor dem Einbau auf Position 20 und stieg dann um 2,4 Positionen. Die URLs haben also durchschnittlich ungefähr um die gleiche Differenz ihre Position verändert, nur eben einmal verbessert und einmal verschlechtert. Der unterschiedliche Verlauf zeigt sich auch bei der Rankingentwicklung in Abbildung 23. Allerdings ist die Schwankung hier deutlich größer, als die Durchschnittswerte vermuten lassen. Im zweiten Fall (blaue Linie) wurde wohl zunächst Duplicate Content erkannt und abgestraft, das Urteil wurde dann aber wohl wieder revidiert und die Seite stieg wieder stark im Ranking. Allgemein scheint Google bei dieser Stufe Probleme zu haben, den Duplicate Content zu erkennen und es ist nicht eindeutig, wie dieser hier bewertet wird.

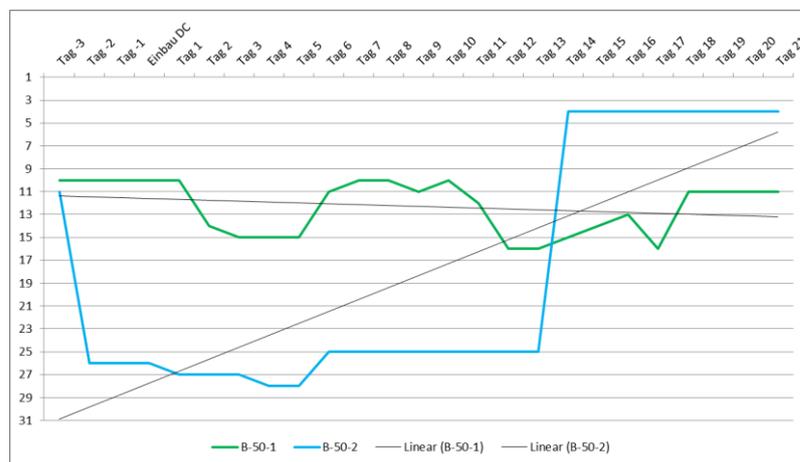


Abbildung 23: Rankingentwicklung der URLs mit 50 % internem Duplicate Content

70 %: Bei den 70 %-URLs gestaltet sich das Bild eindeutig. Es haben sich beide URLs im Durchschnittsranking verschlechtert und es liegt ein ähnlicher Verlauf vor (siehe Abbildung 24). Die Differenz der Positionen und die Höhe der Ausschläge sind zwar unterschiedlich, dies liegt aber an der unterschiedlichen Ausgangsposition. So stand die erste URL (blaue Linie) vorher auf Position 3 und fiel um durchschnittlich 16,9 Positionen. Die zweite URL (violette Linie) stand vorher schon recht weit unten auf Position 25 und verschlechterte sich dann nur um durchschnittlich 0,4 Positionen. Ein Anteil von 70 % internem Duplicate Content scheint sich also eindeutig negativ auf das Ranking auszuwirken und eine mögliche Toleranzgrenze somit zu überschreiten. Auch wenn die Seiten nach 14 Tagen noch einmal gestiegen sind, liegt hier ein einheitliches Muster zur Behandlung vor.

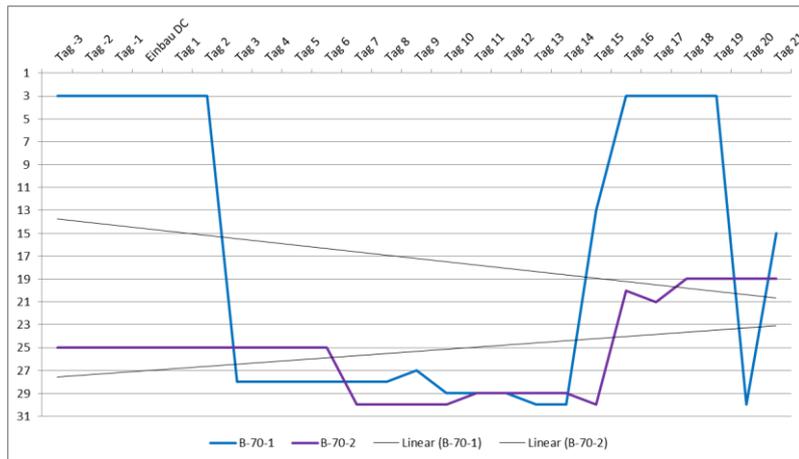


Abbildung 24: Rankingentwicklung der URLs mit 70 % internem Duplicate Content

90 %: Bei den URLs mit 90 % internem Duplicate Content wurde eine andere Entwicklung erwartet, als beobachtet wurde. Durch die Erkenntnisse aus den vorherigen Tests ist davon auszugehen, dass sich das Ranking der URLs stark verschlechtert hat. Dies ist aber nicht der Fall. Die erste URL stand vor dem Einbau von Duplicate Content durchschnittlich auf Position 6,6 und stieg dann um 3,8 Positionen. Die zweite URL stand vor dem Einbau auf Position 9 und verschlechterte sich dann durchschnittlich um -0,6 Positionen. Die Rankingentwicklung ist ähnlich (siehe Abbildung 25). Die Seiten veränderten ihre Position aber letztendlich kaum, was auch die fast waagerechten Trendlinien zeigen. Es ist unklar, warum Google hier so reagiert hat. Möglicherweise wurde der Duplicate Content nicht erkannt und die URLs haben sich nur aufgrund von Positionsveränderungen anderer URLs mit verschoben.

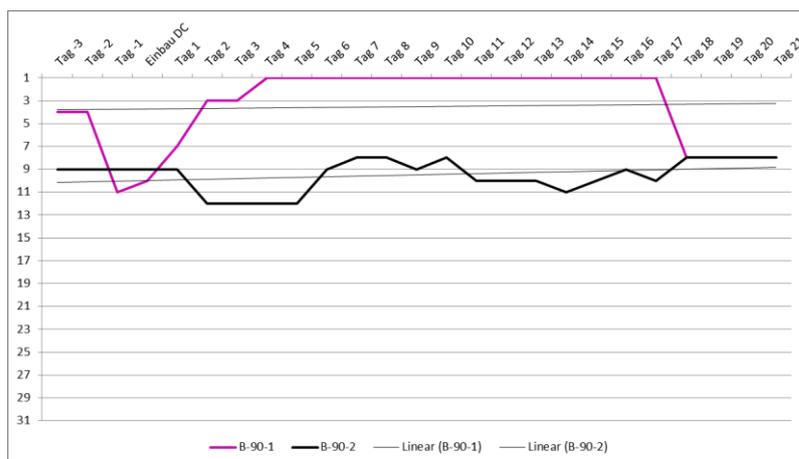


Abbildung 25: Rankingentwicklung der URLs mit 90 % internem Duplicate Content

Zwischenfazit Typ B-URLs

Obwohl sich die 90 %-URLs anders als erwartet entwickelten, scheint die Toleranzgrenze für internen Duplicate Content hier bei 70 % zu liegen. Diese reagierten zumindest eindeutig und einheitlich. Bei den 50 %-URLs scheint es Probleme mit der Einstufung und Erkennung des Duplicate Contents zu geben und bei den 10 und 30 %-URLs gibt es keine negativen Auswirkungen.

Typ B: Duplicate Content intern geklaut

Bei den URLs, von denen Content intern geklaut wurde, wird erwartet, dass einige von ihnen als Original erkannt werden und im Ranking steigen. Auch hier wird wieder angenommen, dass, wenn auf den Content-Klau reagiert wurde, eine ähnliche Rankingentwicklung vorliegt.

Abbildung 26 zeigt die Durchschnittspositionen der URLs, bei denen die jeweiligen Stufen Duplicate Content intern geklaut wurden:

interner DC	vor Einbau DC	nach Einbau DC	Differenz
B-10-1-g	16	17,4	-1,4
B-10-2-g	23	18,2	4,8
B-30-1-g	10	22,2	-12,2
B-30-2-g	17	11,3	5,7
B-50-1-g	11	12,0	-1,0
B-50-2-g	8	22,0	-14,0
B-70-1-g	1	7,1	-6,1
B-70-2-g	6	11,3	-5,3
B-90-1-g	28,6	7,4	21,2
B-90-2-g	18	18,4	-0,4

Abbildung 26: Durchschnittspositionen der beklauten Typ-B-URLs

10 %: Hier liegt kein eindeutiges Ergebnis vor, was die Durchschnittspositionen und die Rankingentwicklung in Abbildung 27 zeigen. Die erste URL stand vor dem Einbau vom Duplicate Content auf Position 16 und verschlechterte das Ranking dann durchschnittlich um 1,4 Positionen. Die zweite URL stand vorher auf Position 23 und verbesserte sich um 4,8 Positionen. Auch das Bild der Rankingentwicklung ist nicht eindeutig. Es erinnert aber an die Entwicklung der Seiten, bei denen 10 % interner Duplicate Content eingebaut wurde. Es liegt die Vermutung nahe, dass der interne Duplicate Content

hier nicht erkannt wurde und sich die URLs nur aufgrund anderer Positionsveränderungen mit verschoben haben.

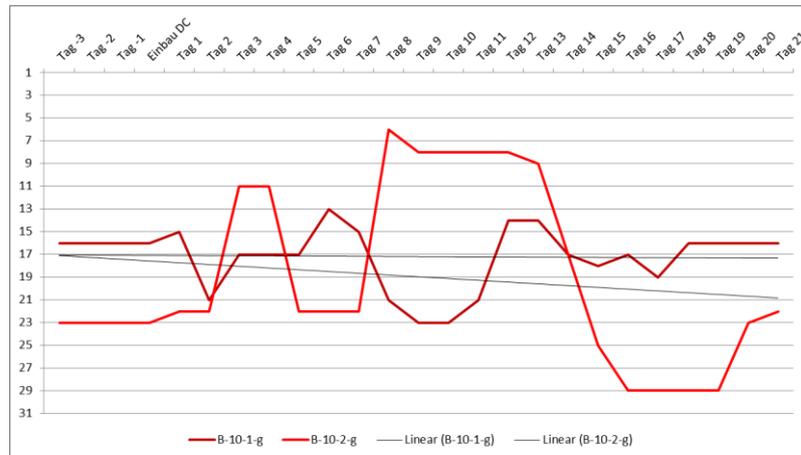


Abbildung 27: Rankingentwicklung der URLs, von denen 10 % Content intern geklaut wurde

30 %: Das Bild gestaltet sich auch hier ähnlich wie bei den URLs, bei denen 10 % interner Duplicate Content geklaut wurde. Die erste URLs stand vorher auf Position 10 und verschlechterte ihr Ranking dann um 12,2 Positionen. Die zweite URL stand vorher auf Position 17 und verbesserte sich um 5,7 Positionen. Auch hier kann beobachtet werden, dass sich die URL, die weiter oben im Ranking stand, verschlechtert hat und die andere ihre Position durchschnittlich verbesserte. Dies bestätigt sich auch in der Rankingentwicklung in Abbildung 28. Auch hier wurde der Content-Klau wohl nicht erkannt.

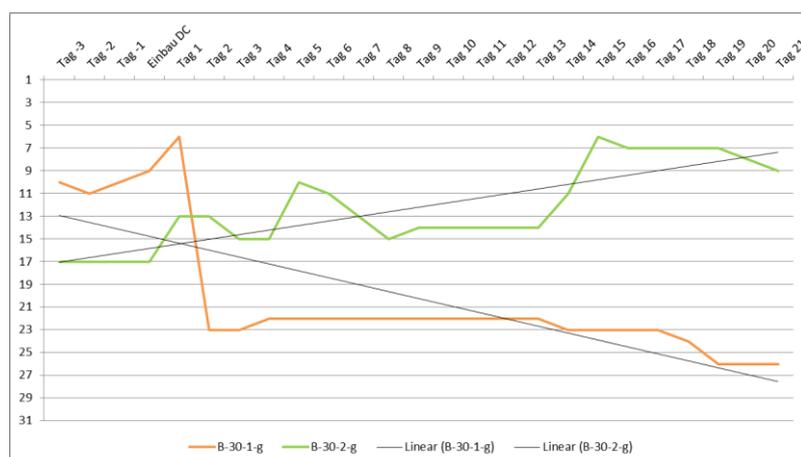


Abbildung 28: Rankingentwicklung der URLs, von denen 30 % Content intern geklaut wurde

50 %: Bei den Durchschnittspositionen haben sich beide URLs verschlechtert, allerdings um unterschiedlich große Differenzen. B-50-1-g stand vorher auf Position 11 und verschlechterte sich um 1 Position. B-50-2-g stand auf 8 und fiel durchschnittlich um 14 Positionen im Ranking. Bei der Rankingentwicklung in Abbildung 29 zeigt sich, dass sich B-50-1-g insgesamt kaum verändert hat (grüne Linie) und B-50-2-g sehr stark gefallen ist (blaue Linie). Bei den URLs, bei denen 50 % interner Duplicate Content eingebaut wurde, ist B-50-2 stark gestiegen. Vielleicht wurde diese fälschlicherweise als Original identifiziert und B-50-2-g als diejenige, die den Content kopiert hat. Die unterschiedliche Behandlung der URLs lässt darauf schließen, dass es auch hier vermutlich keine einheitliche Regel gibt. Insgesamt scheint sich der interne Content-Klau aber eher negativ auf das Ranking der beklauten URL auszuwirken.

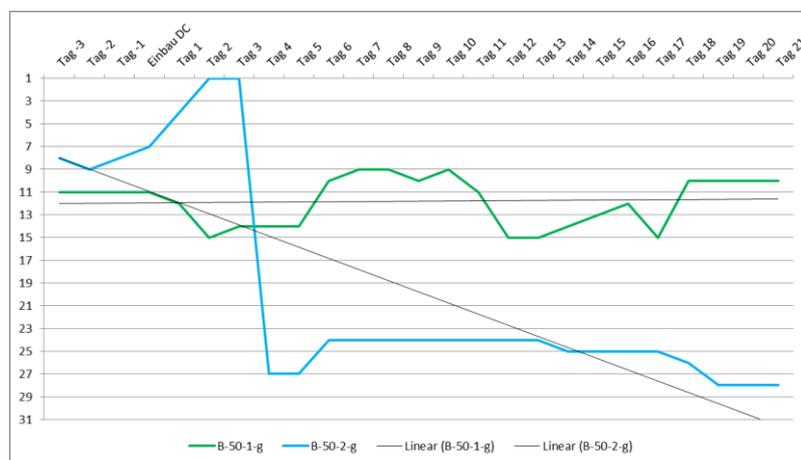


Abbildung 29: Rankingentwicklung der URLs, von denen 50 % Content intern geklaut wurde

70 %: Wie schon bei den Seiten, bei denen 70 % interner Duplicate Content eingebaut wurde, ist das Ergebnis auch bei den Seiten, die bestohlen wurden, eindeutig. Beide URLs standen vorher in den Top 10 und verschlechterten sich danach durchschnittlich um den gleichen Wert. Die erste URL fiel von Position 1 um 6,1 Positionen und die zweite von Position 6 um 5,3 Positionen. Ein ähnlicher Verlauf zeigt sich auch bei der Rankingentwicklung der URLs in Abbildung 30. Die Regeln für die Behandlung scheinen hier eindeutig zu sein. Allerdings müssten die URLs theoretisch als Originalquelle identifiziert werden und im Ranking steigen.

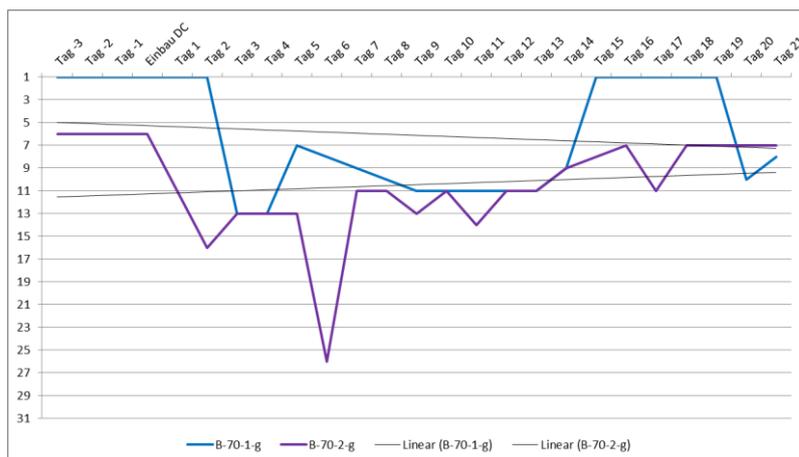


Abbildung 30: Rankingentwicklung der URLs, von denen 70 % Content intern geklaut wurde

90 %: Die erste URL stieg hier von Position 28,6 um durchschnittlich 21,2 Positionen. Das ist die größte Differenz, die überhaupt beobachtet wurde. Google scheint hier wohl erkannt zu haben, dass es sich bei dieser Seite um den Originalcontent handelt, der auf einer anderen Seite intern erneut verwendet wurde. Dieser positive Aufwärtstrend zeigt sich auch in der Rankingentwicklung in Abbildung 31. Die zweite URL hingegen veränderte sich kaum und verschlechterte sich durchschnittlich um 0,4 Positionen (von Position 18).

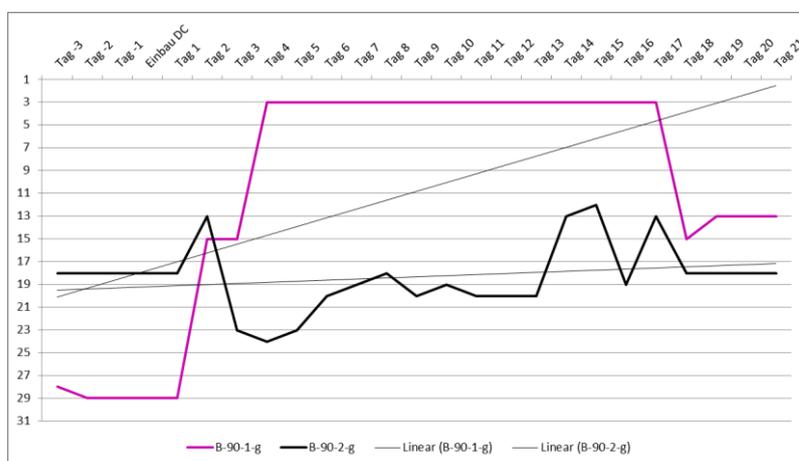


Abbildung 31: Rankingentwicklung der URLs, von denen 90 % Content intern geklaut wurde

Zwischenfazit Typ-B-eklaut

Auch hier wurden wieder die 70 %-Seiten gleich behandelt. Bei den anderen Stufen wurde der Duplicate Content wohl nicht eindeutig erkannt und die Seiten entwickelten sich unterschiedlich. Es hätten theoretisch mehr URLs als Originalquelle identifiziert werden und im Ranking steigen müssen. Insgesamt haben sich hier aber mehr URLs

durchschnittlich im Ranking verschlechtert, als bei den URLs, auf denen der Duplicate Content eingebaut wurde. Vielleicht hat Google bei internem Duplicate Content noch Probleme zu erkennen, welches die Originalseite ist.

Typ C: Externer Duplicate Content eingebaut

Bei den URLs, bei denen externer Duplicate Content eingebaut wurde, werden besonders bei den Seiten mit mehr als 50 % externem Duplicate Content große Rankingverschlechterungen erwartet. Wie zuvor auch, werden hier die Durchschnittspositionen und die Rankingentwicklungen der einzelnen URLs betrachtet. Die Domain, bei der 100 % externer Duplicate Content eingebaut wurde, wird gesondert behandelt (siehe unten).

Abbildung 32 zeigt die Durchschnittspositionen der URLs, bei denen die jeweiligen Stufen Duplicate Content extern eingebaut wurden:

externer DC	vor Einbau DC	nach Einbau DC	Differenz
C-10-1-D2	20	19,1	0,9
C-10-2-D5	25,8	18,3	7,5
C-30-1-D3	10	10	0
C-30-2-D1	21	12,3	8,7
C-50-1-D1	27	12,7	14,3
C-50-2-D4	12	10	2
C-70-1-D6	22,8	27,1	-4,3
C-70-2-D3	30	26,7	3,3
C-90-1-D4	8	17,9	-9,9
C-90-2-D6	7	10,8	-3,8

Abbildung 32: Durchschnittspositionen der Typ-C-URLs

10 und 30 %: Wie beim internen Duplicate Content auch haben sich jeweils beide URLs, bei denen 10 oder 30 % externer Duplicate Content eingebaut wurde im Ranking verbessert. Die Rankingentwicklungen sind jeweils relativ ähnlich und es liegen positive Trends vor (siehe Abbildung 33 und 34). Bei den 30 %-URLs sind die Phasen der Entwicklung jeweils verschoben. Der kleine Anteil von 10 bzw. 30 % externem Duplicate Content wird also auch hier in keiner Weise bestraft. Er wirkt sich sogar eher positiv aus, sofern die Entwicklung auf der Erkennung vom Duplicate Content beruht.

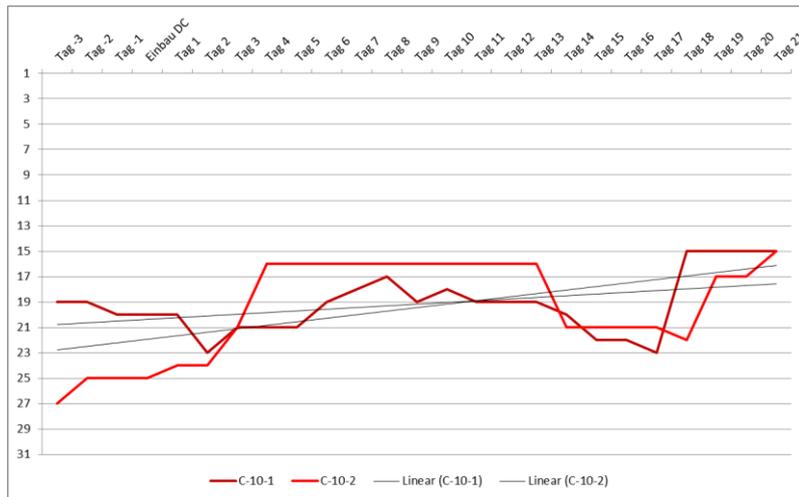


Abbildung 33: Rankingentwicklung der URLs mit 10 % externem Duplicate Content

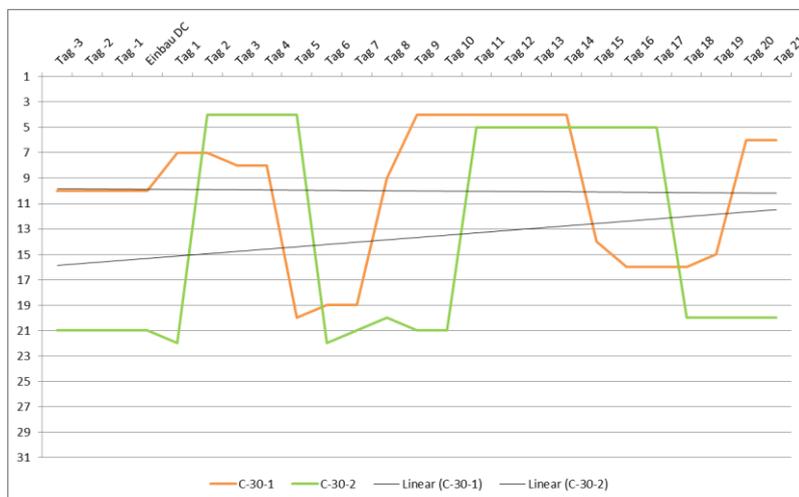


Abbildung 34: Rankingentwicklung der URLs mit 30 % externem Duplicate Content

50 %: Anders als beim internen Duplicate Content haben sich hier beide URLs nach dem Einbau vom Duplicate Content im Ranking verbessert. Die erste URL stieg um durchschnittlich 14,3 Positionen von Rang 27. Die Zweite verbesserte sich von Position 12 um 2 Positionen im Ranking. Auch hier hängt allerdings die Größe der Differenz von der Ausgangsposition ab. Die gesamte Rankingentwicklung ist eher unterschiedlich und erfolgt nicht nach dem gleichen Muster (siehe Abbildung 35). Wie bereits beim internen Duplicate Content, scheint Google bei der 50 %-Grenze auch hier Probleme bei der Behandlung zu haben. Zumindest löst auch dieser Anteil an externem Duplicate Content keine negative Entwicklung aus.

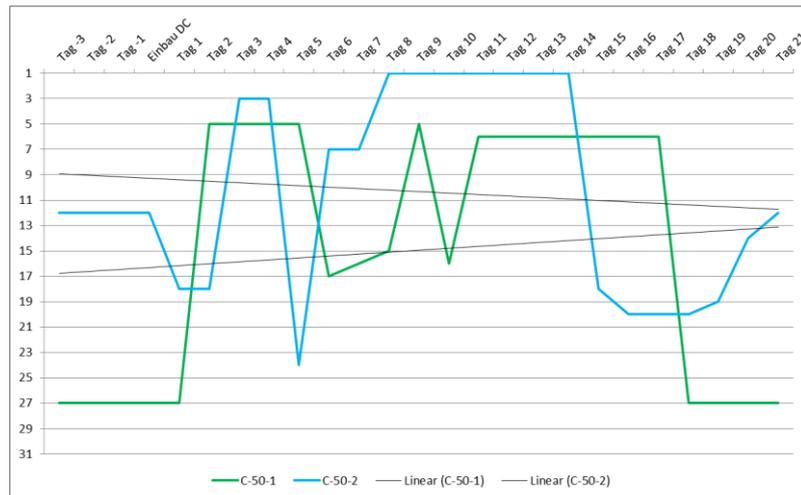


Abbildung 35: Rankingentwicklung der URLs mit 50 % externem Duplicate Content

70 %: Hier gibt es kein eindeutiges Ergebnis. Die erste URL stand vor dem Einbau des Duplicate Contents auf Position 22,8 und verschlechterte sich dann um durchschnittlich 4,3 Positionen. Die zweite URL stand vorher auf Position 30 und stieg dann um 3,3 Positionen. Die Rankingentwicklung zeigt allerdings einige Ähnlichkeiten (siehe Abbildung 36). Es sieht so aus, als ob bei der ersten URL (blaue Linie) der Duplicate Content erkannt wurde und sie im Ranking herabgesetzt wurde. Auch die Trendlinie ist negativ. Die zweite URL stand vorher auf der letzten Position im Ranking und konnte nicht weiter im Ranking fallen. Es ist nur eine Vermutung, dass sie das getan hätte, wenn es noch weiter nach unten gegangen wäre. So liegt kein eindeutiges Ergebnis vor und es ist nicht ganz klar, ob der Duplicate Content bei der ersten URL wirklich als solcher identifiziert und sie darum gezielt im Ranking zurückgesetzt wurde.

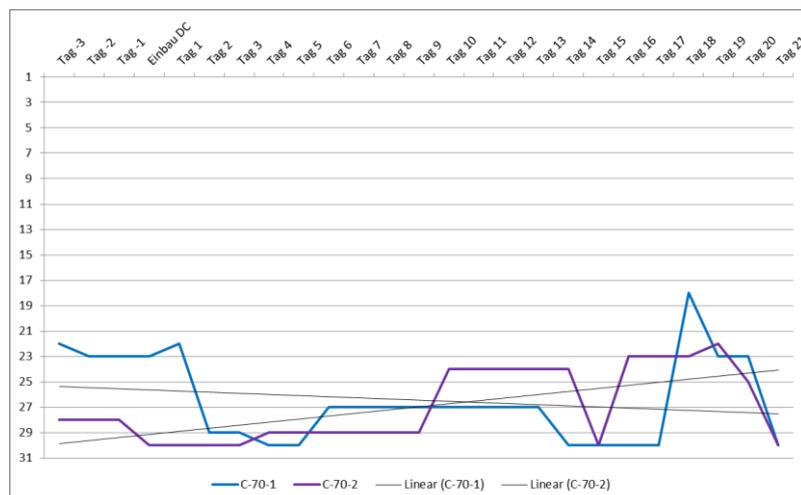


Abbildung 36: Rankingentwicklung der URLs mit 70 % externem Duplicate Content

90 %: Bei den URLs, bei denen 90 % externer Duplicate Content eingebaut wurde, scheint das Ergebnis wieder eindeutig zu sein. Beide URLs verschlechterten ihr Ranking nach dem Einbau des Duplicate Content. Die erste URL stand zuvor auf Position 8 und verschlechterte sich um 9,9 Positionen, die zweite sank um durchschnittlich 3,8 Positionen von Rang 7 ausgehend. Auch in der Rankingentwicklung ergeben sich fast parallele negative Trendlinien (siehe Abbildung 37), obwohl es in der Entwicklung selbst Unterschiede gibt. Dennoch kann davon ausgegangen werden, dass der Duplicate Content hier erkannt wurde und die Seiten daher im Ranking zurückgesetzt wurden.

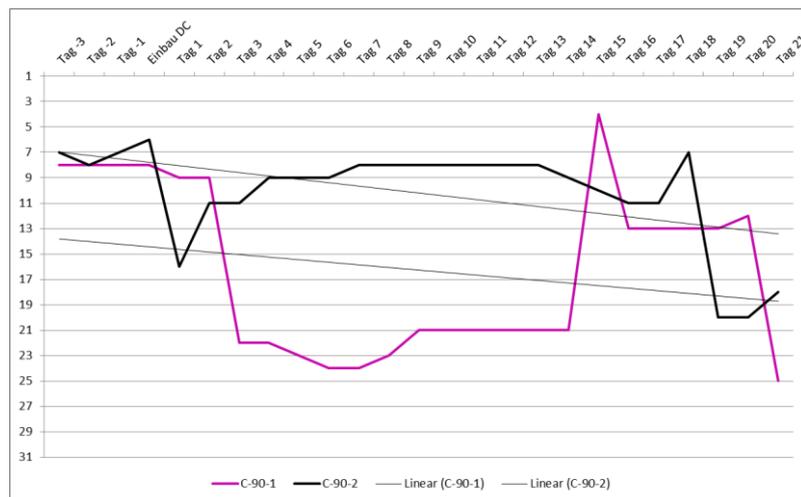


Abbildung 37: Rankingentwicklung der URLs mit 90 % externem Duplicate Content

Zwischenfazit Typ C

Bei den Seiten, bei denen externer Duplicate Content eingebaut wurde, wurden größere negative Veränderungen erwartet. Hier haben sich insgesamt nur 3 URLs im Ranking durchschnittlich verschlechtert. Die Toleranzgrenze für externen Duplicate Content scheint also recht hoch zu liegen und es werden erst Seiten mit großen Anteilen an Duplicate Content abgestraft (mehr als 70 %).

Typ D: Duplicate Content extern geklaut

Bei den Seiten, von denen Content extern geklaut wurde, wird erwartet, dass diese als Originalquelle identifiziert werden und im Ranking steigen. Dies soll durch die Betrachtung der Durchschnittspositionen und Rankingentwicklungen gezeigt werden. Auch hier

werden die URLs, von denen 100 % Content geklaut wurde, gesondert im nächsten Punkt behandelt.

Abbildung 38 zeigt die Durchschnittspositionen der URLs, bei denen die jeweiligen Stufen Duplicate Content extern geklaut wurden:

externer DC	vor Einbau DC	nach Einbau DC	Differenz
D2-10-1	1	21,4	-20,4
D5-10-2	1,6	8,8	-7,2
D3-30-1	16	12,8	3,2
D1-30-2	13	1,1	11,9
D1-50-1	24	2	22
D4-50-2	25	14,7	10,3
D6-70-1	20	11,6	8,4
D3-70-2	27	19,2	7,8
D4-90-1	19	13,6	5,4
D6-90-2	16	6,7	9,3

Abbildung 38: Durchschnittspositionen der Typ-D-URLs

10 %: Es haben sich beide URLs, von denen 10 % Content extern geklaut wurde, im Ranking verschlechtert. Sie standen vorher beide an der Spitze des Rankings. Die erste URL fiel von Rang 1 um durchschnittlich 20,4 Positionen, also eine sehr große Differenz. Die Zweite verschlechterte sich von Position 1,6 um 7,2 Positionen. Diese negative Entwicklung zeigt sich auch deutlich im Rankingverlauf (siehe Abbildung 39). Es ist nicht ganz klar, warum diese URLs so stark fielen, obwohl nur so wenig Content geklaut wurde. Eine Vermutung ist, dass die Ursache für die Verschlechterung nicht der Duplicate Content ist, sondern, dass die URLs von anderen, die im Ranking gestiegen sind, verdrängt wurden. Diese Seiten wurden zumindest anscheinend nicht als Originalquelle erkannt.

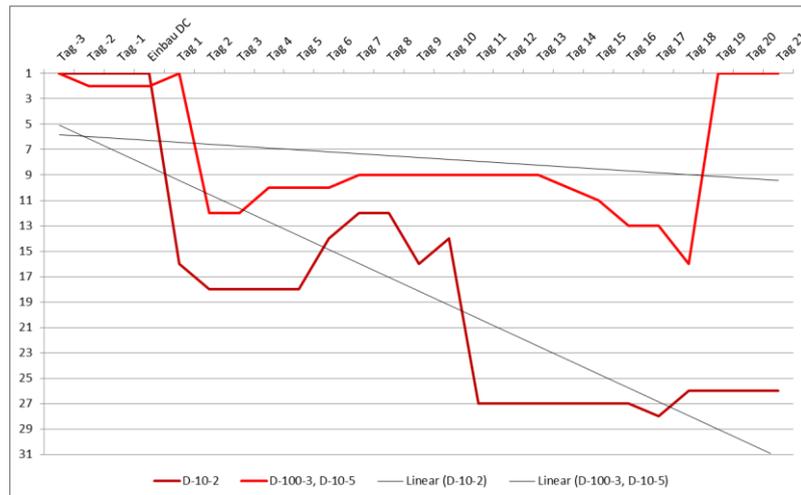


Abbildung 39: Rankingentwicklung der URLs, von denen 10 % Content extern geklaut wurde

30, 50, 70 und 90 %: Hier haben sich alle URLs in ihrer Durchschnittsposition verbessert (Werte können der Abbildung 38 entnommen werden). Dieser positive Trend lässt sich auch bei den meisten Rankingentwicklungen ablesen. Beide 30 %-URLs weisen ähnliche Entwicklungen auf und haben parallele positive Trendlinien, was auf ein gleiches Behandlungsmuster schließen lässt (siehe Abbildung 40). Bei den 50 %-URLs ist die Entwicklung eher unterschiedlich (siehe Abbildung 41). Die erste URL ist dauerhaft gestiegen (grüne Linie), die zweite URL ist dann wieder stark gefallen (blaue Linie). Dennoch weisen beide eine positive Trendlinie auf. Bei den URLs, von denen 70 % Content geklaut wurde, liegt wieder eine eindeutige Entwicklung vor. Beide URLs sind stetig im Ranking gestiegen (siehe Abbildung 42). Sie wurden eindeutig als Originalquelle erkannt. Die 90 %-URLs weisen eher unterschiedliche Entwicklungen auf (siehe Abbildung 43). Die zweite URLs scheint als Original erkannt worden zu sein (schwarze Linie), bei der ersten gibt es immer wieder Schwankungen. Dennoch haben auch hier beide eine positive Trendlinie. Insgesamt kann trotzdem gesagt werden, dass ab den 30 % URLs alle richtig als Originalquelle identifiziert worden und entsprechend im Ranking gestiegen sind. Die Ergebnisse sind hier bis auf kleine Ausnahmen bei 50 und 90 % einheitlich und eindeutig.

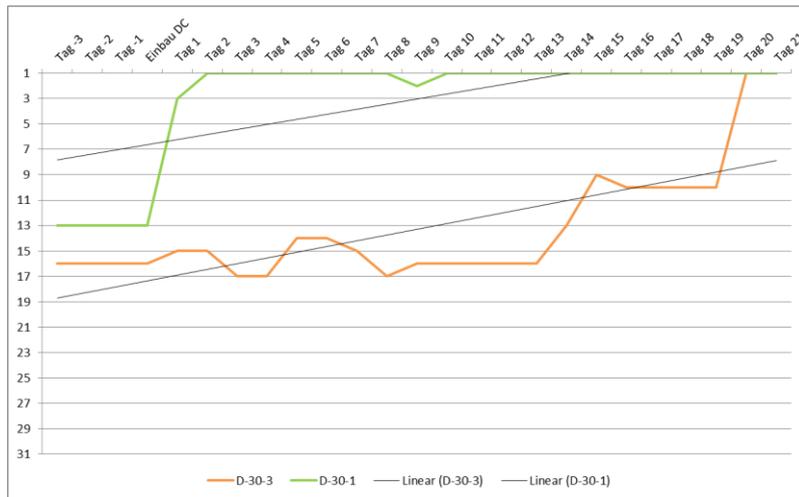


Abbildung 40: Rankingentwicklung der URLs, von denen 30 % Content extern geklaut wurde

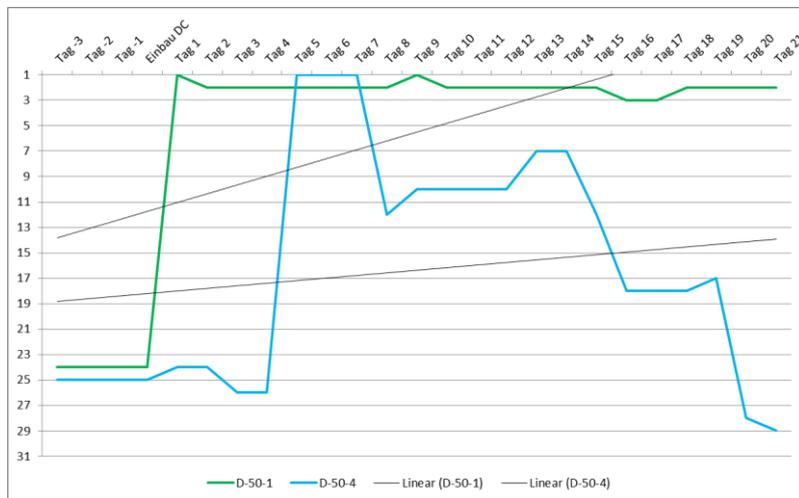


Abbildung 41: Rankingentwicklung der URLs, von denen 50 % Content extern geklaut wurde

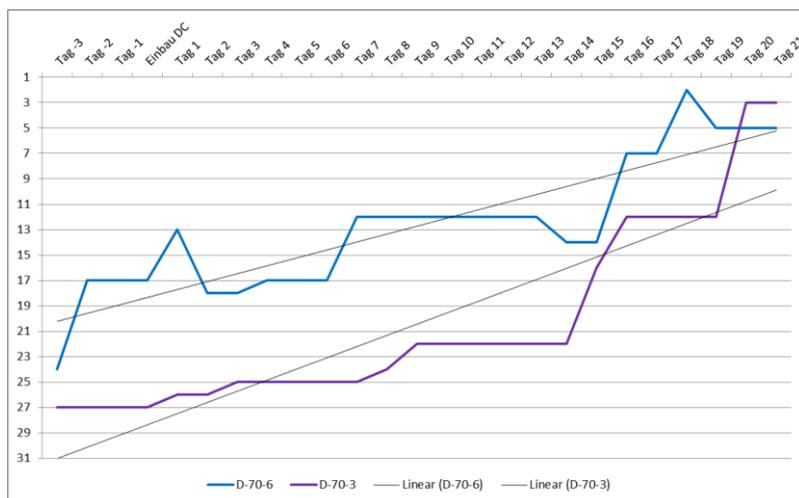


Abbildung 42: Rankingentwicklung der URLs, von denen 70 % Content extern geklaut wurde

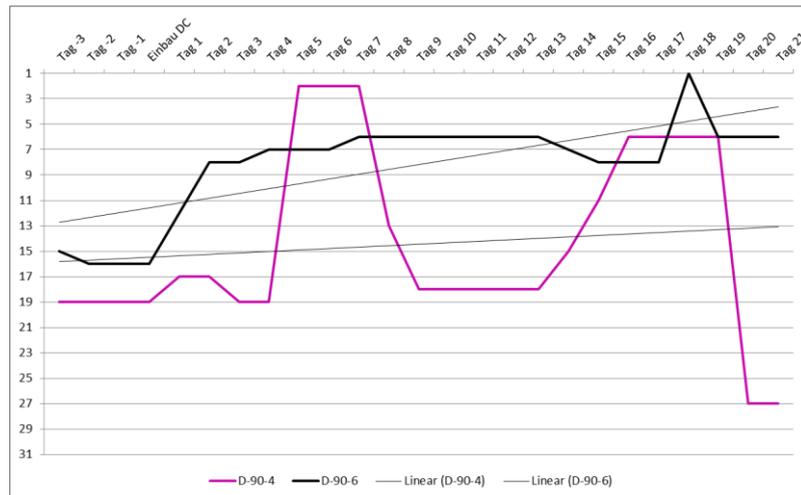


Abbildung 43: Rankingentwicklung der URLs, von denen 90 % Content extern geklaut wurde

Zwischenfazit Typ D

Das Ergebnis der Seiten, von denen Duplicate Content extern geklaut wurde, ist eindeutig: Hier wurden ab 30 % (fast) alle Seiten als Original erkannt und sind im Ranking gestiegen. Die eindeutigsten Ergebnisse liefern die 30 und 70 %-URLs, bei 50 und 90 % gab es bei jeweils einer URL kleine Abweichungen.

100 % externer Duplicate Content

Es werden nun die Domains betrachtet, die mit 100 % externem Duplicate Content belastet wurden. Von einer Domain wurde also der Inhalt von allen drei URLs kopiert und bei einer anderen Domain eingefügt. Es wird erwartet, dass die beklautete Domain als Original erkannt wird und stark im Ranking steigt und dass die andere Domain stark im Ranking fällt. Auch hier werden wieder die Durchschnittspositionen und die Rankingentwicklungen der einzelnen URLs betrachtet.

Abbildung 44 zeigt die Durchschnittspositionen der URLs, bei denen 100 % Duplicate Content extern eingebaut oder geklaut wurde:

	100%	vor Einbau DC	nach Einbau DC	Differenz
C-100-1-D5		17,4	26,4	-9,0
C-100-2-D5		24,4	26,0	-1,6
C-100-3-D5		5,4	25,8	-20,4
D5-100-1		3	10,4	-7,4
D5-100-2		1,4	9,8	-8,4
D5-100-3		1,6	8,8	-7,2

Abbildung 44: Durchschnittspositionen der URLs, bei denen 100 % externer Duplicate Content eingebaut wurde (C-100) und von denen 100 % Content geklaut wurde (D5-100)

Es haben sich durchschnittlich alle URLs im Ranking verschlechtert, egal ob der Content bei ihnen eingebaut oder geklaut wurde. Die URLs, bei denen der Content eingebaut wurde, fielen zum Teil um eine große Differenz. Am meisten verschlechterte sich die dritte URL. Sie lag vorher auf Position 5,4 und fiel dann um durchschnittlich 20,4 Positionen. Am Ende der Testphase lagen alle URLs durchschnittlich zwischen Position 24 und 26, also schon recht weit unten im Ranking. Es kann also durchaus gesagt werden, dass 100 % externer Duplicate Content zuverlässig erkannt und auch stark abgestraft wurde. Dies zeigt sich auch im Verlauf des Rankins in Abbildung 45. Vom Anstieg am Ende abgesehen, ist deutlich zu sehen, dass die Seiten stark abgestraft wurden und sich lange Zeit sogar auf den letzten Positionen 28 bis 30 befanden.

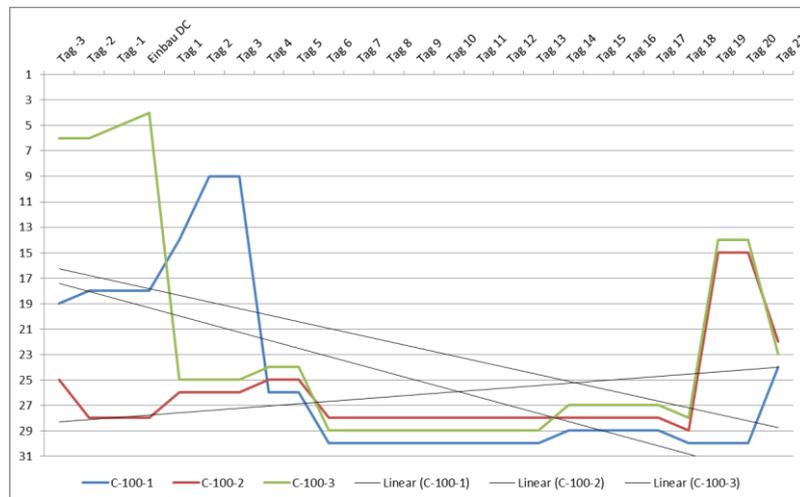


Abbildung 45: Rankingentwicklung der URLs, bei denen 100 % externer Duplicate Content eingebaut wurde

Die URLs, von denen der Content geklaut wurde, standen vor dem Content-Klau an der Spitze des Rankings von Position 1 bis 3. Nach dem Klau lagen sie durchschnittlich auf Position 8 bis 10. Auch sie sind also im Ranking gefallen, wenn auch nicht ganz so

stark. Der Rankingverlauf zeigt eine ähnliche Entwicklung der URLs (siehe Abbildung 46). Zunächst sind die Seiten gefallen, am Ende der Testphase stiegen sie aber wieder auf die Positionen 1 bis 3. Es kann also vermutet werden, dass die URLs am Ende der Testphase noch als Originalquelle erkannt wurden.



Abbildung 46: Rankingentwicklung der URLs, von denen 100 % Content extern geklaut wurde

Zwischenfazit 100 % externer Duplicate Content

Der eingebaute Duplicate Content wurde bei den URLs eindeutig erkannt und abgestraft. Die Originalseiten, von denen der Content geklaut wurde, sind erst am Ende der Testphase als solche identifiziert worden. Hier wurde ein eindeutigeres Ergebnis erwartet.

Gesamtfazit Typ A bis D

Die Ergebnisse des Website-Testes entsprechen zum Teil den Erwartungen, zum Teil aber auch nicht. Die eindeutigsten Ergebnisse gibt es jeweils bei den Seiten mit 70 % internem oder externem Duplicate Content. Hier scheint jeweils die Toleranzgrenze zu liegen, obwohl es bei den 90 % Seiten keine eindeutigen Ergebnisse gibt. Bei den 50 %-Seiten gibt es anscheinend Probleme und zum Teil gegensätzliche Entwicklungen im Ranking. Bei den Seiten mit 10 oder 30 % Duplicate Content gab es entweder kaum Veränderungen oder aber gegenteilige Veränderungen zu den Erwartungen. Der Duplicate Content scheint hier als solcher noch keine Rolle zu spielen.

Bei den Seiten, von denen der Content geklaut wurde, wurden beim externen Duplicate Content zuverlässig fast alle als Original erkannt und sind im Ranking gestiegen. Beim internen Duplicate Content gelang dies nicht. Hier haben sich sogar mehr Seiten verschlechtert als bei denen, wo Duplicate Content eingebaut wurde. Beim internen Duplicate Content scheint es für Google also schwieriger zu sein, zuverlässig zu identifizieren, welches die Originalseite ist.

Die Entwicklung der Seiten mit 100 % externem Duplicate Content entspricht den Erwartungen. Die Seiten, bei denen der Content eingebaut wurde, sind stark im Ranking gefallen und die Seiten, die geklaut wurden, wurden zumindest zum Ende der Testphase hin als Original erkannt.

Durch die Beobachtungen und aufgezeigten Rankingentwicklungen wurde erneut bewiesen, dass Duplicate Content in Bezug auf das Ranking in der Suchmaschine Google eine Rolle spielt. Außerdem wurde gezeigt, dass es eine Bestrafung in Form einer Herabsetzung des Rankings gibt, wenn sich zu viel Duplicate Content auf der Seite befindet. Die Grenze liegt hier vermutlich bei mehr als 50 % und es spielt beim Erreichen der Toleranzgrenze keine Rolle, ob es sich um internen oder externen Duplicate Content handelt, da bei beiden Fällen jeweils die Seiten mit mehr als 50 % Duplicate Content abgestraft wurden. Somit wurde auch die dritte Hypothese bestätigt.

Auch die Veränderungen der Durchschnittspositionen zeigen, dass sich interner Duplicate Content grundsätzlich stärker auf das Ranking auswirkt als externer. Es gab hier mehr URLs, die ihre Position durchschnittlich verschlechtert haben. Dabei spielt es aber anscheinend kaum eine Rolle, ob der interne Duplicate Content bei einer URL eingebaut oder geklaut wird. Negative Veränderungen gab es auf beiden Seiten, bei den geklauten URLs sogar mehr.

Widerlegt werden konnte schließlich die letzte Hypothese, dass die Seiten, wenn die Toleranzgrenze für Duplicate Content erst einmal überschritten ist, immer um die gleiche Anzahl an Positionen im Ranking herabgesetzt werden. Es gibt zwar anscheinend einheitliche Muster in der Behandlung, die Seiten werden aber jeweils um unterschiedlich viele Positionen herabgesetzt. Dies hat zum Teil auch mit der Ausgangsposition vor dem Einbau vom Duplicate Content zu tun.

4.7 Zusammenfassung der Ergebnisse

Im Folgenden werden noch einmal die wichtigsten Ergebnisse des Kapitels 4.6 in kurzen Aussagen zusammengefasst. Dazu wird zunächst in Tabelle 11 gezeigt, welche Hypothesen bestätigt und welche widerlegt wurden.

Hypothese	Ergebnis
Duplicate Content spielt in Bezug auf das Ranking in der Suchmaschine Google eine Rolle.	bestätigt
Es gibt eine Bestrafung in Form einer Herabsetzung des Rankings, wenn sich zu viel Duplicate Content auf der Seite befindet.	bestätigt
Dabei spielt es in Bezug auf das Erreichen der Toleranzgrenze keine Rolle, ob es sich um internen oder externen Duplicate Content handelt.	bestätigt
Grundsätzlich wirkt sich aber interner Duplicate Content stärker auf das Ranking aus, als externer	bestätigt
Wenn erst einmal die Toleranzgrenze für Duplicate Content überschritten ist, macht es keinen Unterschied, wie viel Duplicate Content auf der Seite ist, die Seiten werden dann immer um die gleiche Anzahl an Positionen im Ranking herabgesetzt.	widerlegt

Tabelle 11: Bestätigte und widerlegte Hypothesen der Bachelorarbeit

Es folgen nun die wichtigsten Erkenntnisse:

- Google reagiert in den meisten Fällen sofort auf den Duplicate Content.
- Von Zeit zu Zeit scheint Google die Seiten neu zu bewerten und entsprechend neu einzustufen.
- Die Toleranzgrenze liegt sowohl für den internen als auch für den externen Duplicate Content vermutlich bei mehr als 50 %.
- Bei den Seiten, wo 70 % Duplicate Content eingebaut wurde (egal ob intern oder extern) gab es die eindeutigsten Ergebnisse.
- Bei der Bewertung von Seiten mit 50 % internem oder externem Duplicate Content scheint es Probleme zu geben.
- Die Seiten, von denen extern Content geklaut wurde, wurden fast alle zuverlässig als Originalquelle erkannt.
- Beim internen Duplicate Content gibt es anscheinend noch Probleme bei der Identifikation von Original und Kopie.

5 Tipps und Handlungsempfehlungen auf Basis der Test- ergebnisse

Der Website-Test hat ergeben, dass Google rasch auf internen und externen Duplicate Content reagiert, und dass es starke negative Auswirkungen geben kann, wenn die Toleranzgrenze von mehr als 50 % überschritten wird. Außerdem wurde gezeigt, dass sich interner Duplicate Content stärker auf das Ranking auswirkt als externer.

Die Handlungsempfehlung ist daher eindeutig:

Es sollten auf jeden Fall große Anteile von mehr als 50 % an Duplicate Content auf der Seite vermieden werden, besonders intern.

Um Herauszufinden, ob bei der jeweiligen Seite ein Duplicate Content-Problem vorliegt, können die in Kapitel 3.3 vorgestellten Tools genutzt werden. Außerdem können zur Vermeidung und Lösung des Problems die hier vorgestellten Lösungsmöglichkeiten, jeweils auf die Seite individuell angepasst, angewendet werden.

Schon in Kapitel 2 wurde gezeigt, welche große Rolle guter Content für das Ranking spielt. Die Bedeutung stieg in den letzten Jahren enorm und wird auch wahrscheinlich in den nächsten Jahren weiter steigen. Dies hängt auch direkt mit dem größten Rankingfaktor Backlinks zusammen. Guter und hochwertiger Content bekommt auch gute und hochwertige Links. Das ist eine einfache Regel, die unbedingt befolgt werden sollte, wenn mit einer Seite hoch gerankt werden soll.

Das Kopieren von kompletten Seiten sollte in jedem Fall ganz vermieden werden. Hier zeigten sich wie erwartet sehr große negative Auswirkungen. Es kann sich hier der in Kapitel 3.3 vorgestellten Empfehlung von Google selbst angeschlossen werden:

„Erstellen Sie keine doppelten Seiten, Sub-Domains oder Domains, die im Grunde denselben Inhalt haben“ (MOSKWA 2008).

Aber auch bei kleinen Anteilen gab es besonders beim internen Duplicate Content negative Entwicklungen. Auch wenn diese nicht direkt auf den Duplicate Content zurückzuführen sind, kann es dennoch auch bei so kleinen Anteilen einen Nebeneffekt geben.

Am besten wird es sein, wenn auf einer Seite gar kein Duplicate Content verwendet wird und nur individuelle und einzigartige Inhalte präsentiert werden. Dann ist die Wahrscheinlichkeit am höchsten, dass mit der Seite gut gerankt werden kann. Durch guten Content können auch weitere Faktoren angestoßen werden, die eine Seite ebenfalls im Ranking voranbringen: So werden bei gutem Content mehr Backlinks gesetzt, die Besucher verweilen länger auf der Seite und teilen gute Inhalte in sozialen Netzwerken, dadurch kommen wiederum mehr Besucher auf die Seite und es werden vermutlich noch mehr Backlinks gesetzt. So kann die Seite mit einer Basis von gutem und einzigartigem Content immer weiter im Ranking steigen.

An dieser Stelle sei noch einmal auf die Content- und Qualitätsguidelines von Google verwiesen, die bereits in Kapitel 2.3 vorgestellt wurden. Google gibt hier wertvolle Tipps und Hinweise, was sie selbst unter „gutem Content“ verstehen und was entsprechend hoch gerankt wird.

Abschließend noch einmal alle Tipps auf einen Blick:

- Vermeiden Sie internen Duplicate Content, wenn möglich, ganz auf Ihrer Seite.
- Vermeiden Sie mehr als 50 % externen Duplicate Content auf Ihrer Seite.
- Erstellen Sie individuellen und einzigartigen Content, der für Nutzer interessant ist und neue Informationen bietet.
- Beheben Sie ein mögliches Duplicate Content-Problem möglichst schnell und dauerhaft mit den vorgestellten Lösungsmöglichkeiten.
- Erfahren Sie in den Content- und Qualitätsguidelines von Google, was unter „gutem Content“ verstanden wird.

6 Framework für Agenturen mit Hinweisen für den „idealen Test“

Die Ergebnisse des Website-Testes dieser Bachelorarbeit sind besonders für SEO-Agenturen interessant. Diese haben die Möglichkeit, den Kunden Empfehlungen auf Basis der Ergebnisse zu geben. Durch den Website-Test wissen sie, wie die Suchmaschine Google mit Duplicate Content umgeht und wann dieser zu einem Problem für die Rankings der Kunden wird. Das folgende Kapitel richtet sich nun auch speziell an Agenturen. Es sollen Tipps und Anregungen gegeben werden, um selbst einen Test durchführen zu können und in diesem Gebiet weiter zu forschen. Dazu wird zunächst der ideale Test erläutert. Es werden also weitere Szenarien beschrieben, die über den Test im Rahmen der Bachelorarbeit hinaus, noch untersucht werden könnten. Hierbei geht es nicht immer nur um die Identifikation einer Toleranzgrenze, sondern auch um die Reaktion von Google auf Duplicate Content im Allgemeinen und weitere interessante Aspekte wie etwa der Einfluss von Backlinks auf Duplicate Content. Außerdem werden allgemeine Phasen zur Planung und Durchführung eines Website-Testes erläutert und schließlich ein genaues Framework, also ein Prozessplan, zur Wiederholung des Testes der Bachelorarbeit vorgestellt. Dieses enthält die einzelnen nötigen Schritte und zeigt, wann sie durchgeführt werden müssen. Die Wiederholung des Testes könnte zum Beispiel nach einem neuen großen Update von Google sinnvoll sein.

Wie bereits in Kapitel 4 an mehreren Stellen erläutert wurde, konnten einige Faktoren beim Website-Test aufgrund der kurzen Bearbeitungszeit im Rahmen der Bachelorarbeit nicht untersucht werden. Diese Faktoren sollen an dieser Stelle noch einmal zusammenfassend aufgezeigt werden.

Das Testen von fiktiven Keywords hat viele Vorteile. So müssen sich die Seiten nicht erst gegen Wettbewerber durchsetzen und es gibt keine fremden Einflüsse. Es wird also in einem genau abgegrenzten Raum geprüft, der gezielt manipuliert werden kann. Vielleicht wirkt sich der Duplicate Content durch die fremden Einflüsse und anderen Webseiten aber noch einmal anders auf das Ranking aus. Dies kann nur beim Testen auf ein reales Keyword herausgefunden werden. Daher sollten im idealen Test auch einige Websites auf reale Keywords optimiert werden. Möglich wäre die gleichen Stufen an

Duplicate Content sowohl bei Websites, die auf ein reales Keyword optimiert sind, als auch bei Websites, die auf ein fiktives Keyword optimiert sind, einzubauen und die Entwicklung zu vergleichen. So kann herausgefunden werden, inwieweit fremde Einflüsse und andere Wettbewerber eine Rolle spielen.

Die Wahl des richtigen Keywords spielt für den Test ebenfalls eine wichtige Rolle. So muss nicht nur zwischen fiktivem und realem Keyword gewählt werden, sondern es kann sich auch bewusst für ein Money-, Brand- oder Compound-Keyword entschieden werden. Mit Money-Keywords sind Keywords gemeint, auf die es sich lohnt zu ranken und die häufig gesucht werden. Meist steht hinter ihnen eine Kaufabsicht. Beispiele wären „Smartphone Accessoires“ oder „kostenloses Girokonto“. Brand-Keywords enthalten einen Markennamen, zum Beispiel „iPhone“ oder „Sparkasse“. Compound-Keywords schließlich setzen sich aus einem Money- und einem Brand-Keyword zusammen. Beispiele wären „Smartphone Accessoires iPhone“ oder „kostenloses Girokonto Sparkasse“ (vgl. LINKRESEARCHTOOLS 2013). Es wäre interessant herauszufinden, ob sich Google in Bezug auf Duplicate Content anders verhält, wenn es um ein Money-Keyword oder ein Brand-Keyword geht. Ähnliches gilt für verschiedene Branchen. Wird Duplicate Content in der hart umkämpften Finanzbranche möglicherweise stärker bestraft als in einer wenig umkämpften Branche mit weniger Wettbewerbern? Um dies zu überprüfen, könnten einige Domains mit Duplicate Content auf ein starkes Money-Keyword in einer hart umkämpften Branche optimiert werden (zum Beispiel „kostenloses Girokonto“). Zum Vergleich könnten dann Domains in einer wenig umkämpften Branche mit relativ neutralen Keywords platziert und mit den gleichen Anteilen Duplicate Content versehen werden. Die Auswertung der Rankings zeigt dann, ob es Unterschiede gibt.

Ein weiterer interessanter Faktor sind Backlinks. Sie spielen in Bezug auf das normale Ranking eine wichtige Rolle und sind einer der größten Rankingfaktoren (vgl. SEARCHMETRICS 2013). Backlinks werden von Google als Empfehlungen gewertet. Je mehr Backlinks auf eine Seite zeigen, desto wichtiger und beliebter wird diese eingestuft und entsprechend höher gerankt. Doch wie verhält sich das Ranking, wenn Duplicate Content im Spiel ist? Wird das Original noch erkannt, wenn mehr Backlinks auf die Kopie zeigen? Kann eine Seite mit viel Duplicate Content und vielen hochwertigen Backlinks höher ranken als eine Seite mit Unique Content und wenigen Backlinks? Um diese Fragen zu beantworten, bietet es sich an, im Test für einige Seiten gezielt

Backlinkaufbau zu betreiben. Dabei kann zwischen Originalseiten und kopierten Seiten variiert werden, um hier vergleichen zu können. Backlinkaufbau ist allerdings sehr zeitintensiv und sollte über mehrere Monate betrieben werden. Hier ist also eine detaillierte Planung der Durchführung des Testes wichtig.

Ein weiteres gutes Test-Szenario kann die Rolle des Domainalters gerade auch in Bezug auf die Erkennung des Originals darstellen. Was passiert, wenn eine alte und etablierte Seite Content von einer neuen, jungen Seite klaut? Kann die neue, junge Seite dann trotzdem noch damit ranken? Oder wird tatsächlich die alte und etablierte Seite fälschlicherweise als Originalquelle ausgewählt? Um dies herauszufinden, werden Seiten unterschiedlichen Alters benötigt. Optimaler Weise liegen mehrere Jahre zwischen der Erstellung der Seiten. Dies könnte allerdings zum Problem werden, wenn keine so alte Seite vorliegt, die zu Testzwecken verwendet werden kann.

Im Test im Rahmen der Bachelorarbeit wurde nur die Reaktion von Google bei Duplicate Content in Textform untersucht. Content kann aber auch in anderer Form vorliegen, wie etwa als Bild oder Video. Wie reagiert Google hier? Wird es überhaupt erkannt, wenn ständig das gleiche Bild verwendet wird? Um dies herauszufinden, kann der Test ähnlich wie im Rahmen der Bachelorarbeit aufgebaut werden. Einziger Unterschied ist, dass der Duplicate Content nicht in Form von Text, sondern in Form von Bildern oder Videos vorliegt.

Es ist darüber hinaus sinnvoll, insgesamt mehr Seiten, also URLs, für den Test zu erstellen. Je mehr Seiten beim Test vorhanden sind, desto besser können die Rankings verglichen werden und desto genauer sind die Ergebnisse. Dies gilt besonders bei der Optimierung auf fiktive Keywords. Wenn es mehr rankende URLs gibt, kann es auch größere Veränderungen in den Rankings geben. Die Anzahl der benötigten URLs hängt natürlich auch von den Zielen des Testes ab und schließlich auch davon, wie viele Kapazitäten zur Verfügung stehen.

Zusammenfassend ergeben sich also folgende Szenarien, die über den Test in der Bachelorarbeit mit einem fiktiven Keyword hinaus, noch getestet werden könnten:

- auf ein reales Keyword optimierte Seiten, davon:
 - starkes Money-Keyword in einer hart umkämpften Branche
 - eher neutrales Keyword in einer weniger umkämpften Branche

- Seiten mit Backlinkaufbau und Seiten ohne (auf reales Keyword optimiert)
- junge und alte Seiten im Vergleich
- Testen mit verschiedenen Content-Formen

Wenn alle Faktoren wirklich in einem gemeinsamen Test untersucht werden sollen, müssten sehr viele verschiedene Seiten erstellt werden. Daher würde es sich anbieten, den Test über einen längeren Zeitraum laufen zu lassen und die einzelnen Seiten für mehrere Faktoren zu nutzen. So könnten beispielsweise zunächst die Unterschiede im Vergleich vom fiktiven zum realen Keyword untersucht werden. Später könnte dann für einige Seiten vom realen Keyword Backlinkaufbau betrieben und die Auswirkungen untersucht werden.

So ein umfangreicher Test muss genau geplant werden. Welche Phasen zur Planung und Durchführung des Testes durchlaufen werden müssen, wird im Folgenden zunächst allgemein erläutert (in Anlehnung an BOPP 2011). Im Allgemeinen muss der Test geplant, durchgeführt und ausgewertet werden. Dieser gesamte Prozess lässt sich in acht Phasen unterteilen. Wie diese aufeinander aufbauen, wird anschließend in einem einfachen Phasenmodell gezeigt.

Phase #1: Ziele definieren und Hypothesen aufstellen

Bevor der Test durchgeführt werden kann, muss zunächst geklärt werden, was überhaupt getestet werden soll. So kann es verschiedene Gründe für die Durchführung des Testes geben. Nach einem großen neuen Content-Update von Google könnte überprüft werden, ob auch die Behandlung von Duplicate Content verändert wurde. Als Vergleichswert können die Ergebnisse vom Website-Test im Rahmen dieser Bachelorarbeit verwendet werden. Ein weiterer Anlass wäre die Durchführung einer oder aller beschriebener Szenarien aus dem idealen Test. Auch denkbar wäre die Situation, dass auf Basis von Aussagen von Google selbst oder von Beobachtungen von Rankingentwicklungen verschiedene Hypothesen mit Bezug auf Duplicate Content entwickelt werden, die dann in einem Test überprüft werden sollen. In der ersten Phase muss auf jeden Fall festgelegt werden, welche Ziele mit dem Test erreicht und welche Hypothesen überprüft werden sollen.

Phase #2: Planung des Testes

In der zweiten Phase wird dann der Ablauf des Testes genau geplant. Hier wird festgelegt, wie viele Seiten benötigt werden und wie diese aufgebaut sein sollen, damit die in Phase 1 aufgestellten Hypothesen überprüft werden können. Es ist wichtig, dass hier gedanklich verschiedene Situationen durchgespielt werden und mehrmals untersucht wird, ob das Design so auch wirklich alle benötigten Kriterien erfüllt. Auch sollte hier geprüft werden, ob genügend Zeit und Kapazitäten zur Verfügung stehen, um den Test durchführen zu können. Außerdem sollte das Ende des Untersuchungszeitraums festgelegt werden.

Phase #3: Vorbereitung

Nun wird der Test vorbereitet. Das heißt, dass die in Phase 2 festgelegte Anzahl an Domains registriert werden muss, wenn nicht auf bereits vorhandene Domains zurückgegriffen werden kann. Außerdem muss der Content für die Seiten erstellt werden. Wenn genügend Mittel zur Verfügung stehen, kann auch eine Content-Agentur damit beauftragt werden. Schließlich muss sich für ein Content-Management-System entschieden werden und die entsprechenden Seiten erstellt werden.

Phase #4: Online stellen und stabile Rankings

Wenn alle Seiten erstellt sind und Phase 3 abgeschlossen ist, können die Seiten online gestellt werden. Je nachdem, ob der Zeitpunkt des Launches für den Test eine Rolle spielt, werden die Seiten alle zum gleichen oder an verschiedenen Zeitpunkten online gestellt. Danach muss zunächst abgewartet werden, bis die Seiten indexiert sind und stabile Rankings erreicht haben.

Phase #5: Duplicate Content einbinden

Sind die Rankings stabil, kann der Duplicate Content eingebunden werden. Hier ist es sinnvoll, wenn bereits vorher ein genauer Plan erstellt wurde, wie viel Content von welcher Seite wo eingebunden werden soll. Es bietet sich an, den Duplicate Content bei allen Seiten zum gleichen Zeitpunkt online zu stellen. Dann herrschen überall die gleichen Bedingungen. Es kann aber auch Versuchsanordnungen geben, bei denen dies genau nicht gewünscht ist. Das muss alles in Phase 1 und 2 genau geplant werden.

Phase #6: Rankings beobachten

Nun werden über den in Phase 2 festgelegten Zeitraum die Rankings der Seiten beobachtet. Es bietet sich an, hier regelmäßig Screenshots der entsprechenden Google-Abfragen zu machen, um den Verlauf der Rankings zu dokumentieren. Es kann außerdem mit verschiedenen Tabellen und Diagrammen gearbeitet werden, um die Entwicklung darzustellen. Je nach dem, was beim Test untersucht wird, kann es auch sein, dass in dieser Phase noch weitere Tätigkeiten anfallen. Soll zum Beispiel die Wirkung von Backlinks überprüft werden, muss nun verstärkt Backlinkaufbau betrieben werden.

Phase #7: Auswertung

Es schließt sich die Auswertung an. Hier werden die in Phase 6 gemachten Beobachtungen gründlich ausgewertet und miteinander verglichen. Es wird geprüft, ob Regelmäßigkeiten oder ähnliche Entwicklungen festgestellt werden können. Außerdem wird untersucht, ob die in Phase 1 aufgestellten Hypothesen bestätigt oder widerlegt werden konnten. Die Ergebnisse werden dann schön aufbereitet und können entweder veröffentlicht oder für den internen Gebrauch genutzt werden.

Phase #8: Abschluss des Testes

In der letzten Phase wird der Test schließlich abgeschlossen. Wenn die Seiten nicht mehr für andere Zwecke benötigt werden, können sie abgeschaltet werden. Außerdem sollte hier noch einmal geprüft werden, was im Verlauf des Testes gut geklappt hat, was weniger gut und was bei einem zukünftigen Test vielleicht besser gemacht werden könnte.

Phasenmodell

Das in Abbildung 47 dargestellte einfache Phasenmodell zeigt, wie die einzelnen Phasen des Projektes sequentiell aufeinander aufbauen. Das Modell ist einfach, übersichtlich und leicht nachvollziehbar. Eine neue Phase kann immer erst dann begonnen werden, wenn die vorherige komplett abgeschlossen ist. Dies kann zugleich ein Nachteil sein und in der Realität zu Problemen führen (vgl. KARAVUL / VOIGT 2011). Bei der Durchführung des Website-Testes sollte dies aber kein Problem sein. Es ist auch kaum möglich, Phasen hier parallel stattfinden zu lassen.

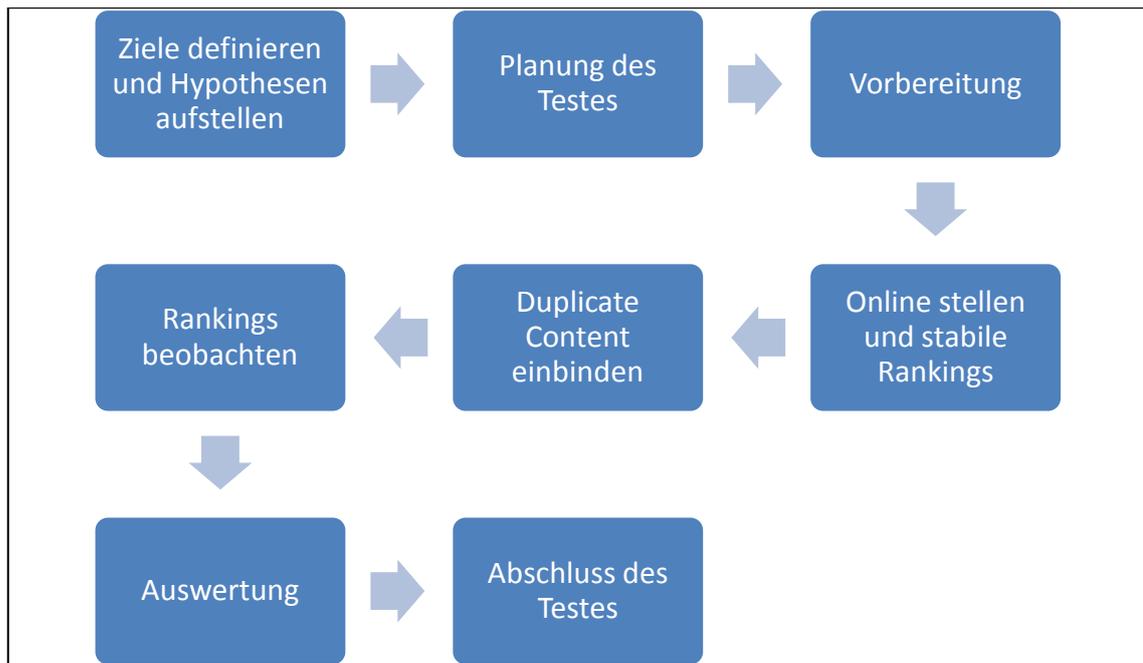


Abbildung 47: Einfaches Phasenmodell zur Durchführung eines Testes

Nachdem nun also die einzelnen Phasen allgemein vorgestellt wurden, wird im Folgenden beschrieben, wie der Website-Test, der im Rahmen dieser Bachelorarbeit durchgeführt wurde, wiederholt werden kann. Dazu wird ein Prozessplan vorgestellt, der genau beschreibt, welche Steps wann durchgeführt werden müssen. In diesem Plan finden sich auch die meisten der oben beschriebenen Phasen wieder.

Prozessplan zur Wiederholung des Website-Testes

Die Wiederholung des Testes bietet sich, wie bereits beschrieben, an, wenn ein neues großes Update von Google durchgeführt wurde. So kann überprüft werden, ob die Ergebnisse des Testes dieser Bachelorarbeit weiterhin gültig sind, oder ob sich etwas am Algorithmus geändert hat und Duplicate Content nun anders behandelt wird. Außerdem könnte der Website-Test auch optimiert auf ein reales Keyword erneut durchgeführt werden, um zu prüfen, ob Google hier genauso wie beim fiktiven Keyword reagiert oder ob fremde Einflüsse eine Rolle spielen.

Für die Wiederholung des Testes gelten selbstverständlich die gleichen unter 4.4 beschriebenen Rahmenbedingungen wie für den Originaltest. Es müssen also 30 Domains mit je drei Texten erstellt werden. Auch hier empfiehlt sich die Optimierung auf drei verschiedene fiktive oder reale Keywords (je nach dem wo die Ergebnisse des Original-

testes überprüft werden sollen). Die Durchführung des Testes ist dabei gleich, egal ob reale oder fiktive Keywords gewählt werden.

Zunächst werden die Domains mit jeweils drei Texten mit Unique Content bestückt. Dieser wird dann nach einiger Zeit teilweise durch Duplicate Content ersetzt. Dabei werden verschiedene Anteilstufen von internem und externem Duplicate Content verwendet und zwar 10%, 30%, 50%, 70% und 90%. Von jeder Stufe werden zwei Domains erstellt, sodass die Ergebnisse miteinander verglichen werden können. Hinzu kommen pro Keyword eine Vergleichsdomain und zwei Domains von denen der Content geklaut wird. Außerdem gibt es noch eine Domain, die 100% externen Duplicate Content enthält, also eine andere Domain komplett kopiert. Es ergibt sich der in Tabelle 12 dargestellte Plan für die benötigten Domains.

Keyword 1	Keyword 2	Keyword 3
Vergleichsdomain 1	Vergleichsdomain 2	Vergleichsdomain 3
Domain, die geklaut wird 1	Domain, die geklaut wird 3	Domain, die geklaut wird 5
Domain, die geklaut wird 2	Domain, die geklaut wird 4	Domain, die geklaut wird 6
10 % interner DC*	10 % interner DC	30 % interner DC
50 % interner DC	30 % interner DC	50 % interner DC
70 % interner DC	70 % interner DC	90 % interner DC
90 % interner DC	30 % externer DC	10 % externer DC
10 % externer DC	50 % externer DC	70 % externer DC
30 % externer DC	70 % externer DC	90 % externer DC
50 % externer DC	90 % externer DC	100 % externer DC

Tabelle 12: Benötigte Domains bei der Optimierung auf drei Keywords (*DC = Duplicate Content)

Der folgende Prozessplan zeigt die Wiederholung des Testes unter gleichen Bedingungen wie beim Originaltest (siehe Abbildung 48). Die oben beschriebene Phase #1 „Ziele definieren und Hypothesen aufstellen“ entfällt, da die gleichen Hypothesen wie beim Originaltest erneut geprüft werden (siehe Kapitel 4.1). Ebenso muss auch Phase #2 „Planung des Testes“ nicht durchgeführt werden, da der Test einfach wiederholt wird und nicht neu geplant werden muss. Alle anderen oben beschriebenen Phasen werden auch hier durchgeführt.

	Keyword 1	Keyword 2	Keyword 3
Woche 1	für ein Keyword entscheiden, Inhalte für Texte überlegen, Domains registrieren		
Woche 2	Texte schreiben		
Woche 3	Seiten erstellen und Launch	Texte schreiben	
Woche 4	stabile Rankings erreichen	Seiten erstellen und Launch	Texte schreiben
Woche 5		stabile Rankings erreichen	Seiten erstellen und Launch
Woche 6	Einbau DC		Rankings beobachten
Woche 7	Rankings beobachten	Einbau DC	
Woche 8		Rankings beobachten	Rankings beobachten
Woche 9			
Woche 10	Auswertung Keyword 1	Auswertung Keyword 2	Rankings beobachten
Woche 11			
Woche 12		Auswertung Keyword 3	
Woche 13	Vergleich der Beobachtungen der einzelnen Keywords und Zusammenfassung der Ergebnisse, Prüfung der Hypothesen		
Woche 14	Abschluss des Testes: was hat gut geklappt, was weniger gut? Ggfs. Löschen der Seiten		

Abbildung 48: Prozessplan zur Wiederholung des Testes

Der Test ist hier für eine Dauer von 14 Wochen und zur Durchführung von einer Person angelegt. Während der Bachelorarbeit wurde der Test innerhalb von 12 Wochen durchgeführt und entsprechend einige Punkte zusammengefasst. Doch in Anbetracht der Tatsache, dass ein Mitarbeiter einer Agentur, der den Test durchführt, vermutlich noch andere Aufgaben erledigen muss, bietet sich die Verlängerung um zwei Wochen durchaus an. Wenn genügend Personal zur Verfügung steht, könnten auch drei Mitarbeiter zusammen den Test durchführen. Jeder Mitarbeiter würde dann ein Keyword bearbeiten. Dann könnten die einzelnen Phasen wie Texte schreiben und Seiten erstellen gleichzeitig stattfinden und der Test könnte also auch in 12 Wochen durchgeführt werden.

Der Prozessplan gibt natürlich nur zeitliche Richtlinien an. Bei Zwischenfällen oder besonderen Situationen muss individuell reagiert werden. Wenn die Rankings nach den hier vorgesehenen zwei Wochen beispielsweise noch nicht stabil sind, sollte mit dem Einbau des Duplicate Content noch etwas gewartet werden.

Zur Beobachtung und besseren Kontrolle der Seiten ist zu empfehlen, die Google Webmaster Tools für jede Domain einzurichten. Hier kann der Indexierungsstatus überprüft und ggfs. Website- und URL-Fehler entdeckt werden. Zur Beobachtung der Rankings bietet es sich an, täglich Screenshots von den entsprechenden Suchergebnisseiten zu machen und diese mit Hilfe von Tabellen und Diagrammen auszuwerten. Wenn ein reales Keyword für den Test ausgewählt wurde, ist vielleicht auch die Anzahl der Zugriffe und Besucher auf den Seiten interessant. Diese lassen sich beispielweise mit Google Analytics auswerten.

Mit Hilfe des beschriebenen Prozessplans sowie den in Kapitel 4 dieser Arbeit beschriebenen Hinweisen zur Durchführung und Auswertung des Originaltestes sollte es für eine SEO-Agentur leicht möglich sein, den Test unter den gleichen Umständen zu wiederholen und die Ergebnisse zu überprüfen. Inwieweit der Prozessplan abgewandelt werden muss, um die anderen oben beschriebenen Szenarien des idealen Testes zu untersuchen, wird im Folgenden erläutert.

Szenario #1: Verschiedene Keyword-Arten

Hierbei geht es darum, den Einfluss von Duplicate Content bei einer auf ein starkes Money-Keyword optimierten Seite in einer hart umkämpften Branche mit einer auf ein neutrales Keyword optimierten Seite in einer wenig umkämpften Branche zu vergleichen. Denkbar wäre auch, hier Branchen zu wählen, in denen die Kunden der Agentur aktiv sind und diese miteinander zu vergleichen.

Für die Durchführung des Testes kann der zuvor beschriebene Prozessplan in leicht abgewandelter Form übernommen werden. Hier reichen zwei Keywords aus, bei denen dann aber die gleichen Phasen hintereinander stattfinden, wie zuvor auch. Der zeitliche Rahmen kann also genauso übernommen werden.

Es verändert sich aber die Anzahl der benötigten Seiten. Wenn zwei Keywords ausgewählt werden, also ein hartes Money-Keyword und ein neutrales Keyword, werden pro Keyword 13 Domains benötigt, also insgesamt 26. Pro Keyword gibt es dann jeweils

zwei Domains mit den bekannten Anteilstufen an Duplicate Content, zwei Domains zum externen Contentklau und eine Vergleichsdomain. Wichtig ist auch hier, dass alle Domains sehr ähnlich aufgebaut werden, damit die Ergebnisse auch wirklich miteinander verglichen werden können. Bei der Vorbereitung, Durchführung und Auswertung der Ergebnisse ändert sich nichts im Vergleich zum zuvor beschriebenen Prozessplan.

Szenario #2: Faktor Backlinks

Es soll hier überprüft werden, inwieweit Backlinks die Reaktion von Google auf Duplicate Content beeinflussen. Dazu werden Seiten mit und ohne Backlinks benötigt. Das Besondere bei diesem Test ist, dass er über einen wesentlich längeren Zeitraum laufen muss, als die bisher beschriebenen, da der Backlinkaufbau sehr zeitintensiv ist. Insgesamt kann für den gesamten Test mit einem Zeitraum von mindestens sechs Monaten gerechnet werden.

Um gute Ergebnisse zu erzielen und viele Vergleichsmöglichkeiten zu haben, werden insgesamt vier verschiedene Arten von Seiten benötigt und zwar:

- Seiten ohne Backlinks, bei denen kein Duplicate Content eingebaut wird
- Seiten ohne Backlinks, bei denen Duplicate Content eingebaut wird
- Seiten mit Backlinks, bei denen kein Duplicate Content eingebaut wird
- Seiten mit Backlinks, bei denen Duplicate Content eingebaut wird

Wenn möglich, sollten Domains für zwei verschiedene reale Keywords erstellt werden. Pro Keyword werden minimal sechs Domains benötigt, jeweils eine ohne Backlinks und mit bzw. ohne Duplicate Content und jeweils zwei mit Backlinks und mit bzw. ohne Duplicate Content. Es sollte dann immer der gleiche Anteil an Duplicate Content eingebaut werden. Wenn verschiedene Stufen an Duplicate Content eingebaut werden sollen, erhöht sich die Anzahl der Seiten entsprechend. Bei den Seiten mit Backlinkaufbau, bei denen Duplicate Content eingebaut wird, kann zudem variiert werden, indem einmal schon vor dem Einbau von Duplicate Content Backlinks aufgebaut werden und einmal erst nach dem Einbau von Duplicate Content damit begonnen wird.

Wie bereits erwähnt, muss für dieses Test-Szenario sehr viel mehr Zeit eingeplant werden. Allein für den Backlinkaufbau sollten mindestens zwei Monate berechnet werden. Hinzu kommt die Zeit für die Beobachtung der Seiten und die Auswertung der Ergebnisse. Die einzelnen Phasen des beschriebenen Projektplans bleiben erhalten, werden

nur entsprechend verlängert. Auch hier stehen zunächst die Vorbereitungen an, also die Wahl der Keywords und die Registrierung von Domains. Danach folgen die Erstellung der Texte und der Launch der Seiten. Dann wird mit dem Backlinkaufbau begonnen. Es ist sicherlich sinnvoll, vorher eine bestimmte Zahl an Backlinks festzulegen, die erreicht werden soll, bevor der Duplicate Content eingebunden wird. Diese gilt dann für alle Seiten, damit die Ergebnisse besser verglichen werden können. Auch sollte darauf geachtet werden, dass die Backlinks in etwa die gleiche Qualität haben. Dazu kann als Richtwert der PageRank der linkgebenden Seite genutzt werden. Nach dem Einbau des Duplicate Contents folgt die zweite Phase des Backlinkaufbaus. Schon während der gesamten Zeit sollten die Rankings der Seiten ständig beobachtet werden. Am Ende folgen schließlich die Auswertung der Ergebnisse und der abschließende Vergleich.

Es ist wichtig, dass im Vorhinein die Ziele und Rahmenbedingungen, wie etwa die Anzahl und Qualität der zu setzenden Backlinks, genau geklärt werden. Außerdem sollten alle Entwicklungen während des gesamten Testes genau dokumentiert werden.

Szenario #3: Junge und alte Seiten

Dieser Test ist vom Umfang her kleiner als die vorher beschriebenen. Es sollen hier junge und ältere Seiten miteinander verglichen und geprüft werden, welche Auswirkung das Domainalter auf die Reaktion von Google auf Duplicate Content hat. Für den Test werden also alte und junge Seiten benötigt. Optimaler Weise liegt schon eine ältere Seite vor (mindestens ein Jahr alt), die zu Testzwecken verwendet werden kann. Wenn ja, dann werden einfach ein bis zwei neue Domains erstellt, die auf das gleiche Keyword optimiert werden. Wenn die Seiten stabile Rankings haben, wird Content von einer der neuen Seite auf der alten Seite eingebaut. Danach wird beobachtet, wie Google darauf reagiert und geprüft, ob die neue Seite trotzdem als Original erkannt wird. Es reicht hier, wenn der Test über einige Wochen läuft.

Wenn keine ältere Seite vorliegt, die für den Test verwendet werden kann, gestaltet sich das Szenario etwas schwieriger. Dann muss eine Domain registriert werden, die danach mindestens ein halbes, besser ein ganzes Jahr, gepflegt und regelmäßig mit neuem Content bestückt wird, sodass sie die Aufmerksamkeit vom Crawler auf sich zieht und bald als gute, vertrauenswürdige Quelle eingestuft wird. Wenn dies erreicht ist, wird wie oben beschrieben der Test fortgeführt und die neuen Seiten werden online gestellt. Danach wird der Duplicate Content eingebunden und die Reaktionen beobachtet. Wenn

also keine ältere Seite vorliegt, dauert der Test wesentlich länger und ist mit mehr Aufwand verbunden. Es ist dann aber wichtig, darauf zu achten, dass die ältere Seite auch wirklich lang genug online ist, bevor die neuen Seiten online gestellt werden und der Duplicate Content eingebunden wird. Wenn nur ein paar Wochen zwischen der Registrierung der jungen und der alten Domain liegen, dürfte es keinen großen Unterschied bei der Reaktion geben.

Szenario #4: Einfluss von verschiedenen Content-Formen

Bei diesem Szenario soll überprüft werden, ob Google auch Duplicate Content in Form von Bildern oder Videos erkennt und ähnlich wie beim Text darauf reagiert. Dazu müssen auch hier zunächst einige Seiten mit Unique Content erstellt werden. Wenn stabile Rankings erreicht sind, wird der Duplicate Content in Form von Bildern oder Videos eingebaut und die Reaktionen beobachtet.

Möglich wäre neun Domains zu wählen und auf das gleiche Keyword zu optimieren. Davon dient dann eine als Vergleichsdomain. Bei zwei Domains werden auf jeder Seite die gleichen Bilder oder Videos aber mit individuellem Text eingebunden (interner Duplicate Content). Zwei weitere Domains verwenden die gleichen Bilder oder Videos wie eine andere Domain, von der sie geklaut werden (externer Duplicate Content). Hinzu kommen noch zwei weitere Domains, die ebenfalls Bilder oder Videos von einer weiteren Domain verwenden, aber beispielsweise nur 50 %. Je nachdem, ob weitere Anteilstufen von Duplicate Content oder andere Content-Formen geprüft werden sollen, kommen entsprechend weitere Seiten hinzu.

Für den Test kann wie im oben beschriebenen Prozessplan ein Zeitraum von insgesamt 12 bis 14 Wochen gewählt werden, abhängig davon wie schnell und ob Google reagiert.

Auf Basis der hier gegebenen Hinweise und Tipps können natürlich weitere Test-Szenarien entwickelt und auch andere Hypothesen geprüft werden. Der beschriebene ideale Test und die Prozesspläne sollen vielmehr als Hilfestellung zur Durchführung eines Testes dienen und zum Weiterdenken anregen.

7 Fazit

Das große Ziel der Bachelorarbeit wurde erfüllt und es konnte mit Hilfe des Website-Testes eine Toleranzgrenze von internem und externem Duplicate Content bei Google identifiziert werden. Diese liegt vermutlich jeweils bei einem Anteil von mehr als 50 %.

Darüber hinaus konnten weitere interessante Erkenntnisse gewonnen werden und vier der fünf Hypothesen wurden bestätigt. So konnte gezeigt werden, dass Duplicate Content in Bezug auf das Ranking in der Suchmaschine Google eine Rolle spielt und es eine Bestrafung in Form einer Herabsetzung des Rankings gibt, wenn sich zu viel Duplicate Content auf einer Seite befindet. Außerdem wurde bewiesen, dass es in Bezug auf das Erreichen der Toleranzgrenze keine Rolle spielt, ob es sich um internen oder externen Duplicate Content handelt, und dass sich aber interner Duplicate Content grundsätzlich stärker auf das Ranking auswirkt als externer. Widerlegt werden konnte die These, dass die Seiten immer um die gleiche Anzahl an Positionen im Ranking zurückgesetzt werden unabhängig davon, wie viel Duplicate Content sich auf der Seite befindet.

Bei einer Wiederholung des Website-Testes könnten aber dennoch einige Sachen verbessert werden. So sollten alle URLs jeweils sofort über die Google Webmaster Tools an den Index gesendet werden. Außerdem ist es im Nachhinein unglücklich, dass beim Keyword „Jahreszeitentechnologie“ keine stabilen Rankings erreicht werden konnten, bevor der Duplicate Content eingebaut wurde. Ein Grund dafür könnten die wahllos gesetzten internen Links sein. Hier sollte bei einer Wiederholung des Testes vielleicht genauer darauf geachtet werden. Ansonsten lief der Test reibungslos und hat seinen Zweck voll erfüllt.

Über den Test hinaus, wurde in Kapitel 2 der Arbeit außerdem ausführlich über die Arbeitsweise einer Suchmaschine und die verschiedenen Rankingfaktoren informiert. Es wurde hier deutlich, welche große Rolle guter Content für das Erreichen einer Topposition im Ranking spielt. Dies kann nicht oft genug hervorgehoben werden, da der Content wahrscheinlich in Zukunft eine noch wichtigere Rolle spielen wird.

Daraufhin wurden in Kapitel 3 die verschiedenen Arten von Duplicate Content vorgestellt und die Sichtweise von Google erläutert. Google steht für Vielfältigkeit und Relevanz in den SERPs, darum stört Duplicate Content hier und soll entfernt werden. Des

Weiteren wurde eine Auswahl an Beispiele von Duplicate Content gezeigt und Lösungsmöglichkeiten präsentiert.

Nach der ausführlichen Erläuterung des Website-Testes mit dem Aufbau der Seiten, den Rahmenbedingungen, der Durchführung des Testes, den Beobachtungen und Ergebnissen, folgten zwei Kapitel, die sich speziell an Webseitenbetreiber bzw. SEO-Agenturen richten. Es wurden zunächst Handlungsempfehlungen und Tipps für Webseitenbetreiber gegeben, die sich leicht in einem Satz zusammenfassen lassen:

Vermeiden Sie Duplicate Content jeglicher Art und erstellen Sie stattdessen individuelle und einzigartige Inhalte mit großem Mehrwert für die Nutzer.

In Kapitel 6 wurde dann ein Framework für Agenturen vorgestellt, damit der Test beispielsweise nach einem neuen Update von Google noch einmal wiederholt werden kann oder neue Aspekte untersucht werden können. Im Rahmen der Bearbeitungszeit der Bachelorarbeit war es leider nicht möglich, alle interessanten Aspekte zu untersuchen. Darum wurde ein idealer Test entwickelt, der Anregungen gibt und zeigt, welche Faktoren noch untersucht werden können, wie zum Beispiel der Einfluss von Backlinks.

Es wäre schön, wenn die Erkenntnisse der Bachelorarbeit zu einem besseren Verständnis von der Funktionsweise von Google und besonders von der Reaktion auf Duplicate Content beitragen würden. Vielleicht tragen sie ja dazu bei, die ewige Diskussion um die Behandlung von Duplicate Content zu beenden oder zumindest in eine andere Richtung zu lenken.

Auch wäre es wünschenswert, wenn in anderen Tests noch einige Aspekte des in Kapitel 6 erläuterten idealen Testes untersucht werden, damit das Bild vervollständigt wird und auch die fehlenden Aspekte, die im Rahmen der Bachelorarbeit nicht untersucht werden konnten, bearbeitet werden.

Literatur

ARONICA 2012

Aronica, Janet: October 2012 Traffic Sources Report : StumbleUpon Share of Traffic Down 53% Since July [online]. In: blog.shareaholic.com – URL: <https://blog.shareaholic.com/2012/11/shareaholic-data-report-pinterest-stumbleupon/> (Abruf: 2013-07-16)

BATELLE 2005

Batelle, John: Google Update “Bourbon” [online]. In: John Batelles’s Search-blog – URL: http://battellemedia.com/archives/2005/06/google_update_bourbon.php (Abruf: 2013-07-16)

BOPP 2011

Bopp, Drazen: Projektmanagement : 5 Projektphasen führen zum Erfolg [online]. In: experto.de – URL: <http://www.experto.de/b2b/unternehmen/unternehmensstrategie/projektmanagement-5-projektphasen-fuehren-zum-erfolg.html> (Abruf: 2013-08-15)

CUTTS 2012

Cutts, Matt: Another step to reward high-quality sites [online]. In: Google Webmaster Central Blog – URL: <http://googlewebmastercentral.blogspot.de/2012/04/another-step-to-reward-high-quality.html> (Abruf: 2013-07-16)

CUTTS 2013

Cutts, Matt: How does required duplicate content (terms and conditions, etc.) affect search [online]?. In: GoogleWebmasterHelp Videos – URL: <http://www.youtube.com/watch?v=Vi-wkEeOKxM> (Abruf: 2013-07-28)

DEAN 2013

Dean, Brian: Google’s 200 Ranking Factors : The Complete List [online]. In: Backlinko – URL: <http://backlinko.com/google-ranking-factors> (Abruf: 2013-07-16)

DE VALK 2013

De Valk, Joost: Duplicate Content : causes and solutions [online]. In: yoast.com – URL: <http://yoast.com/articles/duplicate-content/> (Abruf: 2013-08-12)

DOPICHAJ 2009

Dopichaj, Philipp: Ranking-Verfahren für Web-Suchmaschinen. In: Lewandowski, Dirk (Hrsg.): Handbuch Internet-Suchmaschinen : Nutzerorientierung in Wissenschaft und Praxis. Heidelberg : Akademische Verlagsgesellschaft Aka GmbH, 2009. S. 101-115

DZIKI 2010

Dziki, Julian: Duplicate Content : Schluss damit [online]! In: SEOKRATIE – URL: <http://www.seokratie.de/duplicate-content-schluss-damit/> (Abruf: 2013-07-28)

DZIKI 2012

Dziki, Julian: Google Richtlinien für Webmaster aktualisiert [online]. In: SEOKRATIE – URL: <http://www.seokratie.de/google-richtlinien-fur-webmaster-aktualisiert/> (Abruf: 2013-07-16)

ENGE 2010

Enge, Eric: Four Ways To Safely Syndicate Content [online]. In: SearchEngineLand – URL: <http://searchengineland.com/four-ways-to-safely-syndicate-content-32483> (Abruf: 2013-08-12)

E-TEACHING.ORG 2008

E-teaching.org: Was ist ein Experiment [online]? In: e-teaching.org – URL: <http://www.e-teaching.org/didaktik/qualitaet/experiment/> (Abruf: 2013-07-24)

FREES / VAN EIMEREN 2012

Frees, Beate ; Van Eimeren, Birgit: 76 Prozent der Deutschen online : neue Nutzungssituationen durch mobile Endgeräte [online]. In: Media Perspektiven – URL: http://www.ard-zdf-onlinestudie.de/fileadmin/Online12/0708-2012_Eimeren_Frees.pdf (Abruf: 2013-07-16)

GOOGLE I 2013

Google: Richtlinien für Webmaster [online]. In: Google Webmaster-Tools Hilfe – URL: <https://support.google.com/webmasters/answer/35769?hl=de> (Abruf: 2013-07-16)

GOOGLE II 2013

Google: Content duplizieren [online]. In: Google Webmaster-Tools Hilfe – URL: <https://support.google.com/webmasters/answer/66359?hl=de> (Abruf: 2013-08-12)

GOOGLE III 2013

Google: Scraped Content [online]. In: Google Webmaster-Tools Hilfe – URL: <https://support.google.com/webmasters/answer/2721312?hl=en> (Abruf: 2013-08-12)

GOOGLE IV 2013

Google: Thin content with little or no added value [online]. In: Google Webmaster-Tools Hilfe – URL: <https://support.google.com/webmasters/answer/2604719> (Abruf: 2013-08-12)

GOOGLE V 2013

Google: Kanonisierung [online]. In: Google Webmaster-Tools Hilfe – URL: https://support.google.com/webmasters/answer/139066?hl=de&ref_topic=2371375 (Abruf: 2013-08-12)

GOOGLE VI 2013

Google: Über rel="canonical" [online]. In: Google Webmaster-Tools Hilfe – URL: <https://support.google.com/webmasters/answer/139394?hl=de> (Abruf: 2013-08-12)

GOOGLE VII 2013

Google: 301-Weiterleitungen [online]. In: Google Webmaster-Tools Hilfe – URL: <https://support.google.com/webmasters/answer/93633?hl=de> (Abruf: 2013-08-12)

GOOGLE VIII 2013

Google: Seitennummerierung [online]. In: Google Webmaster-Tools Hilfe – URL: <https://support.google.com/webmasters/answer/1663744?hl=de> (Abruf: 2013-08-12)

GORNIAK et al. 2010

Górniak, M. ; Petzoldt, J. ; Schäfer, K. ; Weßels, N.: Abhängige Variable [online]. In: TU Dresden – URL: <http://elearning.tu-dresden.de/versuchsplanung/e35/e287/e590> (Abruf: 2013-07-24)

GORNIAK et al. 2013

Górniak, M. ; Petzoldt, J. ; Schäfer, K. ; Weßels, N.: Störvariable [online]. In: TU Dresden – URL: <http://elearning.tu-dresden.de/versuchsplanung/e35/e287/e603/> (Abruf: 2013-07-24)

HARGRAVE 2010

Hargrave, John: Why Duplicate Content Is Good For You [online]. In: Search Engine Journal – URL: <http://www.searchenginejournal.com/why-duplicate-content-is-good-for-you/22270/> (Abruf: 2013-07-28)

JANSSEN 2013

Janssen, Cory: Content Scraping [online]. In: technopedia – URL: <http://www.techopedia.com/definition/27564/content-scraping> (Abruf: 2013-08-12)

KARAVUL / VOIGT 2011

Karavul, Berekat ; Voigt, Dirk: Projektphasen und Meilensteine [online]. In: Projektmanagement Handbuch – URL: <http://www.projektmanagementhandbuch.de/projektplanung/projektphasen-und-meilensteine/> (Abruf: 2013-08-15)

KUBITZ I 2012

Kubitz, Eric: Die größten SEO-Fehler der Welt [online]. In: SEO BOOK – URL: <http://www.seo-book.de/allgemein/auswertung-2012-die-grosten-seo-fehler-der-welt> (Abruf: 2013-08-12)

KUBITZ II 2012

Kubitz, Eric: Aktuelle Übersicht : das Canonical Tag von Google [online]. In: SEO BOOK – URL: <http://www.seo-book.de/onpage/ubersicht-das-canonical-tag-von-google> (Abruf: 2013-08-12).

LANDRY 2011

Landry, Tommy: An SEO Experiment Gone Awry (Mea Culpa) [online]. In: Return on Now – URL: <http://returnnonnow.com/2011/06/seo-experiment-awry-mea-culpa/> (Abruf: 2013-07-28)

LETSPROMOTEIT.DE 2007

Letspromoteit.de: Artikelverzeichnis : Der Duplicate Content Test [online]. In: Lets Promote It – URL: <http://www.letspromoteit.de/web-marketing/artikelverzeichnis-der-duplicate-content-test> (Abruf: 2013-07-28)

LEWANDOWSKI 2005

Lewandowski, Dirk: Web Information Retrieval : Technologien zur Informationssuche im Internet. Frankfurt am Main : DGI, 2005

LEWANDOWSKI 2013

Lewandowski, Dirk: Suchmaschinen. In: Kuhlen, Rainer ; Semar, Wolfgang ; Strauch, Dietmar (Hrsg.): *Grundlagen der praktischen Information und Dokumentation*. Berlin : de Gruyter, 2013. S. 494-508

LIENERT / RAATZ 1998

Lienert, Gustav Adolf ; Raatz, Ulrich: Testaufbau und Testanalyse. 6. Auflage, Weinheim : Psychologie Verlags Union, 1998

LINKRESEARCHTOOLS 2013

LinkResearchTools: Keyword Classification [online]. In: LinkResearchTools – URL: <http://www.linkresearchtools.com/tools/cla/keyword-classification/> (Abruf: 2013-08-15)

MARSHALL 2011

Marshall, Paul: Is Syndicated Content Duplicate Content [online]? In: Search Engine Journal – URL: <http://www.searchenginejournal.com/syndicated-content-duplicate-content/28833/> (Abruf: 2013-08-12)

MCGONAGLE I 2009

McGonagle, Daniel: Duplicate Content Penalty Myths and Some Tests I'm Running [online]. In: Linkvana Reviews – URL: <http://linkvanareviews.com/duplicate-content-penalty-myths-and-some-tests-im-running> (Abruf: 2013-07-28)

MCGONAGLE II 2009

McGonagle, Daniel: Duplicate Content Penalty Test + quick update [online]. In: Linkvana Reviews – URL: <http://linkvanareviews.com/duplicate-content-penalty-test-quick-update> (Abruf: 2013-07-28)

MEYERS I 2011

Meyers, Peter J.: Duplicate Content in a Post-Panda World [online]. In: Moz.com – URL: <http://moz.com/blog/duplicate-content-in-a-post-panda-world> (Abruf: 2013-08-12)

MEYERS II 2011

Meyers, Peter J.: 6 Extreme Canonical Tricks [online]. In: Moz.com – URL: <http://moz.com/blog/6-extreme-canonical-tricks> (Abruf: 2013-07-28)

MISSFELDT 2013

Mißfeldt, Martin: Duplicate Content Penalty nur ein Seo-Gespens [online]?! In: tagSeoBlog – URL: <http://www.tagseoblog.de/duplicate-content-penalty-nur-ein-seo-gespenst> (Abruf: 2013-07-28)

MOSKWA 2008

Moskwa, Susan: Die “Duplicate Content Penalty” – entmystifiziert [online]! In: Google Blog Webmaster-Zentrale – URL: <http://googlewebmastercentral-de.blogspot.de/2008/09/die-duplicate-content-penalty.html> (Abruf: 2013-08-12)

MOZ 2013

Moz: Google Algorithm Change History [online]. In: MOZ.com – URL: <http://moz.com/google-algorithm-change> (Abruf: 2013-07-16)

NETMARK 2013

Netmark: Google Ranking Factors [online]. In: Netmark – URL: <http://www.netmark.com/google-ranking-factors> (Abruf: 2013-07-16)

ONPAGE.ORG 2013

Onpage.org: Duplicate Content [online]. In: OnPageWiki – URL: http://de.onpage.org/wiki/Duplicate_Content (Abruf: 2013-08-12)

PACHER 2011

Pacher, Jörg: Bessere Qualität in den Suchergebnissen [online]. In: Google Blog Webmaster-Zentrale – URL: <http://googlewebmastercentral-de.blogspot.de/2011/03/bessere-qualitat-in-den-suchergebnissen.html> (Abruf: 2013-07-16)

PETROVIC 2012

Petrovic, Dan: How I Hijacked Rand Fishkin’s Blog [online]. In: DEJANSEO – URL: <http://dejanseo.com.au/hijacked/> (Abruf: 2013-07-28)

RIXECKER 2013

Rixecker, Kim: SEO : Die wichtigsten Ranking-Faktoren 2013 [online]. In: t3n – URL: <http://t3n.de/news/wichtigsten-ranking-faktoren-2013-479404/> (Abruf: 2013-07-16)

SCHMIDT-VOIGT 2012

Schmidt-Voigt, Nikolas: Zuviel Inhalt für eine Seite : Paginierung für SEO [online]. In: Kawumba – URL: <http://www.kawumba.de/zuviel-inhalt-fuer-eine-seite-paginierung-fuer-seo/> (Abruf: 2013-08-12)

SCHWARTZ I 2013

Schwartz, Barry: Google's Panda Dance : Matt Cutts Confirms Panda Rolls Out Monthly Over 10 Of 30 Days [online]. In: Search Engine Land – URL: <http://searchengineland.com/googles-panda-dance-matt-cutts-confirms-panda-rolls-out-monthly-over-10-of-30-days-162950> (Abruf: 2013-07-16)

SCHWARTZ II 2013

Schwartz, Barry: Google's Cutts Says Don't Worry About Duplicate Content But SEOs Don't Trust [online]. In: seroundtable.com – URL: <http://www.seroundtable.com/google-duplicate-content-dont-worry-17115.html> (Abruf: 2013-07-28)

SCOTT 2013

Scott, Allan: 5 häufige Fehler bei der Verwendung von rel=canonical-Tags [online]. In: Google Blog Webmaster-Zentrale – URL: <http://googlewebmastercentral-de.blogspot.de/2013/04/5-haeufige-rel-canonical-fehler.html> (Abruf: 2013-08-12)

SEARCHMETRICS 2012

Searchmetrics: SEO Ranking Faktoren für Google Deutschland 2012 [online]. In: Searchmetrics – URL: <http://www.searchmetrics.com/de/services/whitepaper/seo-ranking-faktoren-deutschland/> (Abruf: 2013-07-16)

SEARCHMETRICS 2013

Searchmetrics: SEO Ranking-Faktoren 2013 für Google Deutschland [online]. In: Searchmetrics – URL: <http://www.searchmetrics.com/de/services/ranking-faktoren-2013/> (Abruf: 2013-07-16)

SEEFELD 2012

Seefeld, Martina: Google's Algorithm Changes [online]. In: t3n.de – URL: http://t3n.de/news/wp-content/uploads/2012/03/seo_google_suchalgorithmus.jpg (Abruf: 2013-07-16)

SEO-UNITED I 2013

SEO-united: Suchmaschinen im Juli 2013 [online]. In: SEO-united.de – URL: <http://www.seo-united.de/suchmaschinen.html> (Abruf: 2013-07-16)

SEO-UNITED II 2013

SEO-united: Content Syndication [online]. In: SEO-united.de – URL: <http://www.seo-united.de/glossar/content-syndication/> (Abruf: 2013-08-12)

SEOUPDATES.INFO 2013

Seoupdates.info: Duplicate Content Checker Tools – Free and Paid [online]. In: seoupdates.info – URL: <http://seoupdates.info/duplicate-content-checker-tools-free-and-paid/> (Abruf: 2013-08-12)

SINGHAL 2011

Singhal, Amit: More guidance on building high-quality sites [online]. In: Google Webmaster Central Blog – URL: <http://googlewebmastercentral.blogspot.de/2011/05/more-guidance-on-building-high-quality.html> (Abruf: 2013-07-16)

SISTRIX I 2013

Sistrix: Google Panda Update [online]. In: Frag Sistrix – URL: <http://www.sistrix.de/frag-sistrix/google-algorithmus-aenderungen/google-panda-update/> (Abruf: 2013-07-16)

SISTRIX II 2013

Sistrix: Google Penguin Update [online]. In: Frag Sistrix – URL: <http://www.sistrix.de/frag-sistrix/google-algorithmus-aenderungen/google-penguin-update/> (Abruf: 2013-07-16)

SISTRIX III 2013

Sistrix: Duplicate Content [online]. In: Frag Sistrix – URL: <http://www.sistrix.de/frag-sistrix/onpage-optimierung/duplicate-content/> (Abruf: 2013-08-12)

STABENOW 2013

Stabenow, Sandra: SEO für internationale Websites : hreflang [online]. In: SEO Trainee – URL: <http://www.seo-trainee.de/seo-fuer-internationale-websites-hreflang/> (Abruf: 2013-08-12)

SULLIVAN 2011

Sullivan, Danny: Why Google Panda Is More A Ranking Factor Than Algorithm Update [online]. In: Search Engine Land – URL: <http://searchengineland.com/why-google-panda-is-more-a-ranking-factor-than-algorithm-update-82564> (Abruf: 2013-07-16)

TROEBEL I 2008

Troebel, Frank: Duplicate Content ein Feldversuch [online]. In: Böser Seo – URL: http://boeser-seo.blog.de/2008/01/13/duplicate_content_ein_feldversuch~3568461/ (Abruf: 2013-07-28)

TROEBEL II 2008

Troebel, Frank: Update Gegrabte Domain und Duplicate Content Feldversuch [online]. In: Böser Seo – URL: http://boeser-seo.blog.de/2008/02/22/update_gegrabte_domain_und_duplicate_co~3764524/ (Abruf: 2013-07-28)

TROEBEL III 2008

Troebel, Frank: Update Gegrabte Domain und Duplicate Content Feldversuch [online]. In: Böser Seo – URL: <http://boeser-seo.blog.de/2008/04/01/update-gegrabte-domain-und-duplicate-co-3981092/> (Abruf: 2013-07-28)

TUKNOV 2013

Tuknov, Mikhail: Top 31 SEO Experiments You Want To Know About [online]. In: Search Engine Journal – URL: <http://www.searchenginejournal.com/top-31-seo-experiment%E2%80%8Bs-you-want-to-know-about/63855/> (Abruf: 2013-07-28)

Anhänge

Übersicht der anfrageabhängigen und anfrageunabhängigen Rankingfaktoren aus Kapitel 2.1:

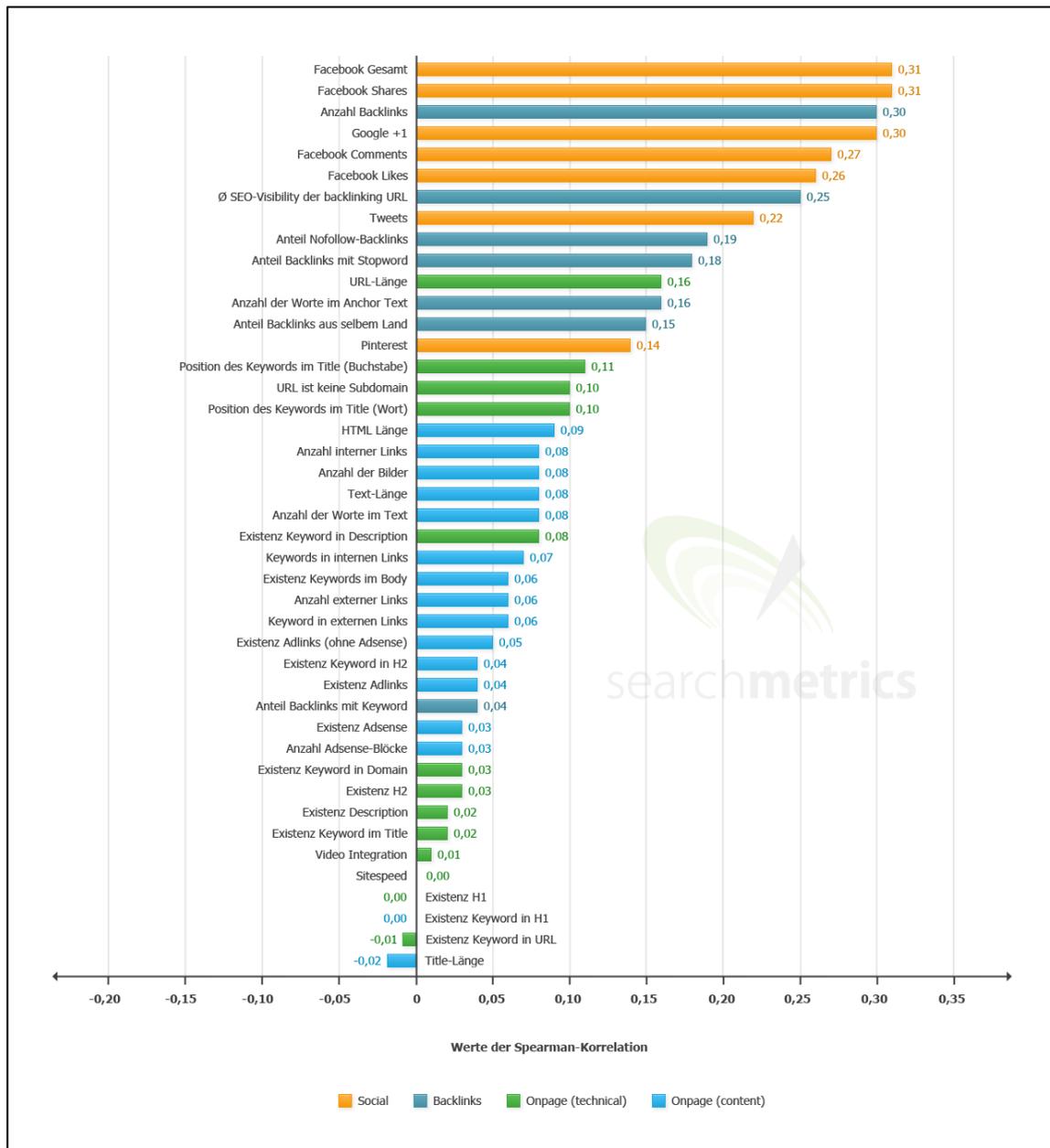
Kriterium	Erläuterung
Dokumentspezifische Wortgewichtung (WDF)	Relative Häufigkeit des Vorkommens eines Worts in einem Dokument.
Wortabstand	Bei Anfragen mit mehreren Suchbegriffen wird der Abstand der Suchbegriffe voneinander berücksichtigt.
Position der Suchbegriffe	An markanten Stellen des Dokuments vorkommende Suchbegriffe werden höher bewertet. Zum Beispiel Vorkommen im Titel, in den Überschriften, in der URL.
Reihenfolge der Suchbegriffe in der Anfrage	In der Anfrage zuerst stehende Begriffe werden als bedeutender angesehen.
Metatags	Vorkommen der Suchbegriffe in den Metatags
Stellung der Suchbegriffe innerhalb des Dokuments	Vorkommen der Suchbegriffe am Beginn des Dokuments wird höher gewertet als späteres Auftreten.
Betonung von Begriffen durch HTML-Elemente	Hervorgehobene Begriffe (fett, kursiv) werden höher bewertet.
Groß-/Kleinschreibung	Dokumente, in denen die Suchbegriffe in exakt der eingegebenen Form vorkommen, werden bevorzugt.
Inverse Dokumenthäufigkeit (IDF)	Relative Häufigkeit des Vorkommens eines Wortes in Dokumenten der gesamten Datenbank; seltene Begriffe werden bevorzugt.
Ankertext	Vorkommen der Suchbegriffe im Linktext eines Dokuments, welches auf das Zieldokument verweist.
Sprache	Dokumente, die in der Sprache des benutzten Länderinterfaces verfasst sind, werden höher bewertet.
Geo-Targeting	Seiten, die ihren „Standort“ in der Nähe des Benutzers haben, werden bevorzugt.

Anhang 01: Anfrageabhängige Faktoren im Ranking einer Web-Suchmaschine (Quelle: LEWANDOWSKI 2005, S. 91)

Kriterium	Erläuterung
Verzeichnisebene	Je höher das Dokument innerhalb der Hierarchie seiner Website steht, desto höher wird es bewertet.
Anzahl eingehende Links	Je mehr Links auf das Dokument verweisen, als desto bedeutender wird es angesehen.
Linkpopularität	Wert für die Autorität / Qualität eines Dokuments wird aufgrund der Verlinkungsstruktur berechnet.
Klickhäufigkeit	Dokumente, die von vielen Benutzern angesehen werden, werden höher bewertet.
Aktualität	Aktuelle Dokumente werden höher bewertet als ältere.
Dokumentlänge	Dokumente ab und bis zu einer gewissen Länge (sinntugend) werden bevorzugt.
Dateiformat	Dokumente im Standardformat HTML werden höher bewertet als solche in anderen Formaten (PDF, Word, usw.)
Größe der Site	Dokumente von umfangreichen Web-Angeboten werden höher bewertet als solche von kleinen Sites.

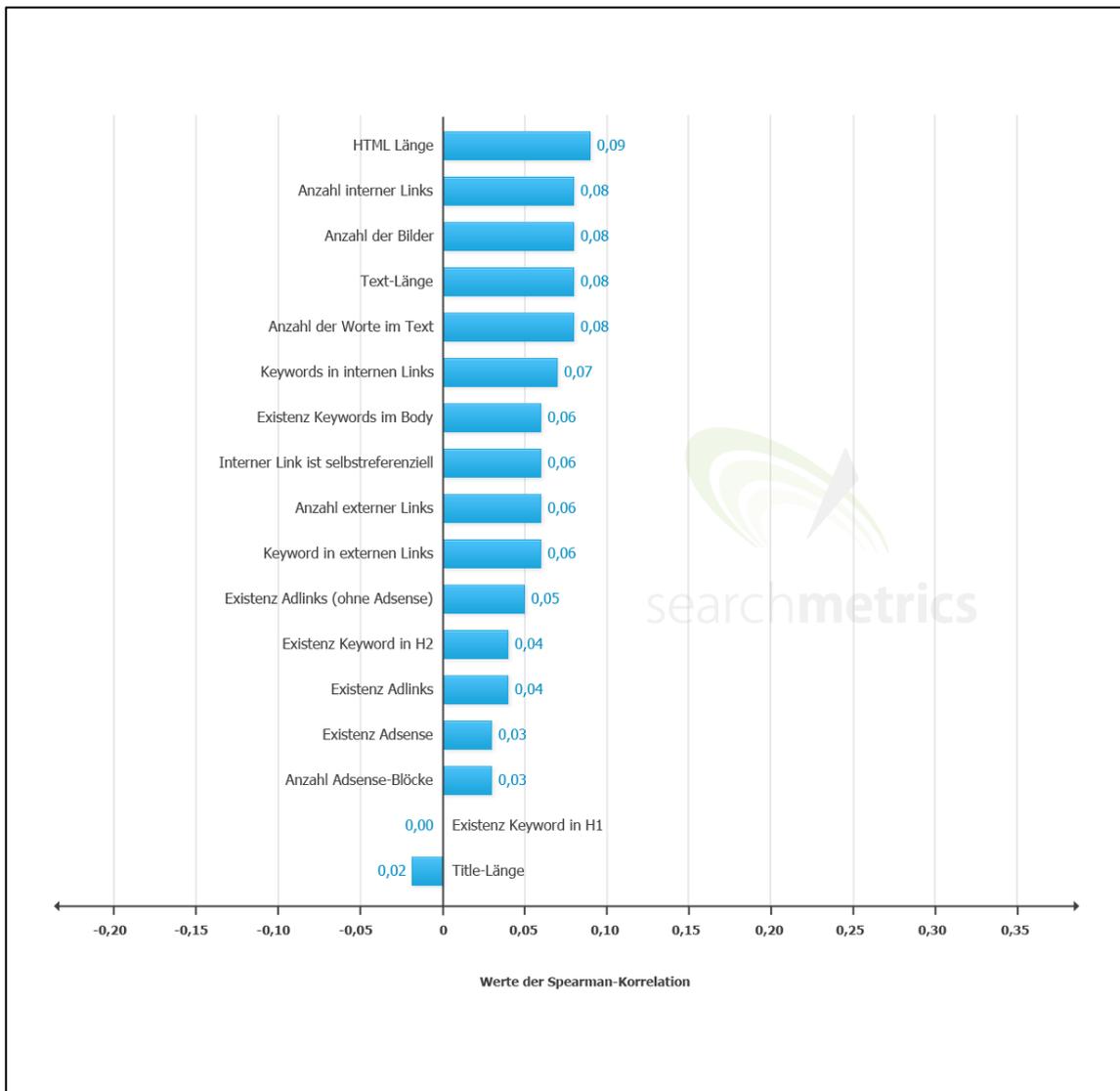
Anhang 02: Anfrageunabhängige Faktoren im Ranking einer Web-Suchmaschine (Quelle: LEWANDOWSKI 2005, S. 94)

Ergebnisse der Analyse von Searchmetrics zu den Google Rankingfaktoren 2013 als Spearman-Korrelation zu Kapitel 2.1:



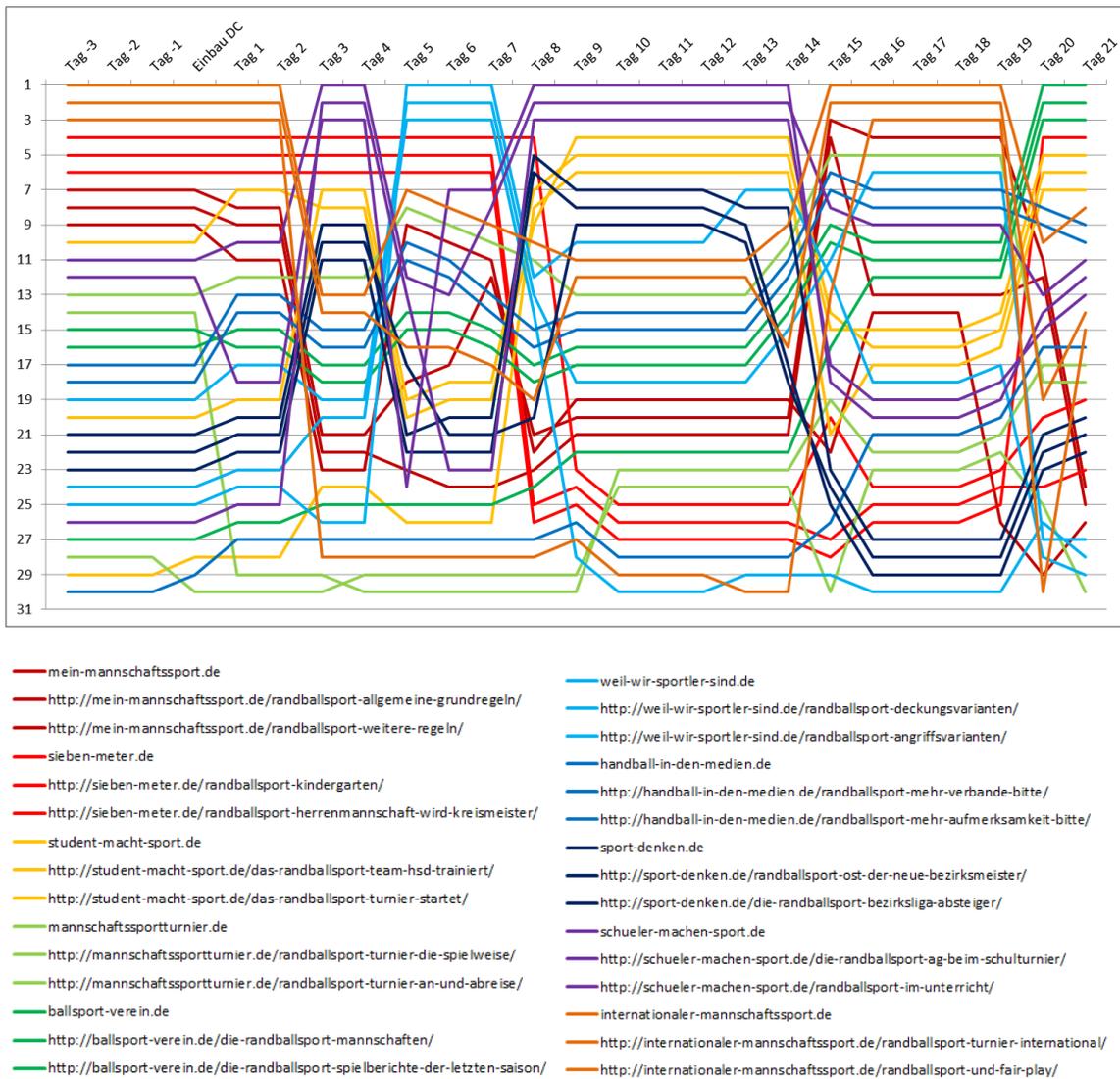
Anhang 03: Searchmetrics Rankingfaktoren 2013 (Quelle: SEARCHMETRICS 2013)

Übersicht der in der Korrelationsanalyse von Searchmetrics 2013 untersuchten Content-Faktoren zu Kapitel 2.2:



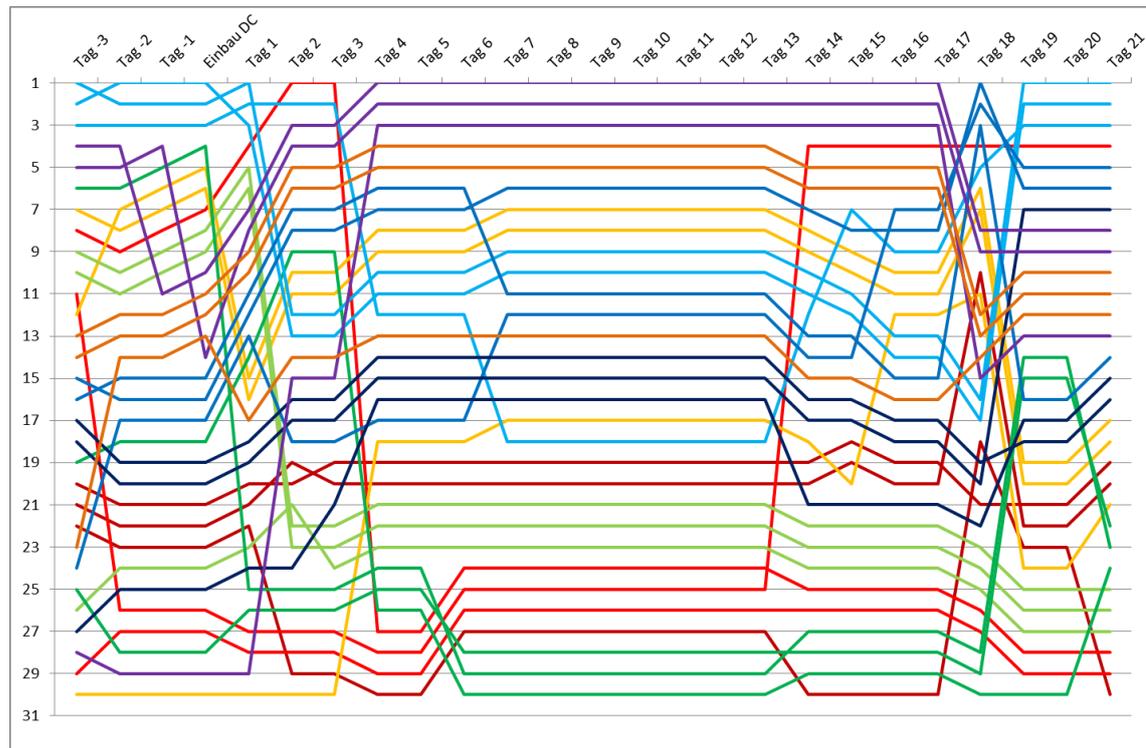
Anhang 04: Korrelation der Content-Faktoren 2013 (Quelle: SEARCHMETRICS 2013)

Entwicklung des Rankings beim Keyword „Randballsport“ nach dem Einbau des Duplicate Contents zu Kapitel 4.6:



Anhang 05: Entwicklung des Rankings im Zeitraum von 21 Tagen nach dem Einbau des Duplicate Contents beim Keyword „Randballsport“

Entwicklung des Rankings beim Keyword „Jahreszeitentechnologie“ nach dem Einbau des Duplicate Contents zu Kapitel 4.6:



- jahreszeitenkraftwerk.de
- <http://jahreszeitenkraftwerk.de/jahreszeitentechnologie-im-herbst/>
- <http://jahreszeitenkraftwerk.de/jahreszeitentechnologie-im-fruehling/>
- mein-voller-kleiderschrank.de
- <http://mein-voller-kleiderschrank.de/jahreszeitentechnologie-die-erfinder/>
- <http://mein-voller-kleiderschrank.de/jahreszeitentechnologie-die-vorteile/>
- draussen-schlafen.de
- <http://draussen-schlafen.de/jahreszeitentechnologie-im-test/>
- <http://draussen-schlafen.de/jahreszeitentechnologie-im-schlafsack-sehr-zu-empfehlen/>
- wetterbeeinflussung.de
- <http://wetterbeeinflussung.de/jahreszeitentechnologie-im-einsatz/>
- <http://wetterbeeinflussung.de/jahreszeitentechnologie-mit-problemen/>
- temperatur-im-auto.de
- <http://temperatur-im-auto.de/jahreszeitentechnologie-nicht-nur-fur-autos/>
- <http://temperatur-im-auto.de/jahreszeitentechnologie-im-test/>
- technik-im-auto.de
- <http://technik-im-auto.de/die-jahreszeitentechnologie-in-der-verwendung/>
- <http://technik-im-auto.de/jahreszeitentechnologie-expertenmeinung/>
- jahresmagazin.de
- <http://jahresmagazin.de/jahreszeitentechnologie-unterwegs-zu-besuch-im-jahreszeitenkraftwerk/>
- <http://jahresmagazin.de/jahreszeitentechnologie-aktiv-der-neue-sleeping-bag-20/>
- terrarium-technik.de
- <http://terrarium-technik.de/jahreszeitentechnologie-einbau-und-moeglichkeiten/>
- <http://terrarium-technik.de/jahreszeitentechnologie-fur-schildkroten/>
- jahresbauer.de
- <http://jahresbauer.de/jahreszeitentechnologie-und-die-vorteile/>
- <http://jahresbauer.de/jahreszeitentechnologie-bei-biobauer-hans/>
- handballspray.de
- <http://handballspray.de/jahreszeitentechnologie-im-gebrauch/>
- <http://handballspray.de/jahreszeitentechnologie-bei-den-stars/>

Anhang 06: Entwicklung des Rankings im Zeitraum von 21 Tagen nach dem Einbau des Duplicate Contents beim Keyword „Jahreszeitentechnologie“

Eigenständigkeitserklärung

Ich versichere, die vorliegende Arbeit selbstständig ohne fremde Hilfe verfasst und keine anderen Quellen und Hilfsmittel als die angegebenen benutzt zu haben. Die aus anderen Werken wörtlich entnommenen Stellen oder dem Sinn nach entlehnten Passagen sind durch Quellenangabe kenntlich gemacht.

Hamburg, den 29.08.2013

Johanna Hartje