

UNDER STANDING SEO

BY FRANZ ENZENHOFER



Copyright © 2018 by Franz Enzenhofer.

All rights reserved. No part of this book may be reproduced or transmitted in any form without permission from the author (contact Franz at fe@f19n.com).

Published by Franz Enzenhofer, Full Stack Optimization (Vienna, Austria) - 3rd Edition (October 2018). Edited by Majella Grawatsch (majella.grawatsch@gmail.com). Printed and bound by Michael Schalk Gesellschaft m.b.H. in Pottendorf, Austria.

ISBN: 978-3-200-05426-4

ABOUT THE AUTHOR

Franz Enzenhofer is a man of few words. This book contains 35,245 of his. He skipped the parts that nobody reads.

Franz Enzenhofer is the masterchief over at www.fullstackoptimization.com, the SEO company with the goal of making all SEO companies – including themselves – redundant.

Franz Enzenhofer refuses to speak about himself in the third person.

ABOUT THIS BOOK

I'm trying to pull a Wittgenstein here, and by that, I mean writing the definitive book on a topic and then declaring the matter solved. Similar to Wittgenstein, I will fail, however while doing so, I will enhance the status quo of an entire discipline: Search Engine Optimization - SEO.

The primary goal of this book is to deliver a framework - that is, a way to think about SEO. If you have online experience, whether you are a manager, developer, editor, product owner, or hacker, you should be able to read and comprehend this book. You should then be able to use what you have learned to make informed SEO decisions for your business. Be aware this book does not waste space replicating any technical specs (those you can easily google yourself), but it will explain what you need to do, when you need to do it, and why you need to do it. Also, this book is about SEO, organic search engine optimization, and not about SEA, Google Ads, AdWords or PPC.

This book is heavy on theory and light on examples. It is up to you to use your imagination, your experience, and your initiative to combine it all in a way that is suitable for your company.

If you are an SEO manager within an SEO agency, then this book is not for you. You are part of the problem, not part of the solution. This book exists to make you redundant.

About the third edition

As the first and second edition were already perfect, the third edition differs only in some clarifications and spelling and punctuation fixes.

Table of Contents

CHAPTER 1: WHAT IS SEO?	11
- What SEO Is Not!.....	11
- What SEO is!.....	14
- “Tell me the latest SEO trick!?!”	15
- R.T.F.M.....	15
- Be Sure To Get Feedback!	17
CHAPTER 2: SEO SETUP	19
- The Goal	19
- Key SEO Areas	21
- SEO Point Of View	25
- SEO Setup Recommendation.....	25
- Common SEO Setup Mistakes	27
CHAPTER 3: SYSTEMATIC SEO	29
- The User	29
- The Webproperty	30
- The User And The Webproperty	31
- The User And Google.....	32
- Distribution	35
- Traffic.....	36
- The User Experience Feedback Loop.....	37
- The Brand Experience Feedback Loop	39
- Targeting	40
- Marketing.....	42
- Inlinks.....	44
- Link Building.....	44
- Brand Building	46
- What You Need To Focus On.....	48
- Feedback	50
- It’s A Model!	53
- Common Mistake: Reverse Engineering Google.....	54

CHAPTER 4: THE SEO PROJECT57

- User Experience..... 57
- A Red Line 58
- Speed 59
- URL Rules 61
- URL Best Practices (Make Your Life Easy)..... 65
- Targeted Page Generation Logic (a.k.a. Targeting)..... 65
- The Quality Criteria (QC) 67
- Editorial Search Demand Research Targeting Page Creation..... 69
- Database Driven Targeting Page Generation..... 70
- Database Crossing Targeting 71
- Query-Driven Targeting 72
- Database Editorial Targeting..... 74
- Retroactive Targeting 75
- Targeted Pages..... 76
- Best Practices For Targeted Pages..... 83

CHAPTER 5: DISTRIBUTION103

- Sitemap Stacking 103
- Sitemap Notifying & robots.txt..... 108
- Private Sitemaps 111
- Getting Crawled And Indexed 112
- Updating Sitemaps 115
- Distribution Quality Assurance 116
- The Startpage, Targeting, And Linking 117

CHAPTER 6: DEADWEIGHT 119

- Edge Cases..... 121

CHAPTER 7: ONGOING SEO PROCESSES123

- Business Case #1: How do we create x% more targeted pages? 124
- Business Case #2: How do we update x% of our targeted pages?..... 127
- Business Case #3: How do we get other sites to point x% more links to our targeted pages? 129

- Business Case #4: How do we ensure that the user experience of our site gets better?..... 135
- Business Case #5: How do we ensure that our site gets more brand recognition? 136
- Ongoing Feedback 137

CHAPTER 8: OTHER THINGS TO THINK ABOUT.....139

- Mobile 139
- Render As Googlebot And Other Tools..... 140
- App Indexing..... 140
- AMP - Accelerated Mobile Pages..... 141
- Multiple Domains And Languages 142
- Going Multidomain!..... 143

SEO Success.....145

CHAPTER 1: WHAT IS SEO?

Before we begin, it is important to understand what SEO is. I define SEO as:

- the business of getting found
- the business of getting found online
- all business activities that contribute to the goal of getting found online

I class the first point above as a slogan, the second one as a clarification of the slogan, with the third being the proper definition. Although the definition sounds rather simple, there are many common misconceptions about SEO. Let's put an end to these misconceptions! We will do this by firstly discussing what SEO isn't, then what it is, followed by the two most important things you should take away from this book.

What SEO Is Not!

1. SEO is not free traffic

SEO does not come for "free." You need to invest time, money and resources! You must have a goal in mind - do you want more traffic/users/clients/customers? To achieve this, you must do things differently. Very differently to how you would do things if SEO did not exist.

2. SEO is not PageRank™

PageRank™ was the original algorithm developed by Larry Page and Sergey Brin, the Google founders, to rank web pages according to their relevance. Their most important principles were:

- pages with more inlinks (also known as backlinks) are more valuable than pages with fewer inlinks, and
- links from pages with more inlinks are more valuable than links from pages with fewer inlinks.

That was in 1998 – that's multiple lifetimes ago concerning the internet. Since then, the term has been used and misused in every possible way until it became an esoteric concept that can be used to prevent or encourage each and every action you see fit. Want to do something? PageRank™! Don't want to do something? PageRank™! Why?

PageRank™! PageRank™ is thought-cancer and a frustrating distraction from our task at hand, so we will not use it. Just google [pagerank thought-cancer] if you want to hear more about it because I could really go on for hours.

Important Note: Whenever a term is within [square brackets], it is a user search query, so something you or any other user types (or sometimes speaks) into Google. It's the same annotation that Google uses in their documentation. The square brackets symbolize the Google Search input field.

3. SEO is not a technology

The internet is a digital medium, and therefore everything on the internet is to some extent technology. However, SEO is not a technological challenge and cannot be “fixed” with technology alone.

4. SEO is not software

You can't download SEO. While some content management systems (CMS), server software, website frameworks and shop systems are, from a technical point of view, more search engine optimized than others, let me make it very clear that you cannot download SEO.

5. SEO is not hardware

Although it is not hardware, SEO does have a hardware dependency. Let me explain: if your website is slow because your server is slow, then SEO is pointless. However, it is also important to note that even the best hardware will not make your site search engine optimized.

6. SEO is not rankings

With modern search engines, there is no consistent “No. 1” ranking in search engine results. Anyone who claims to deliver this is misleading. Users see different results for the same search terms, which vary based on factors such as their device, location, internet connection speed, individual search history, serendipity and more. Your goal cannot be the mere ranking itself. Above the fold rankings (the ones that are visible on a web page without scrolling) are a side effect of successful SEO, but not a goal in their own right.

7. SEO is not a service

You cannot outsource SEO. You can outsource certain tasks, but SEO in its entirety is and always will be the responsibility of your business.

8. SEO is not a tool

Most third party (non-Google) SEO tools (especially all paid SEO tools) deliver more or less made up figures about SEO, so-called “magic numbers.” None of them will do SEO for you, and none of them will lead to SEO success without your effort. Most of the time you will make poor business decisions because of them. Don’t use them.

9. SEO is not a product

You cannot buy SEO.

10. SEO is not a solution

SEO is not the easy solution to wider business problems. If your product, service or anything you want to make money with is shit, then SEO will not save you.

11. SEO is not mere “links”

Although the original PageRank™ algorithm was based on the importance of inlinks, a current ranking is substantially more complex. However, sadly inlinks are still a relevant factor in search, and although they are just a tiny part of what you have to care about, they are not - as it is sometimes claimed - everything.

12. SEO is not a trick

Tricks don’t scale. SEO does.

13. SEO is not art

If you are reading this book, then you are doing so because you want to achieve a goal involving SEO. You will be required to execute tasks, you will check for feedback, and you will iterate. Art is not goal oriented, SEO is. Therefore, SEO is not art.

14. SEO is not a problem

SEO is solved! You just need to do it.

What SEO is!

1. SEO is a business

You will need to do things differently. You will invest time, money, brain share, development resources, content writing, marketing effort and more into this topic. You will need to implement new processes that will change how your company does business.

2. SEO is the business of getting found on the internet

Here's the thing: if you don't want to, you needn't change a thing. But your business won't be found as well as it could be on the internet. SEO is a choice! The choice is yours. It's not the user's choice, nor Google's, nor is it the choice of an overpaid SEO agency or your competition, and it's certainly not the choice of this book. If you don't care about getting found, then there is no need for SEO. It must make sense to you that your users find you. You must care about it. If you don't have a product or don't have a business goal, then create this first. Then, and only then, should you consider SEO. SEO involves all business activities that contribute to the goal of getting found online.

3. SEO involves all business activities that contribute to the goal of getting found online

SEO acts within these areas:

- development
- marketing
- content

You will find out more (much more) about them in the SEO setup chapter to come.

4. SEO is solved!

SEO, like anything that has value, is 5% know-how and 95% execution. What you need to do is clear. You "just" have to do it. In nearly all cases, SEO does not fail because of too little knowledge, but because of flawed or non-existent execution. Knowledge without execution is meaningless for businesses. *Focus on execution!*

5. SEO makes sense!

SEO makes sense, and everybody can and should understand it. So if you encounter

something that claims it is “SEO”, but does not make sense, don’t do it. I use the term “sensemaking” in this book, so in other words, something has to be sensible, rational or logical.

6. SEO is a choice!

There are thousands of good and valid reasons not to do things the search engine optimized way. You can have the best, coolest, most awesome, perfect website without SEO! You can make all the right technical choices, you can create great content and have an effective marketing approach, but the result might not be search engine optimized in a way that leads to sustainable, sensemaking and growing traffic. Only a search engine optimized website is search engine optimized.

“Tell me the latest SEO trick!?”

People often ask me for the most recent SEO trick. And I say to them: “SEO is a business!” and I recommend them one of my excellent SEO workshops or this awesome book or both. However, there is one strategy (with two action points) that is more important than anything else you could do for your business SEO wise. If you do these two things, you will probably be miles ahead of your competition, and you will prevent a lot of frustration, lost time and wasted money. In fact, if you just take two things away from this book, let it be these:

1. R.T.F.M. (Read The F*cking Manual!)
2. if you do something, be sure to get feedback.

R.T.F.M.

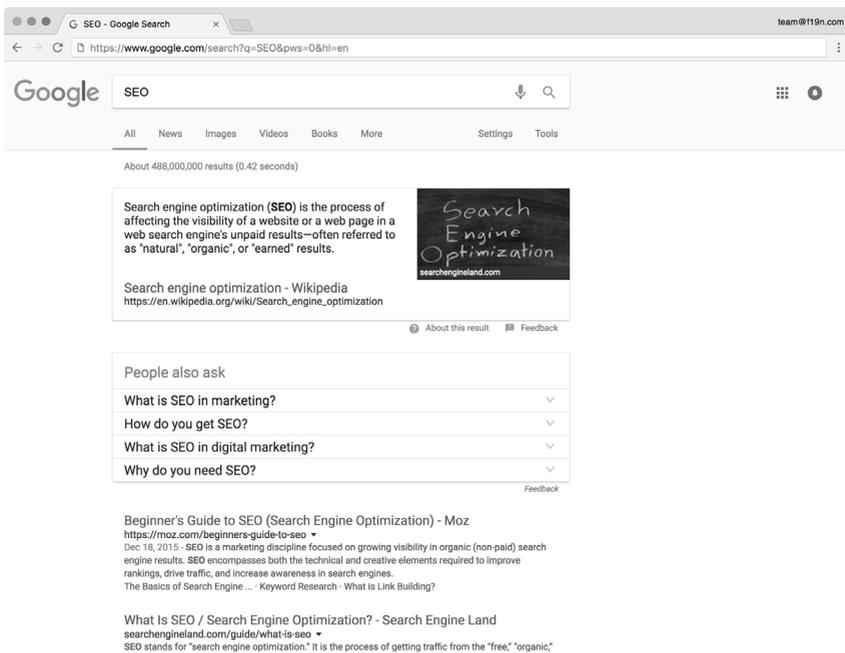
If you want to perform well in Google Search, just read the f*cking official manual. There are more than 400 technical specifications, recommendations, documents, articles and tools about how to do well in Google Search. They are written and published by Google (and yes, I counted them, stopped after 400, so that’s how I know it’s more than 400). That doesn’t take into account specifications (also called “specs”) and recommendations published by initiatives like schema.org, webpagespeedtest.org, and others where Google is a leading member. Do you want to know how to optimize your site for mobile search? There is a spec for that and a Google tool! Multiple languages?

Just check the spec. URLs? There is a spec! Unfortunately, not all specs are in one place, but you can find most of them in these sections of Google’s webproperty:

- <https://webmasters.googleblog.com/>
- <https://developers.google.com/>
- <https://www.google.com/webmasters/>
- <https://support.google.com/webmasters>

Pro-tip: Always read the English versions, as they are the most up-to-date and have the clearest wording.

Now that we’ve discussed R.T.F.M, my next piece of advice is: don’t read SEO blogs! There are 400+ official documents directly from Google. There are 488,000,000 opinions about SEO.



Here’s why you shouldn’t read SEO blogs: 80% of everything that is written about SEO on the internet is bullshit. You have an “80:20 ratio” when it comes to SEO blog

posts: that is, an 80% chance of reducing your understanding of SEO because it is either rubbish or just plain bullshit, and only a 20% chance of it being valuable. However, with Google specs, the chances are much better with a “20:80 ratio”: that is, a 20% chance that the spec is outdated or deprecated, but an 80% chance that the spec will give you valuable and actionable insights. Therefore, it is always a much better investment of your time and resources to read Google specs. Ignore blogs, read the specs!

Is this book part of the 80%? In my workshops, I always start with the line “SEO is 80% bullshit and 20% valuable parts!”. There is always one clever person that asks something like, “How do we determine that this workshop is not part of the 80%?” Well, I can’t decide that for you. Only you can. After you have read this book cover to cover, SEO must make sense to you.

Be Sure To Get Feedback!

Now to the second action point of the “SEO trick” strategy: action and reaction - whenever you do something on your webproperty, make sure you get feedback. As a minimum, always check:

- Google Search Console
- Google Analytics (or another user-friendly, accessible analytics tool)

And, for every actively maintained webproperty, you should review these:

- at least half an hour per webproperty
- after every (development, content or other) release
- at least every two days
- forever

Click around in Google Search Console, browse the tables and beautiful charts, and if you see something that makes your head scratch, please just R.T.F.M. Most of the time there is a “?” right next to the charts, and it leads you to the relevant documentation. If you do these two steps - R.T.F.M and get feedback - right, you will make fewer mistakes, and when (not “if”) you make mistakes, you will recognize them sooner rather than later. And you will learn from them. If you make the right decisions, you will see the positive impact, which will encourage you to learn and do more. You can only win.

CHAPTER 2: SEO SETUP

The SEO setup is a vital part of understanding SEO. All SEO projects that fail, fail because of setup issues. In this chapter, we will discuss your SEO goal, SEO key areas, the SEO point of view, and setup recommendations. If the SEO setup fails, then SEO will fail, too.

The Goal

As we defined previously, SEO is all business activities that contribute to the goal of getting found online. One of the most important parts of SEO setup is your goal, so let's start with that. If you do not have a measurable goal, or you cannot easily measure your progress towards that goal, don't start with SEO. Otherwise, you will:

- a) not reach any goal
- b) have already achieved it
- c) do nothing
- d) do the wrong things.

Your SEO goal must correlate with your business goals. Some companies measure their SEO success via:

- revenue
- customers
- sales
- sign-ups
- visitors
- pageviews

Any Key Performance Indicator (KPI) that works for you and that you can easily and consistently measure is okay. It is important that you know where you are (measure the status quo of your KPI) and define where you want to be using the same metric. My recommendation is to use simple "visits via organic search" as your main metric. The traffic must fulfill some quality criteria. That means it must be sustainable and sensemaking. One off acts that increase traffic for a short period are fine - and are even possible via search - but do not contribute to the overall business goal if they are not

replicable. From a long-term view, the goal must be sustainable traffic growth. When your website is receiving search traffic, it must be possible to keep that traffic, iterate on it and to get more traffic using a new or existing strategy.

However, the traffic must be sensemaking. The users must have searched for words that make some sense for you and your business. They must have to do with your topic. If you show up for [poker bets] but are selling lawn mowers, then this does not make sense. Making sense also means that the traffic must fulfill some minimal value requirements. For example, the organic search traffic should have a bounce back rate below x% and an average time on site of about z min. I will not give figures of x and z at this point because it depends on your individual business and the vertical (the segment of the market) you are in.

Why do I recommend “sensemaking traffic” - a simple and even fluffy metric - as a KPI, if you could use more business relevant metrics, such as net sales, customers, and sign-ups? Well, first of all, visits via organic search are easy to measure on a regular basis. They can be easily assessed via Google Analytics (or another analytics tools) and with Google Search Console.

Look at your number of visits for the last 12 months, find an average, multiply it by 1.5, 2, 3, or 5 and you have a good enough goal for 12 months down the road. I can't say if 50% or 100% growth on your webproperty is a valid traffic goal for your business. Only you know that. If you are just starting with SEO, I would suggest going for the 200% goal, which should be easily achievable if you implement the recommendations in this book.

The second reason I recommend such a simple goal is that as soon as you measure something that happens “later”, such as sign-ups, you are adding in factors that are unrelated to your SEO strategy. The biggest lever to increase sign-ups is - quite often - not more traffic but a simpler sign-up form. Optimizing the sign-up funnel is not SEO but conversion rate optimization (another fascinating area). Is conversion rate optimization complementary to SEO? Sure, but is it SEO? No. Let's keep it simple. You must be able to measure your progress towards your SEO goal easily.

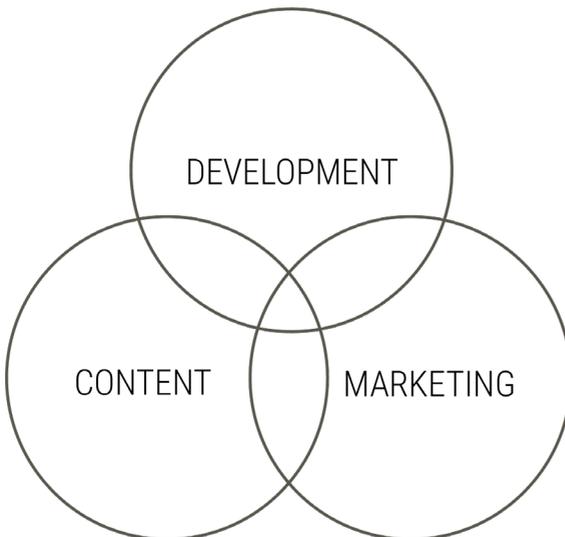
Key SEO Areas

SEO acts in the key areas of:

- development
- marketing
- content

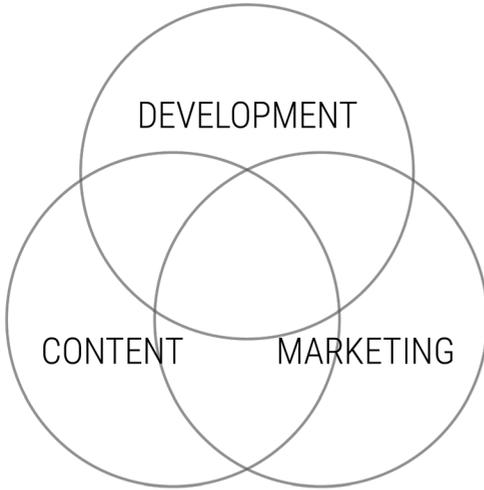
Development ranges from your internal IT department to an external service provider, through to your favorite niece who put together your website. *Marketing* is everyone and everything that communicates your website to the public. That includes traditional marketing, business development, communications, PR, sales and well actually, all of your staff. Yes, I have a very broad notion of marketing! *Content* is a widely used term that involves your product-development team, your database maintainers, editors, journalists, company bloggers... everyone that contributes the stuff that appears on your website.

When these key areas work together, it should look something like this:



The image on the left is the Perfect-World Scenario. Everybody can excel in their own right while working together - fast, direct and unbureaucratically - on critical topics.

However, things can get tricky. Why does it get tricky? Because the thing is, for SEO these areas overlap. For small companies and startups the team structure usually looks like this:



In this case, the team typically sits in one room, and only one person is responsible for two or sometimes all of these areas. At most, the relevant team member is within earshot of the rest of the team. If the marketing team has a cool feature idea, they can discuss it immediately, and development can get to work. The small companies and startup scenario can be search engine optimized. They will - in the current setup - never be the absolute best in any of these areas but, overall, they can make search engine optimized decisions.

If it's a blog, most of the time all of the key functions rest with one person and the scenario will look like this:



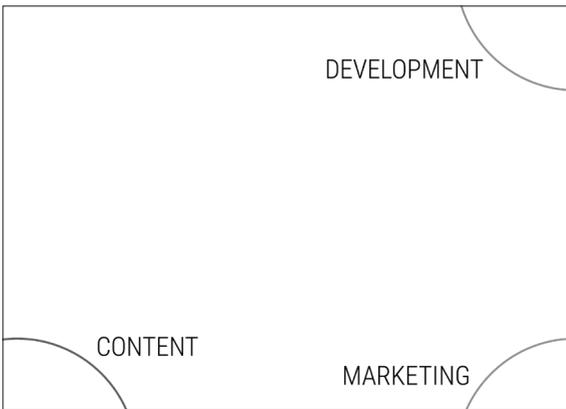
Content creation, technical development (for example, installing a new plugin) and marketing (such as promoting new posts via social media) are all done by the blogger. A blog - if the right technical decisions are made - can be highly search engine optimized. On this note, if you are looking to make a blog search engine optimized, the right technical decisions for you are:

- WordPress
- on your own domain
- on a fast server
- with YOAST SEO Plugin

- W3 Total Cache
- a lightweight, responsive and unbelievable fast theme

With this setup, a blogger will get 80% of SEO platform optimization for an editorial webproperty right and can start focusing on content and marketing immediately.

Then, well, there are the big companies and the scenario looks like this:



The marketing and development teams are not only in separate buildings; they are not allowed - by some internal business process - to talk to each other directly. And even if they did, they are usually chasing very different KPIs. Maybe the marketing and development teams have been in a Cold War-like relationship for years. Perhaps the marketing and content

teams believe they work better separately. Companies such as these cannot be efficiently search engine optimized. You might achieve some quick wins where you suddenly perform with much better traffic. Nevertheless, you will still underperform compared to what you could achieve.

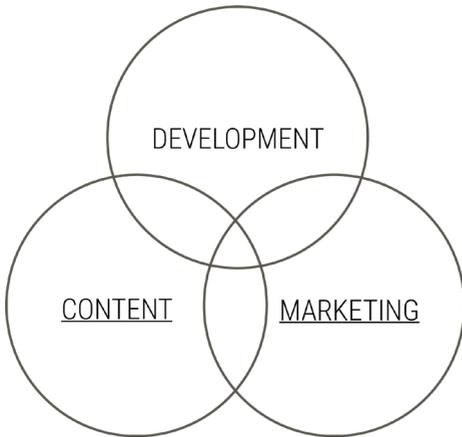
How does this look in your company?

For certain companies - especially small startups - the SEO setup will not be an issue. In successful cases, the founding team wants to be found; they will share a lot of knowledge, and they give each other immediate feedback. For others, it will be an enormous and defining issue of their SEO project.

Neither development, content nor marketing are optional aspects. If you optimize your technical platform, and your content does not fulfill some quality criteria, you will not have sustainable SEO success. If you do not communicate your content via a

marketing channel, you will always underperform. And, if your content does not have unique URLs (for example) because of unresolved technical issues, SEO will not work. Therefore, neither development, content or marketing can be neglected. Every part must be optimize-able! Every part *should* be optimized. If one part fails, SEO will not work. If one area is not interested in getting found, SEO will not work. *All SEO projects that fail, fail because of SEO setup issues.* SEO is solved! The execution is the challenge!

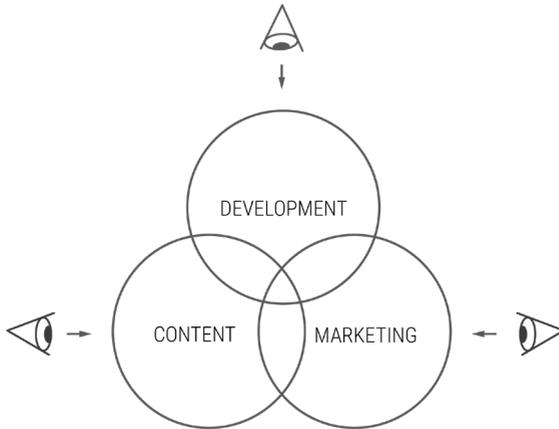
There has been quite a buzz about *content marketing* in the last few years. From an SEO perspective, if you get the development right and have a “good enough” technical platform, then you can focus on content and marketing, otherwise known as “Content Marketing.”



The rise of content marketing has been possible for many editorial-led websites (for example blogs and magazines) because a “good enough” technical platform became available for free, namely WordPress + Yoast + W3 Total Cache. This setup is not perfect (and very often not a feasible option due to other requirements of the platform) but definitely “good enough” for an editorial webproperty from a SEO perspective.

SEO Point Of View

The three SEO key areas of marketing, development and content I have outlined are the reasons that people can make so many different claims about SEO. Quite often these things seem mutually exclusive.



From a content point of view: Google is the biggest content marketplace in the world. From a marketing point of view: search is just another way to promote your company, products, and offerings to an audience. And from a technical point of view: the website must fulfill some technical requirements to get crawled, indexed and displayed by Google. All of

these points are true. None of these viewpoints are more important or valid than the other, nor are they optional.

SEO Setup Recommendation

Now that you understand that SEO is a challenge that crosses several different business areas, how do I recommend implementing it? Here's what you do: you assemble your head(s) of marketing (and everybody who communicates in the name of your company), your development team and your content team in one room. Most of the time, this will be more than three people. From there, make sure they are completely on board with the goal of "getting found." Make it abundantly clear that they will need to change how they work together. Assure them that success in this area will be their success and that they will get their reward (whether it be praise, bonuses, raises or a handshake). Then set their shared KPIs. Don't let them out of that room until you are convinced that they are totally dedicated to achieving the common goal. If one of the involved parties "does

not have time” to meet, give them time. A “does not have time” excuse is always a sign that priorities are not properly set.

The biggest risk is that one part does not want to “get found” or just does not commit to the goal of “getting found.” I once had a developer make it abundantly clear that he had worked at the company for ten years, that the website was his baby, and he could explain exactly, with flawless reasoning, why the website is the way it is. Another time, I encountered a marketing team that decided that they cannot communicate online at all, because “the TV commercial takes up all the budget this year!”. And on another occasion, the editorial team of a news magazine did not show up to my SEO workshop, because they were “independent from the company.” They all had – to them - valid reasons. However, if you want to get found - with the goal of growing sustainable, sensemaking traffic - then the content, development and marketing teams must all want to get found. You cannot solve the problems created by the old way of thinking, by using the old way of thinking.

After your hands-on meeting and once all the stakeholders are in agreement, I recommend naming an SEO manager. But be aware: the role of an SEO manager is a very unthankful one! If the company succeeds SEO wise, it will not be their success alone. It should also be mentioned that the SEO manager will never be granted their own big department. I would even strongly recommend not naming the role “SEO Manager.” Label it product manager, project manager, delivery manager or something entirely different. Regardless of their title, this individual will work across all three departments. The task of an SEO manager is to sync between the departments, to share knowledge, to push SEO tasks, to provide timely feedback on progress, and to double down on all topics concerning SEO.

So once again: the SEO manager is not assigned to their own department (at least not a big, parallel world one) and they will not claim SEO success as their individual success because it isn't. After you have got your SEO setup together and agreed on your goal, you should take the time to make sure everybody understands precisely what SEO is.

Common SEO Setup Mistakes

The SEO Department Mistake

Do not build your own SEO department. SEO must always be a shared responsibility. If you create your own SEO department where you implement parallel structures, for example:

- SEO Programmer vs. IT
- Link Building Team vs. Marketing
- SEO Texter vs. Editorial Team

Then you will

- a) waste of a lot of money
- b) lose a lot of time
- c) underperform.

The CEO SEO Mistake

CEOs are terrible SEOs. They fulfill the first criterion “work within all three areas,” but they are horrible with the second key responsibility of “sharing knowledge” and the third of “delivering timely feedback.” CEOs have a lot to do and their time and resources are not always best invested with SEO.

The Good Looking Wordpress Theme Mistake

If you let a custom themes designer “polish” your theme, there is a very high likelihood that they will screw up your website. If you have just started your first WordPress based website, use a standard, simple, unbelievable fast theme and just add your own logo and colors, nothing more. Speed must be your first priority for choosing and adapting a Wordpress theme.

CHAPTER 3: SYSTEMATIC SEO

This chapter will give you a bird’s view of everything SEO, including the user, the webproperty, Google, distribution, and traffic as well as feedback loops and targeting. All of these topics will be revisited again in detail later in this book.

The User

When in doubt, start with the user. As there is always doubt in the online world, always start with the user.

The user is a strange, mystical creature. They:

- do not have an age but an age range
- do not have a gender (but they are still never an “it”) but a gender distribution
- might use the website on a desktop or mobile (or something in-between)
- do what they want, not what we want
- might be located anywhere in the world, and even then, they could be at home on their couch, at work doing work, at work doing something else, on a bus, or even on a beach somewhere.



As a matter of fact, we don’t know much about the user until we have a closer look. The user is similar to Schrödinger’s Cat: “The User” is all of the things listed above and much more, at least until we have a closer look. Furthermore, we can determine a lot of things about a single user after they have visited our site, but we cannot say much about users in general, and we cannot predict future users. What a bummer. What we can say for certain is that the user is not us. We know the technical, business, and organizational reasons why our website is the way it is; the user doesn’t. We care about our business; the user doesn’t (yet). We (sometimes) read what we publish on our website; the user doesn’t.

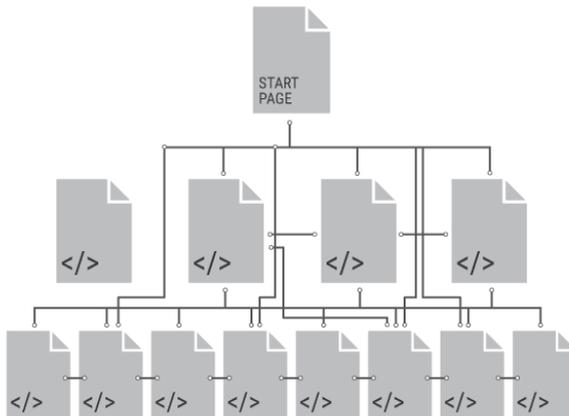
Remember that we are users, too, on other web properties. The user is:

- efficient
- lazy
- clever

Any online business that treats its users like a mindless herd of lemming that just needs to get pushed down a funnel has already lost. The user knows what they want - but we don't. We must make it easy for them to do what they want, because the user is efficient and doesn't like to waste time. On top of this, the user is also lazy and clever, and will always use the easiest means to accomplish their goal.

The Webproperty

Your webproperty might consist of one or more domains and one or more subdomains, and most of the time has a startpage and a lot of other pages. This image below is a symbolic presentation of a webproperty:



Links connect the startpage with pages, and links connect pages to other pages. There are thousands of variations and mutations of this arrangement (startpage only, single page apps, multiple startpages, and multi-language domains to name a few).

The User And The Webproperty

In a perfect world, we wouldn't need to worry about anything. In a perfect world, a user would be interested in what we have to offer - they would simply type the URL into the address bar of the browser, and our startpage would open. They would navigate through our website using our well-thought-out navigation structure and do what we want them to do because we know what's best for them. Of course, this never happens. In most cases, the user does not know us, does not care about us, does not enter via the startpage and completely ignores our fancy navigation logic. And they don't do what we want them to do. What the user does is google.

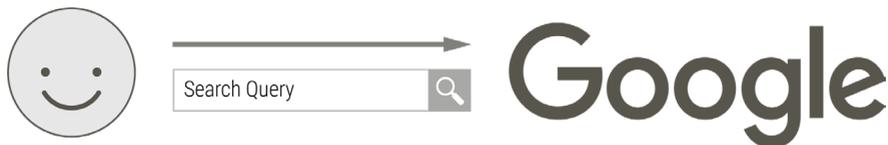


Google is a marketplace! It is by far the biggest marketplace this world has ever seen. Google takes an inventory (its index of web resources like web pages, images, videos, maps, products, locations, knowledge graph data and so on) and matches it with the user's demand.

For all western markets, Google is the dominant search engine and therefore the dominant content marketplace. Most of the principles introduced in this book are also suitable for other search engines, but it hardly matters. Google is the undisputed number one search engine in all western markets. Second place goes to YouTube, third to Amazon and then there are some other minor players. Some people (mostly politicians and media companies) say that Google is a monopoly (in most markets it most likely is), and while it does make for some interesting discussion, the topic is not relevant for SEO. You have to optimize for the market that there is, not for a market you wish to exist. And for us, that market is Google.

The User And Google

The user googles, always, all the time. He or she types (or speaks) a query in the Google Search input field or Chrome omnibar (or the equivalent on the other browsers) or search widget on a phone and thousands of other places. The user does not pause to think about what to write, they just google.



A query can range from a word to multiple words to a full sentence. It's rarely even a thought; it's usually half of a thought, one that the user is still thinking. Quite often it's not even spelled right or in a logical order. But every user that googles has an intent: a "job to be done," something they want to achieve. Everybody does it, all the time, without much thinking. They just google.

We can categorize a user's search intent into "Do-," "Know-" and "Go-" queries.

- 1) Do-queries (action intent): also known as "action" and "transaction" queries. The user wants to accomplish a goal or engage in an activity: download an app, play an online game, book a hotel, order a new coffee machine or find awesome cat videos to watch.
- 2) Know-queries (information intent): the user wants to find information; they want to know something for example: the best time to travel to Sri Lanka, or what the weather is like in Italy in June.
- 3) Go-queries (navigation intent): the user wants to navigate to an individual web page that they already know about, most of the time the startpage of a particular webproperty such as: Twitter or Facebook.

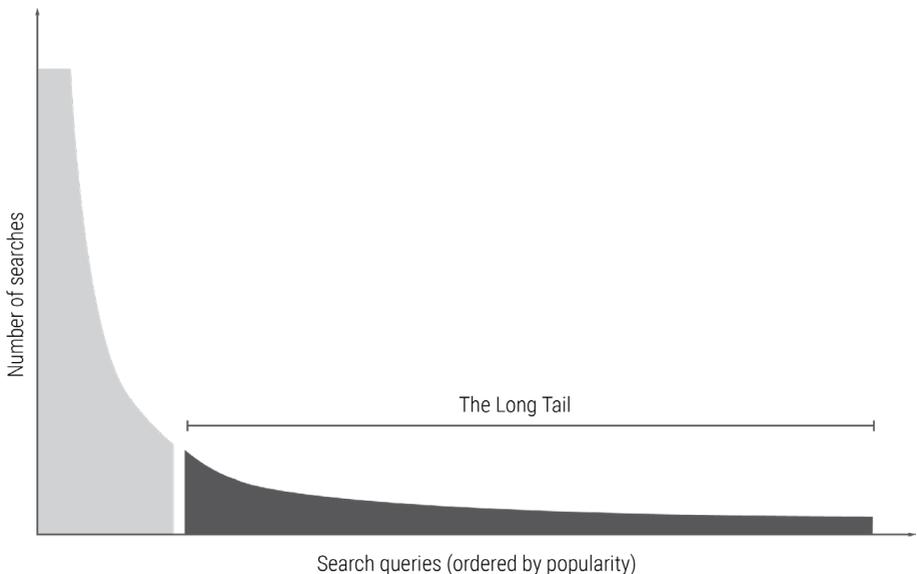
This categorization of do-, know-, and go-queries does not include all varieties of human search behavior, but it is a good starting point for our purposes.

Google takes these chaotic, fragmented and often nonsensical queries, and enriches it

with everything they know about the user. Everything they know about the user can be quite a lot: their location, device, internet speed, current search session, previous search history plus lots of other things we don't know they know, which will have an influence. Google takes a best guess at what the specific intent is, and then matches it to their inventory, which is a huge index of everything that is online. Everything online? Well, no, not exactly. Google matches it to their own inventory: that is, everything that Google has discovered, crawled and indexed. So if we want our pages to be included in that inventory, we need to distribute our content to Google.

Whenever you see a Google query, you must not see these words as mere “keywords.” They are much more important. They are a textual representation of the user’s search intention. If your goal is to have sustainable traffic, then it’s your job to fulfill their intent. You can - for a short time - gain traffic with targeting “keywords,” but you can only maintain sustainable traffic by meeting the user’s search intent. When we talk about queries or search queries in this book, I always mean the “textual representation of the user’s intent.” I don’t mean “keywords.”

Here is what search demand looks like:



In every vertical (the segment of the market you are active in, for example, music, insurances, cars, vegan, dating, poker), there are always search queries that are frequently searched. We call these high volume queries. For the instances for the verticals given above, [shakira], [car insurance], [tesla], [vegan new york] or [online poker] would be high volume queries.

Now, in every vertical, there are always queries that will be searched very infrequently, from a few times a day to once in the whole of eternity. We call these “low volume queries.” So for the instances above, examples would be: [shakira worst video using oil], [car insurance Ohio drunk driving], [should I buy an e-card mastercard], [vegetables for breakfast berghausen austria], [online poker strategy minimal budget 10\$]. The funny thing is that for each and every vertical if you add the search demand of all the high volume queries together, and compare them to the sum of all the low volume queries, you will find that the total amount of searches for the “low volume” queries is much, much bigger than the sum of the “high volume” queries.

The sum of the low volume queries is called the long tail. The sum of the “high volume” queries is referred to as the short tail. For every online vertical, the long tail is much bigger than the short tail. As people search in more diverse ways, the long tail gets longer and longer. To be successful at SEO means to have a strategy (or more than one strategy) to target both the long tail and the short tail. If you target just one of them, you will underperform.

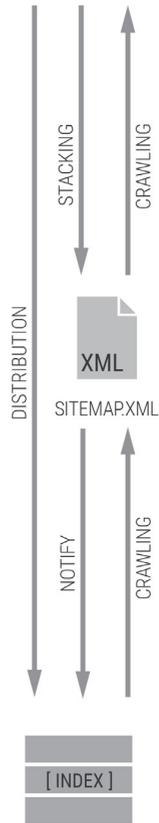
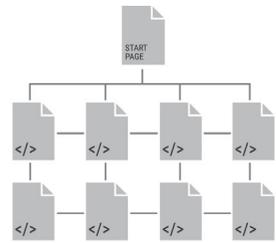
Note: some people say that there are “only short tail verticals.” I tell these people that they just don’t have enough imagination and don’t look hard enough. Sometimes it’s not very obvious, but there is always a long tail.

Distribution

Distribution is the process of communicating our inventory to Google. This is so Google can fulfill its role as a content marketplace and match the user queries with suitable content, with our content as the inventory.

It can be quite technical, but at its simplest, it is a list of all your targeted page URLs. Your targeted page URLs are all the pages that you want to be found for (there will be much, much more about targeted pages later). Most of the time, the list is in an XML format called “sitemap.xml.” You place all your targeted page URLs into a list (called stacking) and notify Google about the existence of this list. Google crawls the list, fills its discovery queue with these URLs, and then crawls your content. If it fulfills a Google internal quality criteria, Google will even index your content (which means it will make a searchable copy of your content within their database). That’s what distribution is. You can optimize your distribution in quite a lot of ways (again, more about this later). What is important is that you have a working, maintainable distribution logic to deliver your content to Google. If distribution does not work, SEO does not work. That’s why we have to make sure distribution works.

Note: do you really need a sitemap.xml? As Google is crawling you anyway? Simple answer is yes, you need a systematic distribution - a sitemap.xml logic. It is the only scalable way to tell if Google has crawled and indexed your content. If you don’t have it, your distribution pretty much depends on luck. And luck does not scale.

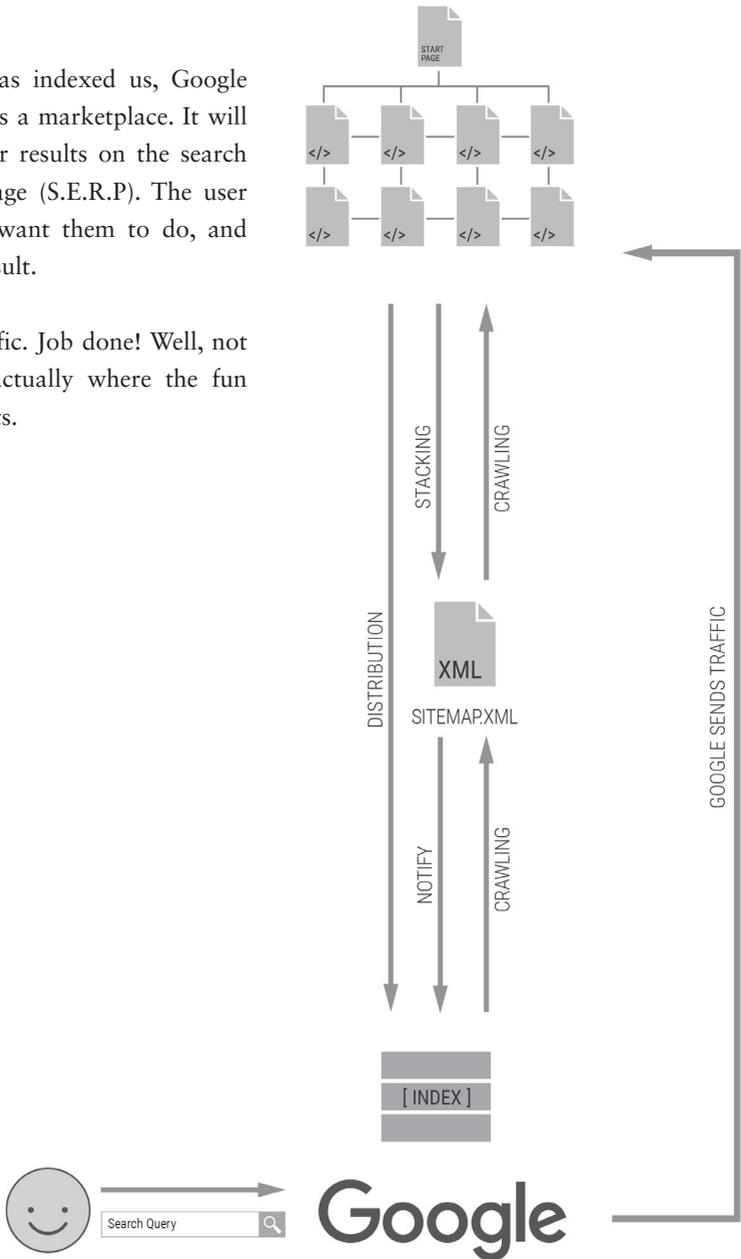


Google

Traffic

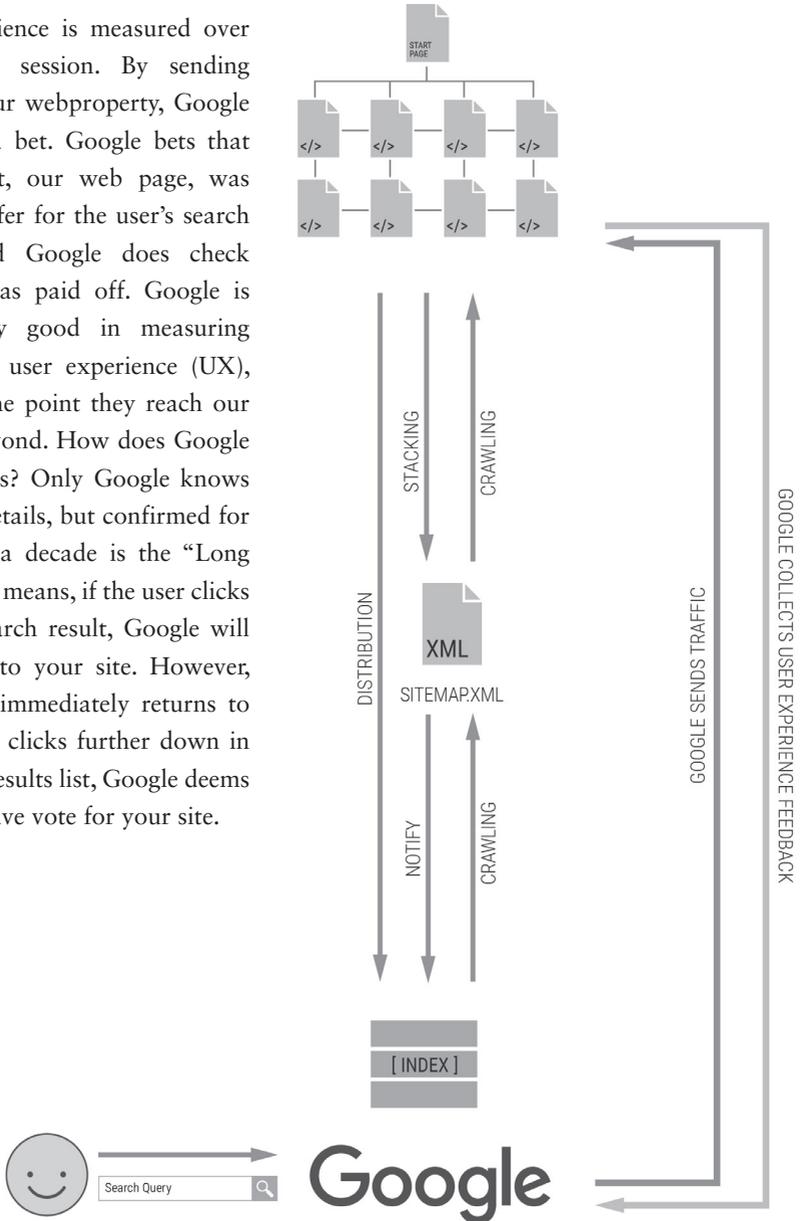
After Google has indexed us, Google fulfills its role as a marketplace. It will then display our results on the search engine result page (S.E.R.P). The user does what we want them to do, and clicks on our result.

Yes, we get traffic. Job done! Well, not quite. This is actually where the fun (and work) starts.



The User Experience Feedback Loop

User Experience is measured over one search session. By sending traffic to our webproperty, Google has made a bet. Google bets that our content, our web page, was the right offer for the user's search intent. And Google does check if its bet has paid off. Google is unbelievably good in measuring their users' user experience (UX), including the point they reach our site and beyond. How does Google measure this? Only Google knows the exact details, but confirmed for more than a decade is the "Long Click." This means, if the user clicks on your search result, Google will send them to your site. However, if the user immediately returns to Google and clicks further down in the search results list, Google deems this a negative vote for your site.



Also, it is suspected that data for measuring user experience is drawn from:

- Google DNS
- Google Chrome (including desktop, Android, and iOS)
- Google Data Saver
- Android, ChromeOS, ...

In its terms and conditions of the now deprecated Google Toolbar, it was even stated very visible that they track their users' online behavior if a certain feature was enabled. Google has denied that Google Analytics and Google AdSense data is used to determine user experience signals for Google Search. Fact is that we don't know exactly how Google collects UX data, but they do, and they are magnificent at it.

It is surprisingly easy to create a website that attracts lots of traffic for a short time. It is much harder to create a webproperty that maintains this traffic and manages to improve on it. If the user experience is negative from Google's point of view, you will not keep your traffic. Plus, there is also another kind of feedback loop.

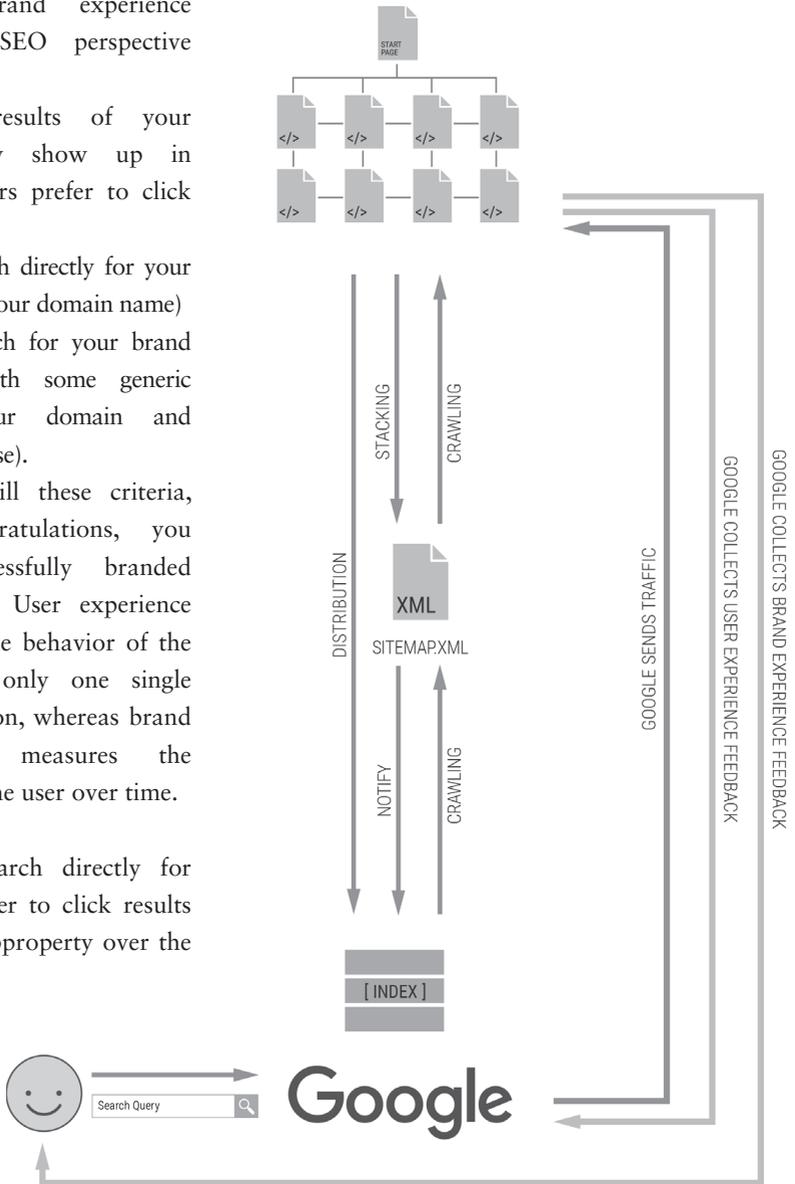
The Brand Experience Feedback Loop

Positive brand experience from an SEO perspective means that:

- when results of your webproperty show up in Google, users prefer to click on them
- users search directly for your brand (i.e.: your domain name)
- users search for your brand together with some generic words (your domain and something else).

If you fulfill these criteria, then congratulations, you have successfully branded your users. User experience looks for the behavior of the user over only one single search session, whereas brand experience measures the impact on the user over time.

If users search directly for you or prefer to click results of your webproperty over the



competition, then this is a very valuable vote for you. On the other hand, if users never search for your brand, or even worse, actively avoid clicking on you in search results, then this is a negative vote for you. You have to “brand” the user as soon as they visit your site. It is crucial that the goal of the users search (“job to be done”) was fulfilled, that they had a good user experience and that they remember your brand (most of the time the name of your website and your domain) and want to visit you again.

It is quite easy for any halfway sophisticated SEO to create a website that attracts lots of traffic for a short amount of time. It takes hard work to create a webproperty that maintains this traffic and increases it. If your brand experience is less than desirable and the users do not begin actively searching for you and clicking on your results, you will not keep nor increase your traffic in the long term.

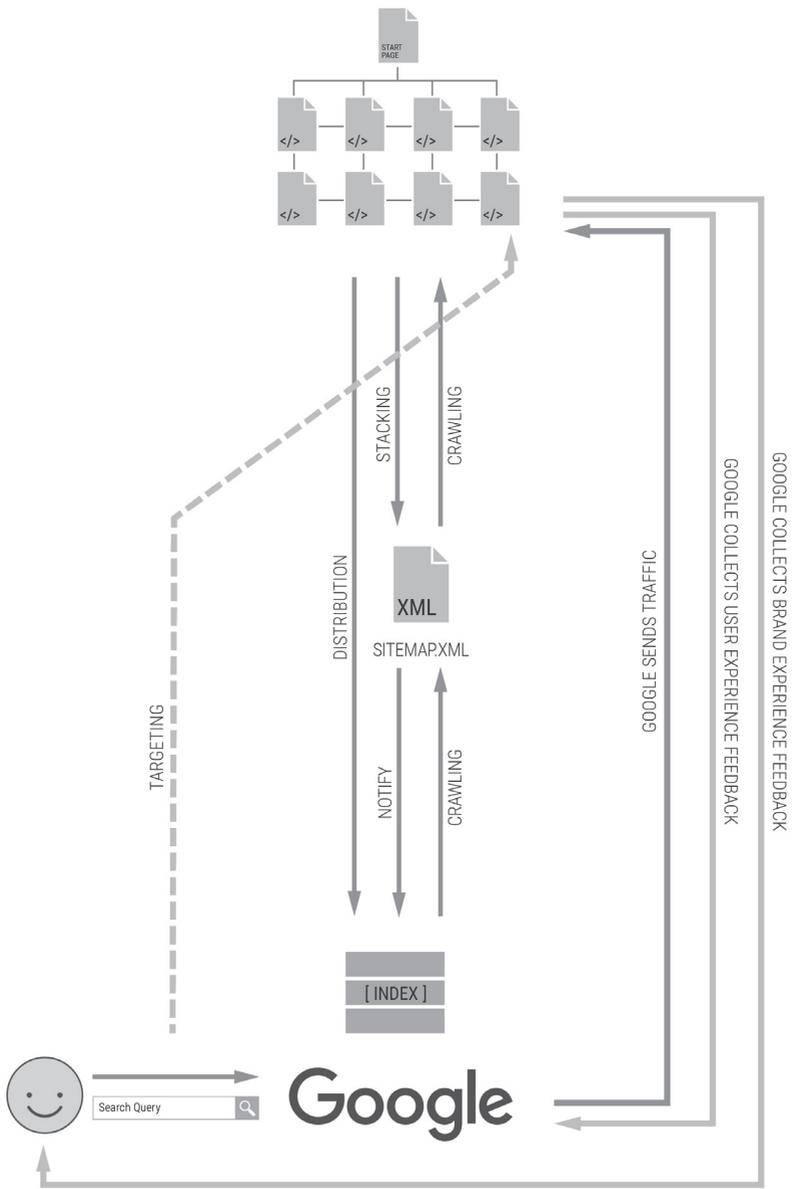
Targeting

You will only be found for a certain user query if you have a page that wants to be found for that particular user query. It is a two-step process, and therefore you have to:

- a) determine/guess/make a hypothesis about what the users might search for
- b) create targeted pages for these searches

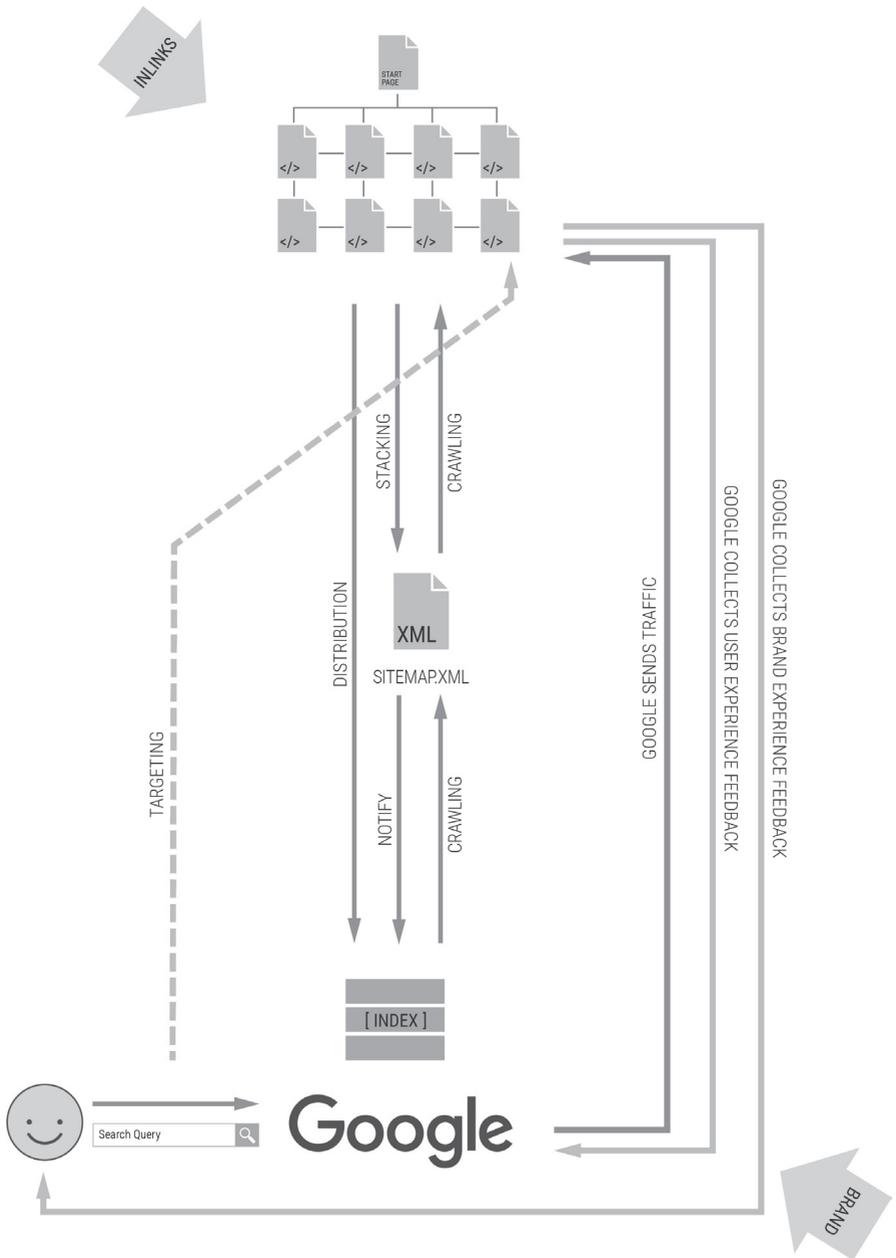
This process is called “targeting.” However, you should know there is a variation for when it is done editorially, and that is commonly referred to as “Keyword Research.”

We target a certain user query and the intent behind it, and create a “targeted” page for that user query and the intent behind it. Targeting can be a manual process or an automatic process, or something in between. Regardless of how we do it, the goal is always the same: to create pages for stuff that people actually search for, might search for or will search for in the future.



Marketing

As we discussed earlier, it is important to remember that the term “marketing” should be used in the broadest way possible. It is everyone and everything that communicates your website to the public. With this in mind, we are now going to add two more arrows to our diagram: branding and inlinks.



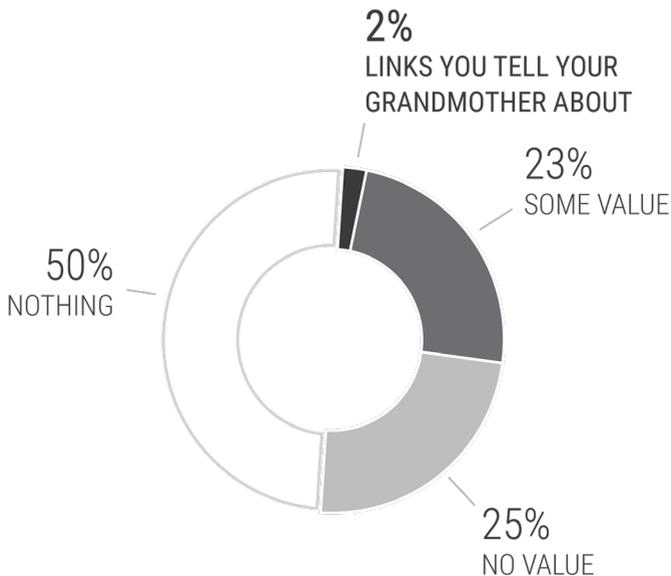
Inlinks

Inlinks (also called inbound links, backlinks or simply links) are links pointing to your webproperty from other webproperties. They are still a factor for Google to determine site and page relevance. That is unfortunate, because it places a dependency on external factors, and must be considered a resource black hole.

Nevertheless, the main thing to remember is this: although you need inlinks, you don't need a fixed number - not 1,000, not 10,000 nor 100,000. You need a healthy growth over time. That means a variety of different domains must link - *in a sensemaking way* - to pages on your site. Never mind if it's ten websites this month and only four the next. What matters is growth over time and that many different websites link to yours. A site without a growth of inlinks will always underperform in the long term.

Link Building

The truth about link building is that 75% of all the resources you invest into it will be wasted. The pie chart below shows what happens to your link building investments.



You can see from the pie chart that 50% of everything you do will lead to nothing. That's 50% of your crafted content that will not get any form of recognition in the form of links. That's 50% of all your emails that will go unanswered, 50% of all your pitches that will be ignored and 50% of your link bait stunts (that is, creating pages or web apps uniquely for the purpose of getting linked) that will not work at all.

Of the remaining 50%, half will lead to links that will have no positive impact on your website's SEO performance. If you do it very badly (and the links are spammy), they might even have a negative impact. Therefore, you are left with 25%. Two percent of that will be the "links you tell your grandmother about," so for instance when an international magazine covers you, you were mentioned on the BBC (with a link of course), or the leading blog of your niche wrote a review about you. These are highly valuable links - links that will have a direct impact on your website's traffic. Today you may be performing ok, then the very next day you get a "link you tell your grandmother about" from a major source which will make your site perform much better. The remaining 23% are the links that will have some positive impact on your site and help your overall growth over time.

But no matter how you look at it, 75% of all your invested resources into link building will be wasted. There are tools out there that promise to tell the difference between "no or negative impact links and positive impact links." They are useful for identifying those big impact "links you told your grandmother about" (which you already told your grandmother about). For classifying the remaining inlinks into links that have some value vs. those that don't, these tools only work half of the time, so you might as well flip a coin.

So, what should you do? If you do link building for the sake of link building, the quality of your links will deteriorate quite quickly. At first, they will just not be sensemaking. Then they will become spammy, and then just plain spam. ***So don't do link building!*** Do marketing, do PR, do business development as normal, but with *links in mind!*

You don't need a link building team. Instead, aim to fill the wasted 75% of your resources with additional value. Make sure that every person that communicates in the

name of your company is thinking about links too. Public relations should think about links for example; all press releases should link to valuable, and relevant targeted pages. Don't get me wrong; I do not recommend writing a press release for sharing links, but if you have a story to tell that might fit the format of a classic press release, make sure that you link back to the correct parts of your site. Marketing, done right, has immense value even without any links (for example, building brand awareness). Therefore it's great when you can boost its value by making sure there is at least something to be linked to.

Make sure your team is keeping an eye out for any opportunity to link to your site. In sales, your clients can link to you, ideally with an awesome recommendation. Even in business development or customer support, there will be opportunities if you are looking for them. Making links part of the routine activity of business divisions, with their own quality standards, will ensure that the links remain "sensemaking" and are of a high quality.

Even with the business integrated approach I have suggested above, half of all the links you get will have no impact. But at least you will not squander your invested efforts as you will still receive value with these activities. Anyway, compared to what really matters (Online Brand Building, see next section) linkbuilding is just an annoying nuisance.

Brand Building

Any form of marketing can be used to push brand awareness. It is unbelievably important for sustainable SEO success that users start:

- to google your brand
- to google your brand with generic terms
- to prefer to click on your pages if they see them in the SERPs.

These points above can all be achieved by providing a good user experience (that is, fulfilling the user's search intent and communicating your value clearly) once they're on your site. But there's more you can do.

It is quite common that following offline campaigns (for example, extensive TV ads) not only does the branded search traffic increase, but also the unbranded organic search

traffic. After the advertising campaign is over, unbranded organic search traffic stays high while the branded search traffic goes down again. The same effect can also be seen to a lesser extent in radio, print and billboard campaigns.

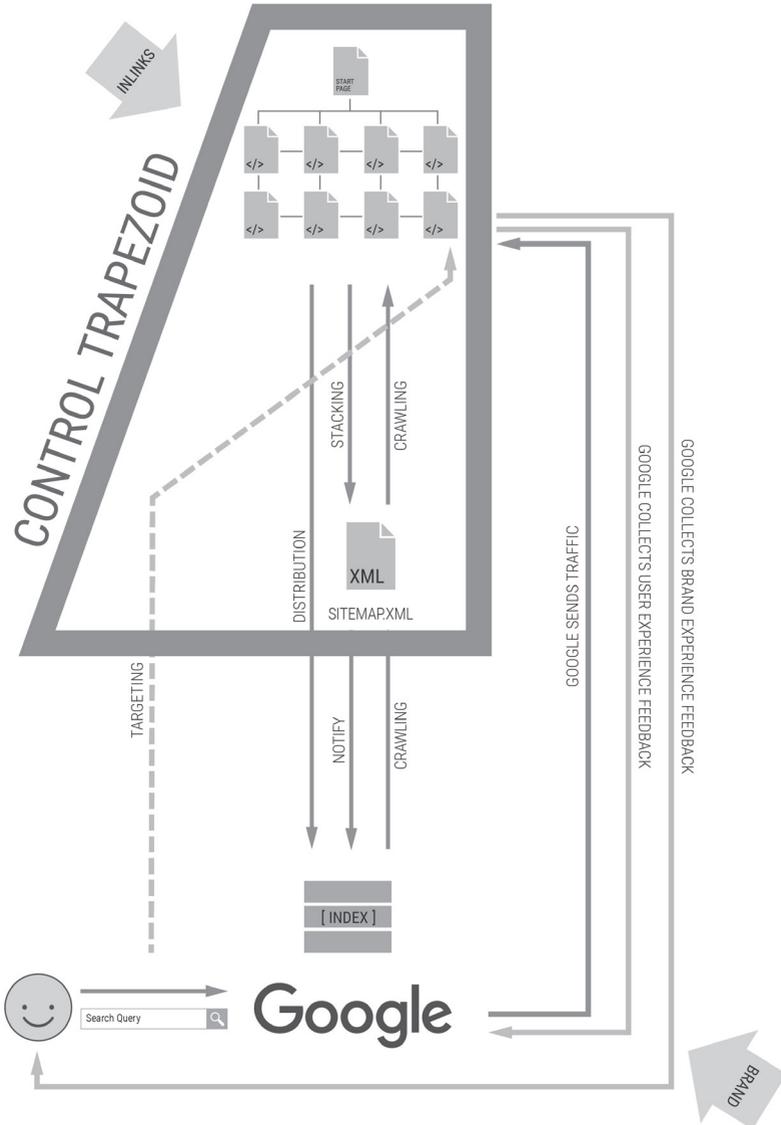
Google has repeatedly stated that paid search listings - Adwords - do not have a direct positive impact on organic, unpaid traffic. However, as any good SEA (Search Engine Advertising person) will tell you, it is “possible” that if you have an extensive AdWords campaign your organic traffic - assuming you have an SEO optimized platform - also rises. After having observed and tested this myself multiple times, my conclusion is that this is due to the brand effect.

Similarly, Google states that their ad network, AdSense, does not have a direct positive impact on organic search performance. But if you use AdSense as a retargeting channel, so basically stalking the users with your brand throughout the web after they have visited you once, it is an excellent way to remind users of your brand. However, be very careful, because retargeting can quickly become very annoying (and this, in turn, is negative branding).

Also, social media is a great way to build a brand. It can be another great channel for businesses. If you are just starting out, it’s sometimes even faster and easier to gain users via social media rather than SEO. I strongly recommend for everyone to have a social media strategy as well as a search strategy, but *social media is not SEO!* Social media is a great way to build a fanbase and position your brand. If they like you - and follow you - on Facebook, there is a higher possibility that they will click on your site when they see it in their search results. They might actually google your website once in a while. However, whenever you do social media - and you should - don’t do it just because of SEO; social media is a channel in its own right.

What You Need To Focus On

It's essential you focus on things that are within your direct control. Let's discuss the diagram below.



Everything within the trapezoid is within your control. That includes:

- the user experience and brand experience once the user reaches your site
- the pages of your site and how they interlink with each other
- what pages you have and what pages you don't have
- what queries you target with what pages
- the distribution until the sitemap.xml

My recommendation is to focus strongly on the things within your control. 100% of everything you invest in the area within the trapezoid will be 100% efficient, minus only your own internal overhead.

Google is not within your control. You cannot tell Google what to do. So don't focus on Google. Google is a black box; you cannot 100% understand Google nor can you reverse engineer Google. What you can focus on is what and how you distribute to Google (quantity and quality wise). You can make educated estimates on what the users demand (as in, search demand). The user is also not within your control. You cannot force the user to use the words you want them to use. Yes, it's true you can influence them via brand building, but that is far from controlling the user. However, the queries you choose to target *are* within your control. Inlinks are not within your direct control. That is work that somebody else, another webproperty, has to do. You can influence it, but you can't do it yourself. (Technically, you could build some kind of link farm, with millions of pages with hundreds of different domains within your control - all interlinking each other. But let's just say it doesn't work anymore and will get you kicked out of Google in no time).

Feedback

Now that you know what you can control, remember that getting feedback is critical to success. Previously, I mentioned two major things I want you to take away from this book. One of them is: if you do something be sure to get some feedback. We can also turn this sentence around: if there is no form of feedback available, then don't do it. Thankfully there are a lot of avenues to access SEO feedback, and it's all readily attainable with free tools. Let's look at a few of the major feedback and analysis tools that will help you optimize your pages and strategies.

Google Analytics (GA) or another easily accessible analytics solution is essential for your webproperty. Easily accessible means that you - or anyone who has anything to do with SEO (development, marketing, content) can access the data without too much hassle and can easily make a meaningful analysis of this data, even without in-depth training. If you have an analytics solution that

- a) does not work
- b) only a select few can access
- c) that people hate

it does not count as "easily accessible analytics solution." Google Analytics is free and easy to configure, has a huge, and mostly helpful community and I highly recommend implementing it.

Google Search Console (in short: GSC) can unearth 80% of everything that is wrong (or right) from an SEO perspective with your webproperty. GSC does not need installation; you just need to verify it on your site. (Simply google [Google Search Console] and follow the instructions on how to verify it.) No paid tools can come close to the insights available in GSC.

Important note: Add all protocol domain variations to GSC. That means if your site is *https://www.example.com* you should add the following to GSC:

- *https://www.example.com*
- *https://example.com*
- *http://www.example.com*
- *http://example.com*

Protocol and (Sub)Domain SEO issues are quite common (for example, you get pages of the www-subdomain indexed on the non-www-subdomain version), and GSC provides an easy way to find these errors. GSC is the most important SEO feedback tool. I will often refer to GSC, but I will not explain each and every metric as this would be a book on its own (If “Google Search Console - The Missing Manual” existed, I would buy it). Until then, it’s up to you to check out GSC and whenever you have a question, just R.T.F.M. (it’s usually hidden in GSC reports behind a circled “?”).

Google Search Console has various metrics that relate to distribution, for example, submitted pages via sitemap.xml, indexed pages from sitemap.xml, indexed pages overall, crawling stats, and more. If you want to know if your distribution is working, just check GSC. What’s not in there is the logic that you use to decide what you list in the sitemap.xml. That must be controlled via some internal logic (for example, custom logfiles or just looking into the xml files).

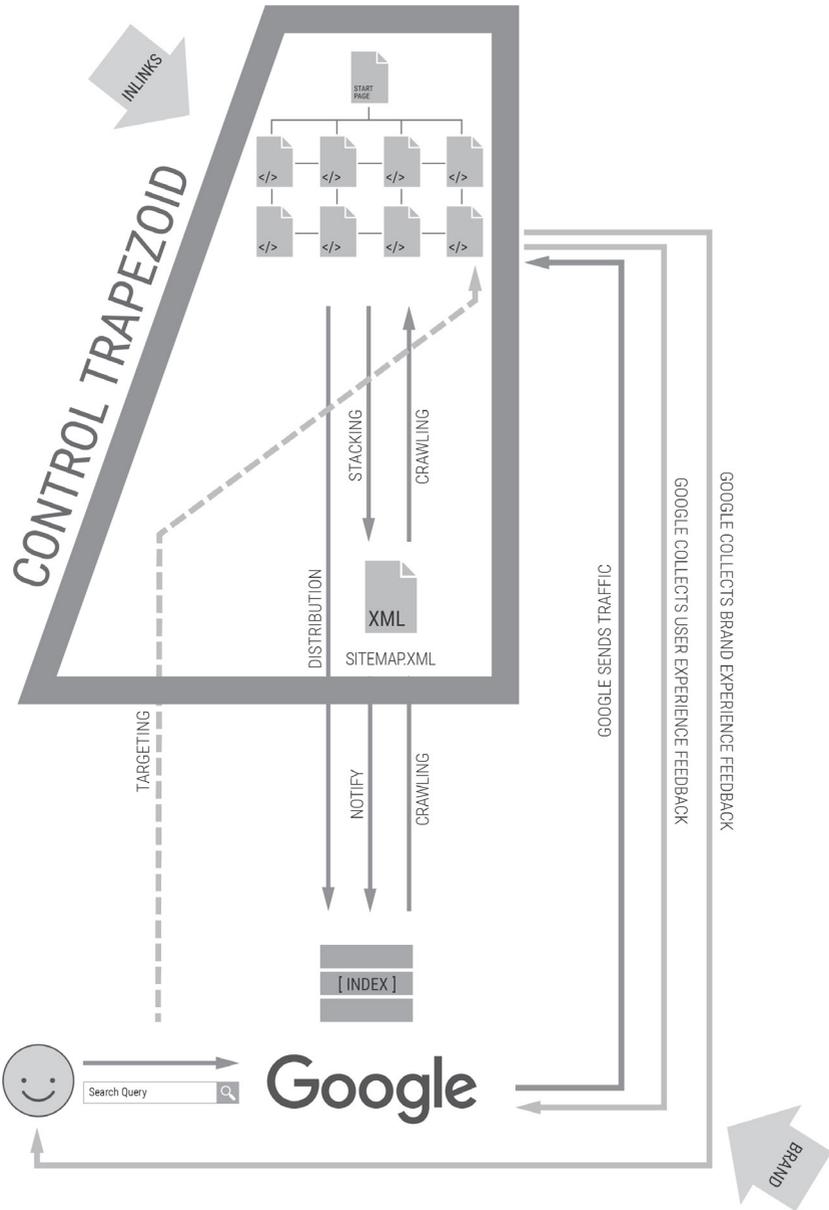
I.e.: what queries you show up for in Google (otherwise known as impressions) and what pages you show up for (top pages) is listed in GSC. The impressions that lead to clicks are also listed in GSC. You can filter those impressions and clicks for countries, devices, search type and more. The user experience for users coming from search can be measured via Google Analytics. What the users do on the page and if their search goal is fulfilled can also be analyzed with GA. If people actively google your brand and preferably click on your results, GSC covers you. A lot of onpage feedback can be found in GSC, as well as a lot of onsite stuff (for example, most internal linked pages, duplicate title tags, and so on). You can check in GSC how many links your site has and also get some great link data via GA referrers. If you want to identify the terms the user searches for and for which you do not show up yet, you won’t have much luck with GA and GSC but try:

- Google Autocomplete / Google Suggest
- Google Trends
- Google Adwords Keyword Planner

The data from the tools mentioned above can help you identify what new targeted pages you should create.

If not everything of what I mentioned here makes sense yet, don't worry. It will - by the end of this book. Right now, the only thing I want to make clear, is that there is always data and the possibility to check the impact of your actions. And you really need to look. Not regularly looking at GA and GSC is like driving with your eyes closed.

It's A Model!



The diagram on the previous page is not SEO. It's a model, a framework, a way to *think* about SEO. And it's damn useful! (SEO is what you literally do.) It's the way I propose you think about SEO for your business. Why? It's measurable and manageable.

It shows what impacts what and gives you a lot of leverage points. It's relatively straightforward (all models are simplifications). Everything that has to do with SEO has a place somewhere in this diagram. If you cannot pinpoint it in there, it doesn't have anything to do with SEO.

Note: If you disagree, please come up with a better model. That is called progress, and I'm looking forward to it.

Common Mistake: Reverse Engineering Google

Do not try to reverse engineer Google. You will not succeed. Even if we were to take everything we know about Google and make estimated guesses about everything we are aware that we do not know, we would still not have anything even approaching the complexity of Google. There is just too much that we don't know that we don't know! And it continuously grows. Let's go even further and imagine we get our hands on the source code of Google, and we implement what we found, we would still not have Google. We do not have the same underlying data, nor the same index, nor the same user search history as Google does. We will never 100% understand Google.

You cannot understand Google, but you can understand search, and if you understand search you can understand SEO. There are a multitude of blogs and "experts" who claim they figured out Mayday, Caffeine, Panda, Panda 2.0, Magic Unicorn, Venice, Penguin or any other stupid animal name Google uses for their publicly announced algorithm updates. These so-called "SEO experts" don't know what they don't know. Additionally, nearly all SEO algorithm updates are not directly actionable for webmasters as we cannot change how Google works. So more or less just ignore them.

I say more or less because each event (algorithm update) might be part of a trend. Sometimes such a trend is quite a good reason to change how we interact with Google

and the users (for example, mobile, speed, HTTPS). But if Google does this, Google will communicate this, so again, you can (must!) ignore all the SEO blog posts by “SEO experts.”

Tip: don't look for stories or explanations, look for trends.

I recommend that you read every new announcement, mostly on

- developers.google.com
- support.google.com > search
- webmasters.googleblog.com

Then:

- For heaven's sake, do not read any SEO blogs about the same topic
- Do nothing! Don't overreact based on a single event
- Look for trends (in your traffic, in SERPs, in future or past announcements)
- Think about what you could do better based on the trends and the new information
- Do better.

I could go on about this topic for hours and hours, or pages and pages. It would be an entertaining read but not very valuable for you. You just need to remember:

- Google is a marketplace
- You cannot reverse engineer the marketplace
- You can, however, look at what goes in (search demand, your inventory) and what comes out (SERPs, traffic)
- Algorithm updates are not immediately actionable
- Look for trends, and...

Don't read SEO blogs!

CHAPTER 4: THE SEO PROJECT

Every SEO project has two parts: the SEO platform and SEO processes. In this chapter, we will cover in detail the SEO project including the platform, the importance of your site speed, URLs, targeting, and targeted pages.

Establishing the SEO platform is mostly, but not entirely, a technical task. It focuses on optimizing the technical infrastructure and application so that it allows for scalable SEO instead of hindering it. Although you can always improve the platform, at some point it will be good enough to enable sustainable traffic growth via search. Before we get into the details, it's important to remember:

- everything accessible pre-login is within SEO scope
- everything exclusively accessible behind a login is not within SEO scope

And, for a healthy SEO platform you will need:

- targeted pages
- distribution (sitemap.xml)
- a startpage

Everything else is “deadweight” and should either not exist or not be indexed.

User Experience

Remember my favorite rules from chapter three? Always start with the user! If your users come to your site via search but don't use your site and bounce back immediately, you a) probably did not create much value for your users, b) did not create any value for your business, and c) will not be able to sustain this traffic for very long.

User Experience is not SEO. It's a massive discipline with lots of different goals and techniques, and it's still evolving. From a pure SEO perspective, you should always aim that your user experience is better, or at least on par, with that of the competition. As you probably don't have the usage metrics of their websites, it's best to look at their site. Ask yourself what they are doing better, assess if it would make sense on your site, then copy and iterate on it. That is the way the internet works.

A Red Line

A site with a bounce rate (where users are coming from search and are only viewing one page in one session) of more than 70% will not survive. That is the red line. If you have a bounce rate of 70% or above, do not waste time on SEO. Focus on fixing your site and your value proposition first.

That is the red line because it means that fewer than 30% of your users view more than one page. It means seven out of every ten visitors do absolutely nothing, leaving immediately and generally go straight back to Google and on to your competitors. But even then people say “What if they bounce forward or do nothing? Aren’t only the bounces back to Google harmful?” My answer is, “Yes, you are right. If the user stays on your site or bounces forward to another site, this is not a negative SEO vote – but it still means your site sucks!” If you have a bounce rate of more than 70% in any vertical, then it’s not the user, it’s not the vertical, it’s you. If you insist that you only have such a high bounce rate because the “goal of the user was fulfilled,” then I say you must work on your value proposition. Nevertheless, I have never seen a website that couldn’t lower its bounce rate below the red line with some simple user experience improvements.

In general, the UX of your site should get better and not worse. Better means:

- a gradual decrease in bounce rate
- a gradual increase in time on site
- a gradual increase in average pageviews

Note: These are metrics that you can easily measure and act upon. That’s why they are important. These are not the metrics that Google uses. They have their own way of collecting UX data, data that we do not have. Google does not use Google Analytics to measure the UX values on your site. So no, it does not help to “tweak” GA to show better usage values, that is just lying to yourself, and lying to yourself doesn’t work in the long term.

To achieve better UX values over time, you need to deliver direct feedback to the people that work with your site (your marketing, development and content teams). Give those people access to your metrics and make it part of their job description to watch, respond and improve these valuable numbers. Unfortunately, I often see that it’s the other way

round. Usage metrics get gradually worse with every release or new feature. You must always strive and fight to get better UX, and the people that work with the site need to know and care about THEIR site's usage metrics.

So, what the most important aspect of the user experience? Quite often it's speed.

Speed

Speed is something we must deliver. Speed is part of UX and brand experience and a crucial topic in its own right. Or from a Google point of view: Google does not like to send users to slow pages, as the users would use Google less. And Google does not like that. Google can measure speed both directly and via the user experience feedback loop. Fast websites perform better in search. They have better UX and deliver a better brand experience. Slow websites are synonymous with having a "clunky, old, outdated" brand and will always underperform. So, what exactly is defined as "fast"? Here is what you should be aiming for on desktop and mobile sites (both HTTP and HTTPS):

- first byte: < 200ms
- start render: < 700ms
- render above the fold: < 1000ms

The first byte is the time it takes for your server to return something to the user's browser for the first time. The start render is the moment when the browser first starts to render something visible to the user, so they know something is happening. The render above the fold is when the visible content - without scrolling in the browser - has been correctly rendered.

You will have a hard time to get the first byte of < 200ms on mobile, but you should still achieve the one second above the fold as there is much less fold. The most important metrics are the "start render" and "render above the fold" times. However, the start render can be misleading because the first visible content is not necessarily useful to the user (for example, it may just be a background color).

There is an acceptable zone if these web performance values are doubled. It's not fast, it's not slow, it's OK-ish, but you should and could do better.

So, what are “OK-ish” speed metrics?

- first byte: < 400ms
- start render: < 1.4 sec
- render above the fold: < 2000ms

Anything above these values is “kaput” - a great German word meaning completely and utterly broken. Your site will never achieve sustainable, growing traffic.

So, your goal from an SEO perspective and a UX perspective is to have a fast site, not an OK-ish one. If you have a slow website (anything above the acceptable values), then stop everything and fix your site, because your site is broken. In this case there just is no more important business decision you can make right now than to make your site fast.

So how can you test the speed of your site?

www.webpagetest.org is a Google supported project. Just enter your page URLs (obviously not only the startpage but all kinds of pages on your site), choose a location and a browser and start testing. Note that “render above the fold” is basically the same as “Page Speed Index” on webpagetest.org. Other very important webperformance tools are

- Google Page Speed Insights
- Google Lighthouse

as there is not one single tool, not one single metric that can capture the whole role of webperformance for user experience.

Basically Google Page Speed Insights tells you how many webperformance best practices are in place on the webpages you test, and if there is enough data: how fast users experience your webpages. Google Lighthouse - included in the Developer Tools of every Google Chrome browser - calculate metrics how fast mobile users can view and interact with your website.

And how can you make your site faster?

To get you started, I recommend this approach: Get rid of onpage deadweight! Delete anything you do not need on your web pages. Unnecessary icons? Delete. Small

decorative images? Delete. External fonts? I hate fonts, burn them. If you have third party rubbish that nobody uses, get rid of it. Evaluate everything on your page with a critical stance, and if necessary, delete whole page sections. Speed is more important than elaborate design.

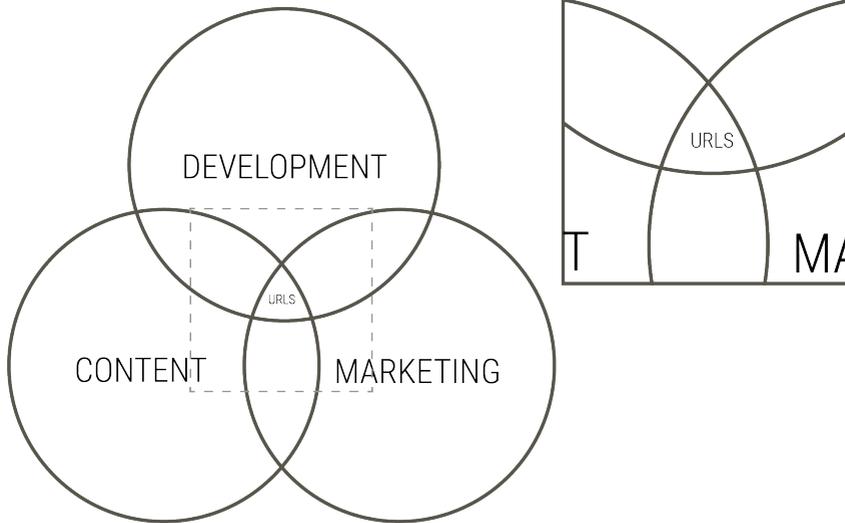
Note: you can have a fast site with fonts and icons, but it is hard. If you are suffering from slow speeds, my recommendation is to delete that rubbish. Your users do not need it, and it's holding you back. Once you have a fast site, you can think about adding fonts and other stuff that you need because of "reasons." You can test a variety of your pages and get recommendations on the steps to take using Google Page Speed Insights. Proceed with Google Page Speed Insights and read all the recommendations of this tool and do all of them, even the minor suggestions - you will find they add up to substantial improvements. After each improvement, check your speed with webpagetest.org. Do not stop until each and every value is in the "green" on Page Speed Insights. Page Speed Insights is the most direct way that Google communicates to you if you have a speed issue. After you have fixed all this, go to Google Lighthouse, start an audit and follow the recommendations in the "Opportunities" and "Diagnostics" sections. Don't be afraid of the blue, underlined "Learn more" links. Click them, learn more, fix things.

Always make sure that your developers (server, backend, and frontend) care about speed. A new feature, a new design, a new database server must never make the site slower. *If you have to choose between a new feature and speed, choose speed!* The speed topic will stay with you as long as you run a website; it will never end. Having a fast website is a significant competitive advantage, having a slow website is online business suicide. Next up: URLs - be sure to get them right.

URL Rules

SEOs love to talk about URLs. Why? URLs are at the intersection of development, marketing, and content. The URL is analogous to the API endpoint of your technical application. It's a defining part of your website. All URLs on a webproperty must be according to the rules listed below. Follow these rules, and you will keep your site manageable, logical, and prevent naming collisions. You will also be able to analyze

different site sections using Google Analytics and Google Search Console.



URL Rules

- 1) URLs must be unique.
- 2) URLs must be permanent.
- 3) URLs must be manageable.
- 4) URLs must be scalable.
- 5) URLs must be short.
- 6) URLs should include a short variation of the targeted page.

The URL rules listed above are by culminated priority. That means rule number one is more important than rules two to six put together; rule number two is more important than rules three to six put together and so on. Rules number five and six are always a tradeoff, with six being the least important rule. Search engine optimized URLs must fulfill all URL rules to be truly optimized. Let's discuss each rule in more detail.

URL Rule 1: URLs must be unique (one URL = one resource, one resource = one URL)

A URL points to one and one thing only (it is a one to one relationship). A URL does not

point to two things (it is not a one to many relationship), and one resource (for example, an HTML page) has only one URL. Remember, this rule is more important than all other URL rules put together. If you do not have unique URLs, you cannot systematically optimize your website.

URL Rule 2: URLs must be permanent (they do not change, they have no dependencies on anything)

Once established, your URLs must not change. They must point to one thing and one thing only. However, it is sometimes unavoidable that URLs change. Be aware, if you change a URL, you are, from an SEO perspective, deleting the old page and creating an entirely new page with the same content. You are essentially starting from scratch again. So make sure that URL changes are extremely rare. The good news, as you probably know, is that you can migrate from one URL to the another. For this, you have to point a HTTP 301 permanent redirect from the old URL to the new URL (just one direct redirect, not redirect chains). Then you have to change all links (within your direct or indirect influence) to point to the new URL. A successful migration will have no permanent negative SEO impact. Nevertheless, most migrations for complex medium or large web properties are not successful. Therefore, the only good reason to proactively change URLs is when you cannot guarantee URL rule number one (uniqueness). In such cases, it's best practice to move to a completely search engine optimized URL logic altogether and then stick to it. This rule is more important than rules three, four, five and six combined.

Pro-tip: Never have categories in the URL.

It's perfectly fine to have categories or any other taxonomy or hierarchy as part of one's information design, but they do not belong in the URL. Why? Categories always change. There is no such thing as a perfect taxonomy. You can promise, swear and even believe that your categories will never change, and then six months later the categories change. Any kind of hierarchy in the URLs are a major dependency and will lead to a lot of headaches, work and SEO misery down the road.

URL Rule 3: URLs must be manageable

Keep it simple. One URL logic per site section. No exceptions, ever, and I mean it. You

should know looking at a URL exactly what it is about, why it exists and why the URL looks the way it does. This rule is more important than URL rule four, five and six combined.

Manageable equals measurable. A bonus of having manageable URL logic is that you can easily make statements about parts of your site (using Google Search Console, Google Analytics and internal logs), rather than your whole site.

Rule 4: URLs must be scalable

There must be no limit on how many URLs you can create with one set of URL logic. It must not matter if you have ten or ten thousand of these pages, the URL logic must still work. This rule is more important than URL rule five and six combined.

To be manageable, measurable, and scalable, all pages URLs should include a short namespace that hints at their page type. Aim to give every page type its own namespace. For example:

https://www.example.com/a/understanding-seo (where “a” is an article page), or
https://www.example.com/c/seo (where “c” is a category page), or
https://www.example.com/t/stuff (where “t” is another page type)

Remember, do not write the whole name of the namespace in there (we want URLs to be permanent! As soon as you name things you have a dependency that might change over time).

URL Rule 5: URLs must be short

Do not confuse “short” with the “shortest possible.” The URL must only include the parts that are necessary to render the page. Nothing more, nothing less. “Short” is a fluffy metric, I know. If you want a hard metric: If your URL is longer than 70 characters, then your URL is probably too long.

Keep in mind: If you need an ID to render a page (because of URL rule three and four), use the ID in the URL. If you can - in any way - substitute the ID with a short variation of the targeted phrase (which we will be discussing shortly), then ditch the ID. If you

cannot get rid of the ID, then the variation of the targeted phrase is - most of the time - overhead. Most of the time, because if you manage to keep the ID and targeted phrase short, you might be good to go. Rule 5 is always considered a trade off with URL rule number six.

URL Rule 6: URLs should include a short variation of the targeted phrase

We will talk more about targeted phrases shortly, but know that this is the least important of all the URL rules. It's a tradeoff with rule number five (keeping it short), and that's why you should never include your whole targeted phrase into the URL. It should be a short sensemaking variation or only a part of the targeted phrase.

URL Best Practices (Make Your Life Easy)

Now that we've outlined the rules, here's a few additional best practices for URLs:

- URLs should always be lowercase
- only use the letters a-z, the numbers 0-9 and “-” in URLs
- always use a single dash “-” as a word separator
- never use “_” underscores, they are not considered word separators
- there should be no double spaces in URLs (“ ”, “-----”, “++” are always substituted with one “-”)
- there should be no spaces (“-”) at the end or beginning of a URL

Designing the optimal URL logic should be a one time effort. At the end of the day, even after you have dissected all the URL rules and best practices listed above, try to remember *you shouldn't overdo the f*cking URLs*. Implement URL rule one to four once so that they work for all eternity, and run automated tests to make sure that they do. Try hard to get rule five and six right. And then live with the outcome. See the “Best Practices for Targeted Pages” section for some URL best practice examples.

Targeted Page Generation Logic (a.k.a. Targeting)

Now that we've discussed speed and URLs, it's time to talk about pages that target user queries. If we want to get found for a query a user searches, we need a page that targets this query. This user's query becomes our “Targeted Search Intent Phrase”, or “Targeted

Phrase”: the phrase we want a specific page to be found for. Every targeted page targets exactly one dominant targeted phrase. I use the term “Targeted Phrase” instead of just “keyword” because “keyword” implies that we are putting keywords into a text, which we are not. Also, “keyword” is confusing as there is also the meta-keyword tag which is not good for anything. It is important that you realize that on these pages you are really targeting search intent, whereby we need to work with the textual representation of this intent.

For example, if you want to get found for [toyota insurance] (the *dominant targeted phrase*) you do not need to create extra pages for slight variations, such as:

- Toyota insurances
- Toyota car insurances
- Toyota car insurance

“Toyota insurance” is the dominant targeted phrase that you need to worry about. The others are reasonable variations, and Google is fairly good at handling singular and plural, as well as determining that Toyota is in most markets mainly seen as a brand of cars. We will target the slight variations “together” with the dominant targeted phrase. Targeting each and every slight variation (especially singular and plural) of a term with multiple pages can lead to internal competition and a lower performance of these pages. We are not targeting the “keywords” but the search intent behind them. One search intent, one page targeted at one dominant targeted phrase plus slight variations.

Next up, never target misspellings. Google corrects them anyway, and deliberate misspellings are bad UX and will be perceived as spammy. Again, you are not targeting text, you are targeting search intent.

Continuing our example, “Toyota motorcycle insurance” might be another dominant targeted phrase that deserves its own page. It is your call whether something is a targeted phrase in its own right or whether it’s just a slight variation. If the search intent of the user is the same, then it’s a slight variation. If the search intent of the user is a different one, then it’s a targeted phrase in its own right.

The process of targeting involves creating an inventory (pages on your webproperty) for

existing or presumed search demand (user queries). The most common way to decide what queries to target are:

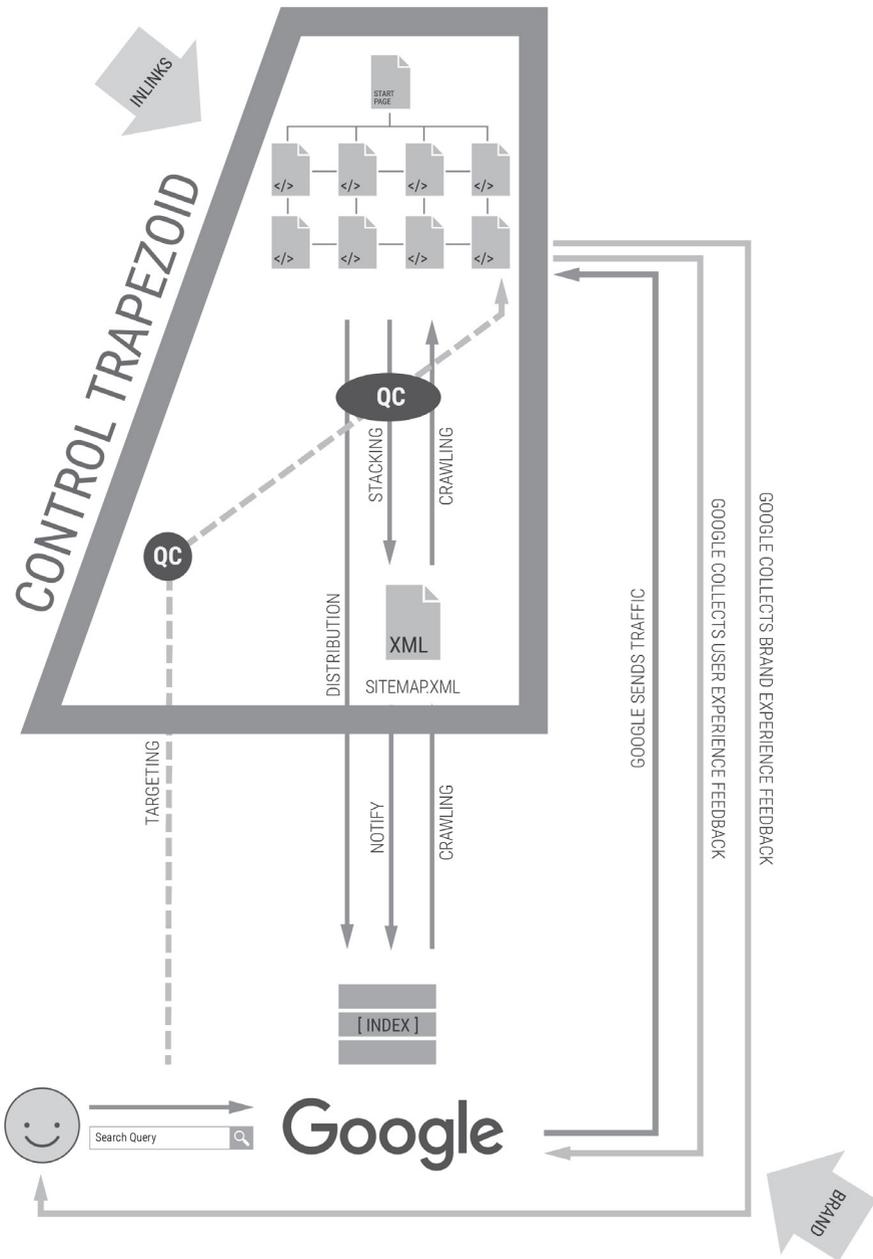
- Editorial Search Demand Research
- Database Driven Targeting
- Database Crossing Targeting
- Query-Driven Targeting

Before understanding how effective targeting works, however, you first need to understand The Quality Criteria.

The Quality Criteria (QC)

The Quality Criteria is an editorial or automatic process that must be applied to every page. The Quality Criteria is what differentiates targeted pages from pages that are a deadweight to your webproperty. And deadweight should either not exist or not get indexed. For example, do you have a page on your site that targets [viagra]? If you are not a spammer or not in the pharmaceutical business then probably not. Why don't you have such a page? Because it doesn't make sense for your business. You have no need for such a page. That is already some kind of editorial QC.

The QC can be applied either during targeting (deciding if a page should exist or not) or during distribution (deciding if we communicate a page to Google). It can be human (an editorial decision, which is a strong QC) or automatic (for example, based on the content of a page we decide if we create and distribute it or not). Regardless of how it is implemented, the important part is that you need QC! You need a quality standard for your website, and you need to apply this criteria to all of your web pages.



The primary job of the QC is to ensure that the page fulfills the quality standard of your company. You would not believe how many websites there are with a good quality product but a horrible, low quality, even spammy website. These websites hurt your company. For most new customers, your website is your company, and a positive first impression is crucial.

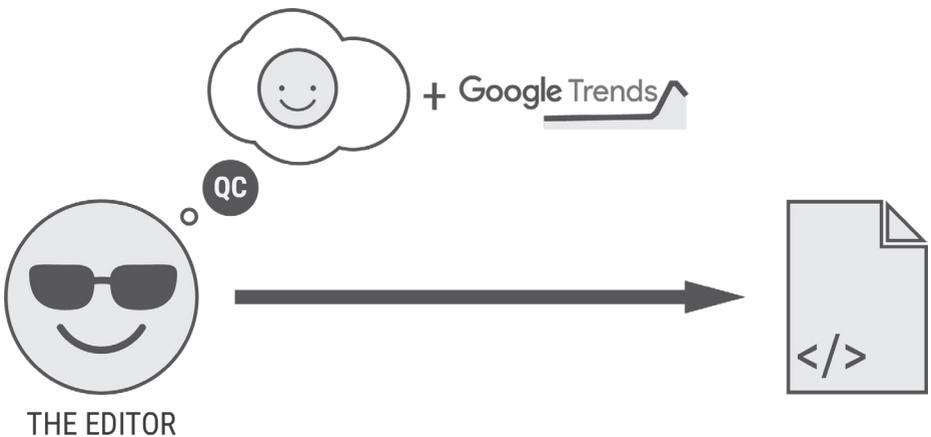
Additionally, the QC must prevent negative SEO impact from:

- off-topic pages
- low quality and spammy pages
- internal duplicate content
- internal duplicate targeting
- external duplicate content

We will revisit the QC constantly whenever we talk about targeting, content, and distribution.

Editorial Search Demand Research Targeting Page Creation

Editorial Search Demand Research Targeting Page Creation is the simple process when an editor (a content person, writer, editor, journalist or copywriter) considers what pages to create “for the user.”



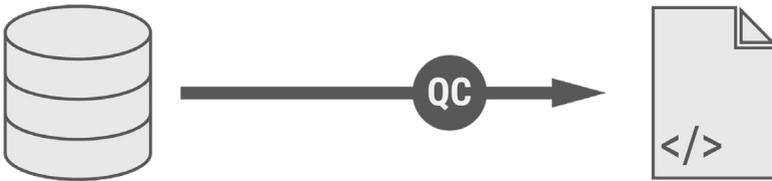
The editor tries to guess a targeted phrase and then checks with

- Google Autocomplete / Google Suggest (the dropdown of recommendations whenever you google something)
- Google Trends
- maybe even in some edge cases the Google Keyword Planner

If the hypothetical targeted phrase makes sense (and if the user actually does search this way), they then go on to create a page with the appropriate content. The QC, in this case, is very strong - it's a human concerned with the best interests of the company. The editor will go on to create unique content (by personally writing it), which is another strong QC. That is the targeting logic for most blogs, magazines, and editorial websites. It's a good, simple process but sadly it scales very poorly.

Database Driven Targeting Page Generation

If you have a database (DB) of content, you may be tempted to generate targeted pages for the data that passes your QC.



For example, if you have a database of business names and addresses, you could decide to create a simple company listing site (like Yellow Pages). You determine that for every database entry, you will create a page targeting that business name. Of course, you apply some minimal QC - if there is no address or telephone number, you don't create a page. Additionally, if you already have a page for that business name, you don't create a new page (there are always a lot of duplicates in big datasets).

Therefore, your QC logic would be:

- business name + telephone number = new page

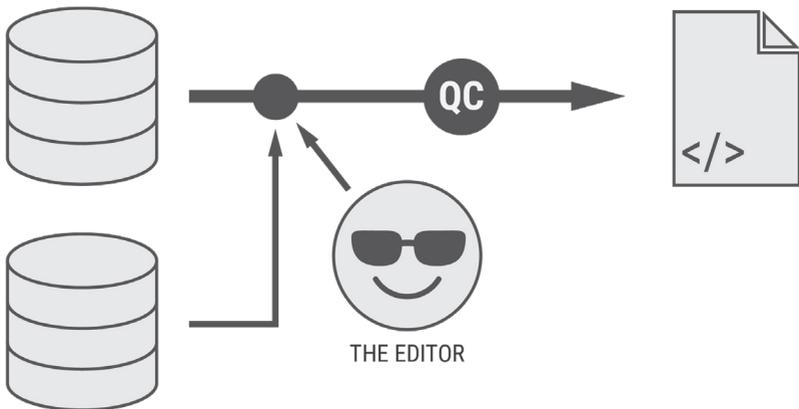
- don't create duplicates

This simple logic completely ignores the case of external duplicate content. If you have “fat content” (that is paragraphs of text such as business descriptions as opposed to just a list of opening hours), you need another quality criteria to filter out all the external duplicate business descriptions. As an alternative, you could have another QC that guarantees that no external duplicate “fat content” exists in the database.



Database Crossing Targeting

The advanced version of database generating content involves combining multiple databases, filtering it through quality criteria and then creating lots and lots of targeted pages.



For example, if you have a set of categories and apply it to your inventory of other content, such as articles, offers, and addresses, you could create a targeted page for each category if the page fulfills a pre-defined QC. A sample QC could be that every category page must have at least three entries. Otherwise, it will not get created.

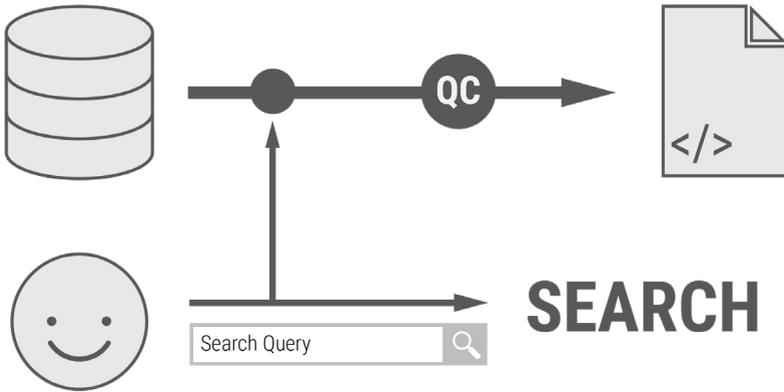
Let's do an example by revisiting our simple business listings site again. We have a DB with business addresses, a separate DB of categories (for instance, broad business types) and a DB of locations (for example, a list of cities). We apply the locations to the addresses so that we group the businesses by location. Additionally, we apply our categories to the location enhanced businesses; now we can create targeted pages for all categories for all locations, for example, "Copy Shops in Berghausen, Austria." Great, at this point we have spammed ourselves with an unbelievable amount of no-content pages (there are no copy shops in Berghausen). That's because we did not apply any QC yet. Ah ha! A sample QC would be that we only create pages where there are a lot of businesses, for example, "only create pages with at least 15 addresses" (a content dependent QC). Note: the number is not the important part here, it can be 2, 10, 20, 100, whatever makes sense. Generally, it's a good idea to start with a high QC and lower it over time.

There are many creative ways to combine databases to create targeted pages, but *never do it without applying hard QC to these pages*. If you cross databases without explicit, outspoken and specified QC, you will spam yourself, and you will suddenly have millions of deadweight pages.

Query-Driven Targeting

Query-driven targeting is very easy to do poorly, and very hard to do right. With query-driven targeting we can use real user queries to create targeted pages. We take a user search query - from any source we can get our hands on, for example, the logs of our internal search, and apply it to our database (so, we make a search of our available material). If the page fulfills a particular requirement (for instance, there are more than x items found, and there is no other page on your site that targets a similar term, and there is no page on your site that shows the same list of items), then we create a targeted

page. If this quite complex QC is not fulfilled, then don't create the page.



Some call these “search pages,” and yes, search pages are query-driven targeted pages but without a QC. If you do it wrong, you will create:

- internal duplicate content
- internal duplicate targeting
- nonsensical pages
- pages with terms that do not align with the goal of your business

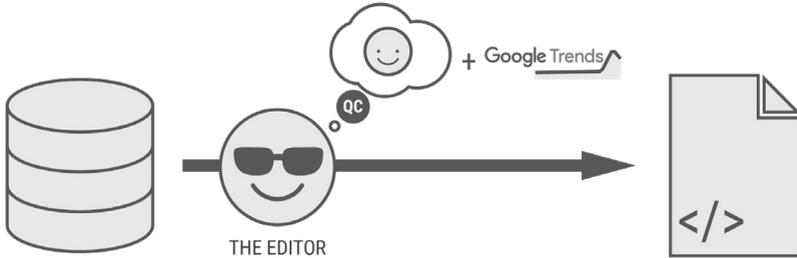
To get this right, make sure you use QC on all queries you target.

Pro-tip: Query-driven targeting is very, very, very hard to do right

Start slow. Begin with a very high QC, a fixed set of queries and lots of quality assurance. Learn, lower the QC, add more queries, learn, and continue. If you don't do it this way, you will spam yourself.

Database Editorial Targeting

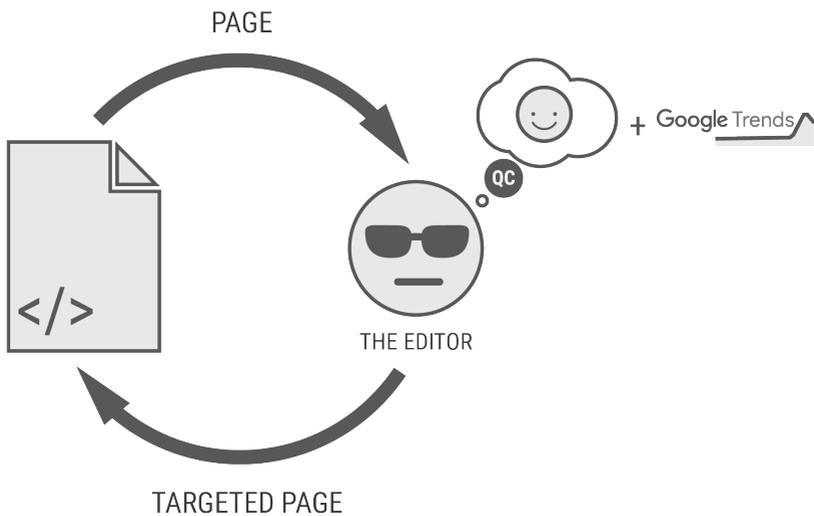
Database Editorial Targeting is another form of targeting, and there are a lot of creative and not-so-creative ways to do it. With this technique, you have an editor who creates list pages (via a topic, theme, or collection) through an editorial filter over a database. A



common use case is the creation of topic pages. I.e.: an editor decides, that there should be a page listing all articles for “mothers day”. He/she goes into the backend and creates a targeted list page for it.

Retroactive Targeting

Retroactive targeting is the process of taking an existing page, so a page that is not targeted at a specific targeted phrase, and retrofitting it to make it become a targeted page. Retroactive targeting is one of the reasons why many people hate SEO. There are a lot of good reasons why pages might exist. An article might be written with a social media audience in mind, or a page might exist because of the top-down navigation (people coming from the startpage and clicking through the navigation). These pages were created to fulfill a particular use case. If you now take a page that fulfills a particular use case and impose a search use case on top of it, you will get a page that fulfills neither use case. In general, you have made the UX of your site worse. Yes, you can, of course, create a page that fulfills a top-down navigation use case and a social media use case and the search intent use case. But in nearly all situations it's not as easy as updating an existing page with a title, headline, and some extra content and then calling it "search engine optimized." It doesn't work like this. If you want to have a targeted page that fulfills top-down navigation and another use case (which is a good idea), then this must be done during the creation of the page, not as a minor, later add-on.



Targeted Pages

So far we've discussed how to go about creating targeted pages, and now we are going to talk about what they consist of. A targeted page is a page which:

- 1) is a HTML page
- 2) is reachable via only one unique URL
- 3) is targeted at one - sitewide unique - dominant phrase
- 4) fulfills a presumed search intent
- 5) fulfills a business utility
- 6) hosts unique content or a unique collection of content
- 7) fulfills a predefined quality criteria (QC)
- 8) displays content and links visible without user interaction
- 9) uses exactly one language
- 10) links to other targeted pages
- 11) gets linked from at least one other targeted page
- 12) is listed in the sitemap.xml.

Let's discuss all these points in detail.

1. A targeted page is a HTML page

Every targeted page must be a HTML document. Even though Google can and does index Flash and PDFs, we only can thoroughly optimize plain old (JS and CSS enhanced) HTML pages. If you deliver the same content via PDF and HTML, the user and Google will nearly always prefer the HTML version. Only HTML pages qualify as optimized (and optimizable) pages.

2. A targeted page is reachable via one unique URL

As URLs are visible in search results, it's one of the primary things the user sees and learns about your page. In addition to this, URLs are actively communicated in PR and marketing, and by support and sales, both online and offline. And when other websites link to you, they will usually copy and paste the URL. URLs are part of the content, and a sensemaking URL should give the page context, in a way similar to a headline. URLs are a major technical decision determining what pages can exist. A targeted page is an indexed URL with content in context. One page, one URL. Also see the URL rules.

3. A targeted page is targeted at one sitewide unique phrase

The one sitewide unique phrase is what we want the page to be found for. There should just be one page for that specific targeted phrase and its underlying search intent on your entire webproperty, never more. If there are two pages on one webproperty that have been designed to be found for the same targeted phrase, it is called “internal competition.” Internal competition is confusing for Google and consequently bad for SEO because:

- Google cannot pinpoint the right resource on our website for the user query with absolute certainty, as there are different options.
- Google can pinpoint the right resource on a competing website (with only one page that targets the query) with absolute certainty.

Therefore, Google will choose the competing “certain” pages over our “uncertain” pages. There’s a chance we may show up in search results, but not as well as we would have with just one targeted page per sitewide unique phrase.

There are three common causes for internal competition. They are

- internal duplicate targeting
- internal duplicate content
- internal duplicate pages

Internal duplicate targeting happens when you create two different pages with different content, but that want to be found for the same targeted phrases. This strategy will underperform. Delete one page or merge them both into one page. Redirect (HTTP 301) the vanished URLs to the new one.

Internal duplicate content happens if you reuse substantial parts of your “fat content” (paragraphs of text) on more than one page. Google cannot lead the user to the one “right” page that content belongs to, and might prefer competing pages where it can easily pinpoint the “right” page. Internal duplicate content is no issue if it happens “sometimes” as Google is pretty good in leading the users to the right pages. However, if internal duplicate content happens systematically on your webproperty, you have a systematic problem. A gray area exists on whether or not internal duplicate content includes when internally reused content is combined with other content to create a unique combination

of content.

Internal duplicate pages, however, happens when you violate URL rule number one (URLs must be unique). These pages will always underperform. Make sure you fulfill the URL rules, then delete the unnecessary URL variations. A failsafe way to prevent unintentional internal duplicate content is the so-called “canonical tag,” which we will discuss later. Always remember that you should have only one page, with one unique URL, for one unique dominant targeted phrase: 1:1:1. It’s also necessary to note that every page we communicate to Google must have a targeted phrase. Otherwise, it’s deadweight, and either should not exist or not be communicated to Google.

4. A targeted page fulfills a presumed search intent

A page must not only want to be found for a particular term, but it must also meet a search intent. We don’t and never will know the exact search intent of the user. However, we should attempt to solve the “job to be done,” or at least point the user in the right direction. Keep in mind; a page can fulfill more than one search intent if the targeted phrase allows multiple search intent interpretations.

If your page does not solve anything for the user, do not create it. Pages that attempt to answer a “know”-query (where the user wants to know something) with a “do”-intent (complete a transaction), or vice versa will not work. However, it does make sense in a lot of cases to fulfill the possible “know” and “do” intent of a query on one page. Also, link further if you do not fulfill one of the presumed search intents of the users, even if it’s to your competition. Everything is better than leading the users to a dead end where the only way out is bouncing back to Google.

5. A targeted page fulfills a business utility

This is obvious but needs to be stated anyway. The page must make sense for your business and help you achieve your goals. Therefore only create pages that will contribute to your business goal.

6. A targeted page hosts unique content or a unique collection of content

Unfortunately, it’s not easy to explain what unique content is. However, it’s much easier

to explain what “duplicate content” is. We’ve covered internal duplicate content, so now let’s discuss external duplicate content.

External duplicate content is whenever you use copy and paste (or a similar process) and take content from another external, Google indexed (or soon to be indexed) web page and publish it on one of your indexable web pages. This can range from small scale copy and pasting to widespread automatic scraping. As you are not the first to publish this piece of content, you are left with a high probability that you will not be seen as the authoritative source for this content. In Google Search, this content has either negative or no value for you. We call this “first one wins.”

If you have lots of pages with external duplicate content on your website, not only do the pages with this duplicate content not perform but also all web pages of your site will be worse off. There is no fixed percentage of when minor external duplicate content becomes major, but for newly launched websites that do not have any brand recognition or lots of inlinks, external duplicate content must be seen as toxic. For more established webproperties it takes much longer until external duplicate content has a whole site negative impact. But external duplicate content pages are still practically useless from a SEO point of view.

Also, “first one wins” is the general rule, but it’s not guaranteed. If the content of a site is mostly external duplicate spam or rubbish, then even unique content first discovered on your site might not be seen as “your” content, as there is a high probability that you are not the authoritative source.

So what if we change or rearrange some words? This question is always asked when I talk about this topic. It does not matter, as soon as copy and paste are used in a paragraph of text it is a duplicate. On the contrary, if you read an article, rewrite it in your own words (without using copy paste) then this is, with a very high likelihood, not external duplicate content. But, what about citations? If the external duplicate content is just a small part of the overall textual content, then don’t worry. The issue is mostly about “paragraphs of text content.” Again, there is no fixed percentage of how much is too much. Sorry to be vague but there are no precise answers here.

However, duplicate content does not apply to “structured content.” Structured content is anything and everything that is key-value pairs, definitions or tables. For example, addresses, telephone numbers, and recipe ingredients. For such structured, factual data there cannot really be an authoritative source, so they don’t contribute to external duplicate content.

There is no negative value in embedding external duplicate images, as images are usually duplicated, but it also has no positive value. It’s a lost opportunity if you embed soulless stock photos instead of showing some creativity (i.e. just use a picture you snapped with your mobile phone).

The conclusion out of all this: External duplicate content on your site has no or negative value for your webproperty. So don’t do it.

Now that we’ve discussed what external duplicate content is, we can return to point number six: A targeted page must host unique content or a unique collection of content. Unique content is the content produced when you are not duplicating external or internal content.

If you combine content - unique, external duplicate, internal duplicate, and structured content - in a new way, then we talk about the “unique combination of content.” Lists of data are fine if a user query asks for a list. For example, [used cars liverpool] is almost certainly looking for a list of used cars (or used car dealerships) in Liverpool. Each list item probably consists of duplicate content, an item that is listed in full on some other page, regardless of whether is it on our site or external. Nonetheless, the list is a unique combination of the content itself. Sometimes these unique combinations of content pages are not in the form of a list. Then we speak of aggregated detail pages, whereby the source of the aggregation can be external or internal. They might be intended to look like editorial articles (but really aren’t), often with multiple short paragraphs of text (so-called text-templates), datasheets or other techniques. Unique (or editorial) content and unique combinations of other content are often mixed on one page.

The major issue regarding a unique combination of content is always utility. If the page makes sense, if the search intent of the user gets fulfilled and if the user has a good user experience, then a unique combination of content is a valid and effective way to drive search traffic.

In my experience, some types of aggregation (for example, generating an article) are very, very hard to do right and very easy to do badly. Most aggregated detail pages are rubbish and must be seen as pure spam. Making sense-making list pages is relatively easy. To be safe, always assure that point number four (a targeted page fulfills a presumed search intent) is always met.

7. A targeted page's content has to fulfill predefined quality criteria (QC)

In addition to the page content being unique, it must also satisfy your predefined quality criteria (QC). It is vital you identify your QC, make sure it gets fulfilled during content creation, and always double check it.

Some parts of the QC are easily quantifiable, for example, a list should always have x number of items; a page must always have at least x pictures and y bullet points, or an address detail page must always display the address and telephone number. But what about the QC that is harder to check? For instance, how do you make sure lists A and B are not the same lists (resulting in internal duplicate content)? With an editorial page QC is relatively easy to do, but with automated pages and potentially millions of targeted list pages, it's significantly more challenging. Just because it is hard does not mean you can skip this point. If you do not check for that, you will inadvertently create massive amounts of internal duplicate content.

8. A targeted page displays content and links visible without user interaction

Contrary to popular belief, Google does render CSS and execute JS. Google attempts to emulate what the user sees, but only what the user sees without clicking or touching counts as valuable content. Content hidden after any user interaction (other than scrolling) is usually devalued for Google and must be seen as wasted (this includes

content in sliders, tabs, accordions, foldout questions and answers). So, we must care about the design of the site and that all our valuable content is visible without user interaction, but we must also care about the HTML markup of the site. The HTML of the page should be coded so that Google can easily identify what the different parts of the page are about.

You cannot show the user one thing and Google something different.

Additionally, the first time visitor referred via Google, must be treated exactly as Google is treated. Everything else is called cloaking and is not a viable long term strategy. As soon as the user leaves the first page and moves within your site, you can do whatever you want as long as you are not misleading. So if you choose to deliver a crappy second pageview user experience, your overall time on site and average pageviews will probably not be up to your competition, and you will not perform for long.

Note: Google used to only use IP addresses from the US and sent no language HTTP headers (Google was American and had no preferred language). That is no longer the (only) case, and so even web pages with geolocation and language detection features must now be renderable for Google. You must treat Google like any other first time visitor.

9. A targeted page only uses one language per page

One page, one language. If your targeted page mixes different languages on one page, it just will not perform. Your main content and your boilerplate (everything like navigation, teasers to other web pages, and other supplemental content) must be in the same language. For example, if your article is in German and your navigation is in English, then your page will always underperform. Similar to the previous point, if you have major “other” language content in the markup of your page (even though it’s not visible on the page due to CSS and JS), this “other” language content will still hinder your performance in Google. However, do not worry if you make a foreign language citation or have some foreign language terms within your content. There is no fixed percentage when something is too much. And do not confuse this with one website, one language. Having a multilanguage site is great, just don’t put more than one language

on one page.

10. A targeted page should link to other targeted pages and 11. A targeted page should be linked from at least one other targeted page

Every targeted page needs at least one link from another targeted page – and vice versa, every targeted page should link to other targeted pages. Pages that do not have a single site-internal link pointing to them do not perform. Pages that do not link to other pages are a one way street and only leave the user with the option to bounce back to Google. You can either have a very sensemaking deliberate logic or simply use serendipity to achieve these goals. As we will see in, “Best Practice #13: List of links to other targeted pages”, I recommend a mix of sensemaking and serendipity interlinking.

12. A targeted page must be listed in the sitemap.xml

Every targeted page must be distributed in a manageable, measurable and scalable way – in other words, every targeted page must be listed in your sitemap.xml. This is important, and we will discuss this topic in the “Distribution” chapter of this book.

Best Practices For Targeted Pages

Besides these requirements, which every targeted page must fulfill, there are some best practices every targeted page should aim to achieve. These best practices are based on my own experience, and I have seen them work with many different clients. Some of them may change if the market (Google, the user, the competition) changes so use them as a starting point, then start to establish your own best practices. *Unquestioned best practices become worst practices over time!*

Best Practice #1: The Targeted Phrase

Define (write down) the targeted phrase of each targeted page. Already very much discussed, every page needs one. There is no single HTML element where you state “this is the targeted phrase” – it is a CONCEPT. The targeting elements listed in this section are in order of priority for their targeting use:

- 1) Title
- 2) H1
- 3) URL
- 4) Above the fold image filename and alt-text
- 5) Anchor text

It's not about stuffing the targeted phrase again and again into these elements - it's about using reasonable sensemaking variations. If a sensemaking variation of the targeted phrase shows up in the URL, title, H1, the image filename and alt-text you can be pretty sure that Google has got the message and knows what the page is about. The targeting elements are a way to make sure that Google knows what the page is about.

Note: Whenever we talk about a sensemaking variation or a short variation of the targeted phrase then partials, synonyms, different (but correct) spellings (do not target or use misspellings), separated and together, singular and plural are all okay. Do not repeat the targeted phrase again and again in the targeting element but, look for, and think about, sensemaking variations. There is a tiny percentage of cases where there aren't any sensemaking variations, but generally, if you can't think of variations, then you aren't thinking hard enough. It's also possible that sometimes we refer to the targeted phrase also as the topic of the page. The topic of a page is used in a wider sense as the targeted phrase (for example, if the targeted phrase of a page is [understanding SEO] than the topic is basically "SEO").

Best Practice #2: Presumed Search Intent

Brainstorm and write down the presumed search intent. If you cannot or will not help the user in some way on their quest, or do not know what the search intent behind a word combination is, then do not even bother to create a targeted page for it. Remember, you are not just targeting words, you are targeting the user. What does the user want from this page? If your user bounces back from your site in disgust, you will not keep that traffic (or the user), and you hurt your future overall performance as well.

Even though this seems quite clear, it is quite often that I come across non-performing "landing pages" created by "renowned SEO agencies" that do everything right but do not solve the "job to be done" of the user. If these pages do not perform, then this is not

because of some esoteric “Google penalty” - it simply means your pages suck.

Best Practice #3: Business Utility

Define and write down the business goal of the page. You must see the value in the traffic that you are targeting. Does it make sense for your business to have specific targeted pages? If you get a lot of traffic, even sustainable traffic, but this traffic has no value for you as it is not your topic and the users will not now, not ever, contribute to your overall business goal, it means these pages do not make sense!

Best Practice #4: HTTPS

Google is now pushing strongly for websites to adopt SSL (a.k.a. https:// with “s” like secure). They announced it as a direct factor for better search traffic performance. From an SEO, user, and technology point of view, HTTPS is the better solution as it’s the safe bet for the future. You should make sure your http:// URLs redirect (serving a HTTP 301) to your https:// URLs. This can be configured across your whole website. Make sure your website is verified with the https:// URL in Google Search Console. Also, google [HTTPS google] for the specs.

Best Practice #5: URL

The URL is one of the targeting elements. To be seen as a truly search engine optimized URL, it must fulfill the URL rules we determined earlier, including the least important one: URL rule number six (URLs should be a variation of the targeted phrase).

So for example, out of these two URLs:

- 1) *<https://www.example.com/content-page/index.aspx?id=0000001>, and*
- 2) *<https://www.example.com/d/understanding-seo>,*

you can be fairly sure that the user and Google both know the second URL gives more context than the first. URL rule number five (URLs should be short - but not necessarily the shortest possible) is a tradeoff with URL rule six. That means if you are in doubt, use only a sensemaking part of the targeted phrase in the URL. Therefore:

<https://www.example.com/d/seo-business>

is better than

<https://www.example.com/d/seo-is-all-business-activities-with-the-goal-of-getting->

found.

If you have a domain name that is part of your targeted phrase, there is no need to repeat it in the URL slug. For example:

https://www.books.com//fishing-books

is not necessary as

https://www.books.com//fishing

is fine. It is better to use a short variation, even if you lose minor context.

Another important aspect of whenever you write a short variation of the targeted phrase into the URL is that the least important URL rule (URLs should include a variation of the targeted phrase) does not violate the most important URL rule (URLs must be unique). So for example, if

- *https://www.example.com/a/1239258295/systematic-SEO* is equal to *https://www.example.com/a/1239258295/SYSTEMATIC-SEO*, which is equal, to
- *https://www.example.com/a/1239258295/sysTeMATic-SEO*, which is equal, to *https://www.example.com/a/1239258295/whatever*, you have screwed up big time!

Whereby the targeted phrase in the slug doesn't determine the content of the page. Remember, 1 URL = 1 resource. My recommendation for such cases is to create another DB that matches "systematic-seo" to "1239258295" so basically URLs such as *https://www.example.com/a/systematic-seo*

become possible. Another common way is to make HTTP 301 redirects from

- *https://www.example.com/a/1239258295/whatever* -> HTTP 301 -> *https://www.example.com/a/1239258295/systematic-seo*
- *https://www.example.com/a/1239258295* -> HTTP 301 -> *https://www.example.com/a/1239258295/systematic-seo*

But, well, this fix is only halfway - neither is the URL as short as it could be. Additionally, the other URLs are still possible; they just get redirected. At one point you will have lots and lots of internal redirects (which slows down the user on the site). This is all because someone just wrote some internal link logic wrong, but due to the redirect fix in place, you do not recognize and change this behavior immediately. (Also you communicate more internal not-targeted-pages-but-redirect URLs which slows down

your distribution). Don't catch too much stuff with redirects, if the URL is wrong just let it fail (404) and fix the root cause (the wrong interlinking).

Last but not least, make sure that your URLs fulfill URL rule number two: URLs must be permanent. That means they do not have any dependencies once they are created (i.e. the headline of the page might be used to recommend a URL, but once the URL is created the headline can change a thousand times, and the URL stays the same). *All targeting elements must be independently editable!* Changing one does not change the others. Also, don't use categories in URLs as there is no perfect hierarchy. Categories always change over time, violating URL rule number two.

One last thing (and yes, we SEOs love to talk about URLs): the URL also shows up on the Google Search Result Page. Therefore, a sensemaking URL that aligns with the search query of the user can increase click-through.

My simple URL formula:

For pages where we have complete editorial control and no dependencies: `https://www.example.com/%short one or two letter namespace%/%short variation targeted phrase%`

- `https://www.example.com/d/understanding-seo`
- `https://www.example.com/l/vienna`

For pages where there are dependencies (i.e. user generated content, external database, ...) `https://www.example.com/%short one or two letter namespace%/%unique id%`

- `https://www.example.com/p/12345`
- `https://www.example.com/f/0001xd`

Best Practice #6: The Canonical

The canonical element identifies the preferred URL of a web page to Google. The canonical is a very powerful tool. If you make a mistake within your canonical logic, there's a chance you destroy your traffic. Nevertheless, with just one canonical tag, you can minimize a variety of potential problems.

The canonical is incorporated as a `<link>` element in the `<head>` of the HTML of your page, and looks something like this:

```
<link rel="canonical" href="https://www.example.com/d/understanding-seo">
```

The canonical gives dominance to the URL the page should have, regardless of what is actually in the address bar of the browser. It is a failsafe way to prevent 80% of everything that can, and will, go wrong with your URL. So, therefore:

- It avoids internal competition with URLs with additional query parameters:
URL: *https://www.example.com/d/understanding-seo?op=345&source=twitter&foo=bar* -> Canonical: `<link rel="canonical" href="https://www.example.com/d/understanding-seo">`
- It avoids problems with different subdomains gone wild:
URL: *https://staging.example.com/d/understanding-seo* -> Canonical: `<link rel="canonical" href="https://www.example.com/d/understanding-seo">`
- And, even despite your best efforts, underlying artifacts (such as page IDs), can make it into URLs:
URL: *https://www.example.com/content-page/index.aspx?id=0000001*
-> Canonical: `<link rel="canonical" href="https://www.example.com/d/understanding-seo">`
URL: *https://www.completelt-other-domain.com/d/understanding-seo*
-> Canonical: `<link rel="canonical" href="https://www.example.com/d/understanding-seo">`

The canonical is page content dependent, not URL dependent. It always shows the URL the content should have, not what it actually has. If Google meets the same page or content via different URLs, Google will summarize all page metrics under the canonical URL. In a way, you can think about the canonical as a redirect that only happens for Google.

Even if there is not a URL variation, the canonical points to itself. The canonical is as static as the content of the page. Make sure that you always include the protocol and the domain in the canonical, too, so that you can fetch (sub)domains gone wild issues (whenever your content is suddenly accessible via two or more different (sub)domains). Quality assurance and checking the canonical after every release is a must.

Known URL variations (for example, because of historical reasons) should always make a HTTP 301 permanent redirect to the correct URL (or just let the wrong/old URL run into an HTTP 404). The canonical is a way to deal with unknown URL variations, a

failsave against stuff, that might go wrong with the URLs.

Best Practice #7: Meta Robots

The meta robots tag is a way to provide Google with instructions on how to deal with your page. There are quite a lot of meta robots tags, so please R.T.F.M. We will just deal with the most important one. This is the instruction to allow Google to index your site: `<meta name="robots" content="index">`, and this is the instruction to tell Google not to index your site: `<meta name="robots" content="noindex">`.

By default, Google sees every page as set to "index" if there is no meta tag. I recommend that every page (type) has an explicitly set meta robots tag. It must be our decision if something is - from our perspective - worth indexing or not.

Any page that is not a targeted page or the startpage, is deadweight. If the page does not fulfill your QC, set it to "noindex." It would be much better not to have the page at all (HTTP 404), but sometimes there are reasons to have pages, even though they are not targeted pages. One good example is "search pages." It makes sense to have an onsite search engine, and it's also ideal for the user to be able to copy and share search page URLs such as <https://www.example.com/search.php?q=polly+wants+a+cracker>. However, you cannot confirm that the search results page fulfills your QC, or doesn't contain internal duplicate targeting or even internal duplicate content with other prior indexed search pages. Therefore, set these pages to "noindex."

Note: you can set the meta robots tag in the HTML of your site or via HTTP header. I strongly recommend setting it in the meta robots tag, as this is much easier for regular testing. People just look more often at the HTML code of a site rather than into the HTTP headers.

There is also another meta robots tag called "nofollow." Never use this. With a `<meta name="robots" content="nofollow">` you devalue all links on your page so that Google will not see the links as a vote. And, although `rel="nofollow"` placed on links makes sense in some edge cases, `<meta name="robots" content="nofollow">` applied to a whole page never, ever, makes sense as you are basically crippling Google's ability to understand your site. Just don't use it.

Best Practice #8: The Title

The <title> is *the most important* targeting element. The title gives the whole page content its context. It is visibly displayed in the Google Search results, and it's the first thing the user sees. Users need to click on the title to read your page.

Note: In this example “Example” is the awesome brand of a tour booking site.

TITLE

10 Best Asia Tours & Trips - Cool People, Awesome Vibes - Example

<https://www.example.com> > Asia tours ▼

250+ trips through Asia with 17023 reviews. Discover the thousands-years old history of China, the stunning beaches of Indonesia and the crazy nightlife of Bangkok. Book your trip now!

A few rules apply here. The title:

- must be sitewide unique
- must align with the content of the page
- should be about 80 characters long
- should include the targeted phrase, and
- should ideally include the brand of the website

A good format is this:

```
<title>%targeting phrase% - %brand%</title>
```

If you are lucky to have a rather short brand (three or four letters) then:

```
<title>%brand%: %targeting phrase%</title>
```

 is also a possibility.

If you have enough space (remember, you are limited to about 80 characters), spark up the title by adding elements that might improve click-throughs. For example: “15 Best Hotels in London - Hotels.com” (as people love numbers). Always make sure that your title is the most direct representation of the targeted phrase. If the title states something different than what you actually want to get found for, you are doing it wrong.

Because every targeted phrase must be unique sitewide, so must be every title. That's

why the title is a good way to discover duplicate pages or duplicate targeting on your site. In Google Search Console go to: Search Appearance > HTML Improvements > Duplicate Title Tags. There you will be able to see pages where there is some title conflict - a sign that something is wrong. Do not freak out about each and every duplicate title, but do solve them if there is a systematic issue on your site that is creating more and more of these duplicate pages.

Best Practices #9: The Meta Description

The meta description is pure marketing as it may show up on the Google Search result page. A good meta description can increase click-through from the SERPs. However, that said, the user will never read this text in its entirety. They will just scan it for a microsecond.

10 Best Asia Tours & Trips - Cool People, Awesome Vibes - Example

<https://www.example.com> › Asia tours ▼

250+ trips through Asia with 17023 reviews. Discover the thousands-years old history of China, the stunning beaches of Indonesia and the crazy nightlife of Bangkok. Book your trip now!

META DESCRIPTION

Other than that it has no impact on the page or its performance. It is not part of the content of the page. Rules for the meta description are as follows:

- it must be sitewide unique
- it should be two to four sentences
- it should be descriptive
- it should include a variation of the targeted phrase

I have experimented a lot with meta descriptions. The only thing that I can say for certain is that whenever you include a number in the meta description - one that aligns with the search intent of the user - it gets clicked significantly more often (think telephone numbers, prices, and street numbers). Sometimes simple unicode bullets like ✓ or similar might also come in handy. If you are too “salesy” in the meta description, Google will not display it to the user. In this case, change the meta description and wait until it shows up. This may take a while.

Best Practice #10: <H1>

The <h1> - the main headline of the page - is the second most important targeting element. It must include a variation of the targeted phrase - and be a good headline! Important: the <h1> must be the biggest textual indicator of what the page is about above the fold. As a matter of fact, it doesn't matter if it is <h1>, <h2> or <h3>. In the case that Google renders the page, it picks the biggest headline, and uses it as the main headline, no matter if it is <h1>, <h2> or <div>. But, as Google does not always render the page, just make sure that <h1> is the biggest headline, and that the headline clearly states what the page is about.

A clear headline is not only an official Google requirement, but it's also about user experience. If the user types in [staten island ferry], sees a search result with a big blue underlined link to "Staten Island Ferry - Example.com," they will most likely click on it. If the user then lands on a page which does not clearly state that it is about Staten Island Ferry, the user will bounce back. Pages with big, descriptive headlines above the fold that actually tell you what the page is about work better for the user, and therefore perform better in Google.

It's best practice *not* to have the title and the <h1> as the exact same text. On the other hand, they should not be completely different as they are both based on the targeted phrase. Using the example of the [staten island ferry]:

- Title: Staten Island Ferry - Example.com
- H1: Staten Island Ferry Ticket Prices and Timetable
- Incorrect H1: Get the next boat!

The <h1> does not have a length restriction, other than the ones that usability sets (a paragraph of text is not a good headline). There should only be one <h1> on a page (as an element and the way it is displayed). Therefore, there should never be a shadow of a doubt what the main headline of the page is.

Best Practice #11: Above the fold image filename and alt-text

The image filename and the image alt-text of the above the fold, main image is the fourth

most important targeting element. We are talking here about big, main images that are an essential part of your main content. We are not talking about thumbnails. (The main thing you should worry about when using thumbnails is to make the ones above the fold load blazingly fast, and, if they are linked, have an alt-text that fits the linked pages.) If, and only if, it makes sense for your page to have an image (and most of the time it does - people love images) and especially if the image has a lot of web page real estate above the fold (most of the time it does), then your image should have the following requirements (we'll take them step by step):

- The image should fit the topic of your page: Google knows what your image displays quite well. The advancements in image recognition by Google are very, very impressive. So displaying something completely off topic can hurt your performance. Never display any porn or erotic images, unless of course, you are targeting adult queries.
- The image should be unique: The main image of a site is usually above the fold and takes up the most real estate, so it should be unique. If you do not have unique images it's a lost opportunity, but not a major problem. So, what exactly is unique? You should be the first one to communicate this image online. So no *stock photos*, which is a good thing, because stock photos *suck*. This usually means you have to take your own photos. Which is easy - just use your mobile phone's camera. From an SEO point of view, and from a user's point of view, these images are good enough in most cases. And, they are much better than the 1,000 stock photos of [happy woman eating salad] (google it). Be creative: in a half day photo shoot, you usually can create one hundred unique, entertaining pics - all of them great unique content. On the other hand, if you use stock photos, you have wasted most of your above the fold real estate (what most users see, and therefore has the most value for Google).
- The image URL must fulfill the URL rules. There is only one small thing that differs between targeted page URLs and image URLs: the fact that you can be more relaxed about the "short" part. It's also fine for properties such as "big," "small," "300x500" to be used in image URLs.
- The image must have a descriptive alt-text that includes a variation of the targeted phrase: The alt-text gives additional context to the image, especially if the image is linked (as this one should be, see next point). Here, it's best practice to use

approximately five very short words that are at the same time descriptive, as well as a sensemaking variation of the targeted phrase. If in doubt go for more descriptive instead of using a non-sensemaking but targeted phrase (if you do image optimization right, these two properties aren't exclusive).

- The image should link to a bigger version of itself: Images are always indexed in the context of the embedding page. Google Image Search prefers larger images (at least one side of the image should be 1200px). If you want (and you should want) to be displayed in Google Image Search and Image Onebox (the Google Images Search embedded results in the normal Google Search), then you must link from your embedded version of the image to the bigger version. The image must be more or less the same content. Slight variations are okay (a small copyright notice, or showing a partial in the small images and the slightly fuller version in big), but showing two completely different images will not work (Google is not stupid). In the end, an embedded image markup should in its simplest form look something like this:

```
<a href="https://images.example.com/i/understanding-SEO-big.jpg">  
  
</a>
```

Note: It's okay to have images on another subdomain. It's best practice to have the smallest images (kb wise) possible, so please google and use [image minifiers].

Best Practice #12: Diversity of Content

Users don't read your content. They really don't. During my workshops, I always ask who can remember the last article they read in full, from top to bottom. Very few people can. Then I ask if they can remember the last article they googled and read in full. In most workshops, no one could, in every other workshop one person could. Even then it was always a GO-query (the user knew of the article and wanted to reach exactly this article).

Users coming from generic know- and do- queries don't read your content! They scan it. They have a "job to be done" in mind and want it to solve in the most efficient,

fastest, laziest way. So how do they consume our content? They quickly scan or skim it, looking for a fragment of information that will let them advance to whatever their goal is. Some of the users like short paragraphs of text, some like a bullet point list, some like infographics, some like a video, some like infoboxes, and some like structured instructions. Others just want to have a list of things to click. There is not one type of user: we are all users, we all like to do things our way. With SEO, you have to cater to all types of users.

Google knows that all users are different. That's why Google strives for diversity on the first SERP by showing different kinds of pages, that include different kinds of content, like:

- an article (paragraphs of text)
- addresses
- bullet points
- tables
- images
- videos
- youtube videos
- structured content like recipes and opening hours
- infoboxes, and
- maps

A simple way to achieve visibility on the first page of the Google SERPs, is to look at what isn't there. If there is no page with a video, then it makes sense to create a page with a video. If there are no results with tabular data, create a targeted page with an extensive table, and so on. You will show up (well, you will show up, if your site is fast too)! Your brand and user experience will determine if you stay there.

In general, you should prepare your content on your targeted pages in as many "sensemaking" ways as possible. This can and should be combined with the SEO process of "How do we update x% of our targeted pages?" which will be discussed later in this book. Don't try to do it all at once, but over time.

What content do I mean when we talk about "diversity of content?"

1. Paragraphs of Text

One or more short paragraphs of text are best structured in a way to make it more easily readable with tags such as `<i>`, ``, ``, `<small>`, `<sub>`, `<sup>`, ``. It's not about the value of these tags; it's all about readability. Keep in mind you must strive for content diversity, not content quantity. If you have a long article (four or five paragraphs of text), this page will not perform significantly better if you just add another paragraph. Important note: Do not push the targeted phrase into the paragraph of text if it does not fit in there. The targeted phrase is necessary for the targeting elements. If it fits naturally into the text - most of the time it does anyway - then there is no issue, but forcing it in there has no positive upside. If an SEO comes your way and tells you to write "SEO texts with relevant keywords in the first paragraph" just get rid of them - they are an idiot. The targeted phrase should just be minded in the targeting elements. All other elements must "just" be on topic and fulfill the user's intent. Stop thinking of the targeted phrase as "keywords," think user search intent.

2. Bullet Points

If you want to list something within an article or body of text, use bullet points `` ``, or an ordered list `` ``. These simple lists offer a better user experience and more value than just another paragraph of text. But there are also other kinds of lists. A list of short teasers, either in the form of headlines, (half)sentences, structured information (for example, prices), and thumbnails for example. Each list item is on its own non unique-internal or external duplicate content. We call targeted pages where most of the main content is a unique combination of content list "targeted list pages," or TLP. Whenever you are creating lots and lots of targeted list pages, you must always be sure that the lists are unique. Each list (independent of ordering) must only exist - internally and externally - once, otherwise it must be seen as internal or external duplicate content.

On a TLP, it's important not to push the list down below the fold. If the user most likely wants a list, give them a list, not an article. The unique combination of content usually links to other targeted pages. Make sure that the whole block is not linked, but certain elements separately (thumbnail, headline) so that you deliver sensemaking internal anchor text. Additionally, a list must really be a list. Each list item must be part

of an unordered list `` or ordered list ``.

A typical “rookie mistake” is to disfigure completely valid targeted list pages with nonsensemaking, general (crap) text (a.k.a SEO text). Maybe a short introduction text before a list might make sense, but most of the time it doesn’t so don’t do it.

3. Tables

If you compare something on your page, or you have any other kind of content that would benefit from a tabular display of the data, use a HTML table.

4. Embeddable Content (Videos, Maps, and More)

Most embedded content is external duplicate content. Pages with embedded content only work if they are placed within the bigger context of other content, making it part of a unique combination of content. If it makes sense for the user experience to embed some relevant content (for instance, a Youtube video, a map, a comments box-widget, or a photo gallery) on the page, then simply do it. Most of the time (and definitely for videos and maps) Google knows that this page has a video or something else embedded. Therefore, these elements add to the diversity of content.

In other cases where Google does not know, can’t crawl, or will not crawl the embedded content, it is still worth it if it has a positive impact on the user experience. But remember always to use asynchronous, non-unblocking embed codes methods. Third party content must not make the page any slower. The speed goals of a page are always absolute, and excuses like “third party crap makes our site slower” do not excuse anything and are not acceptable.

One note about galleries: only the first image that is shown when the site was loaded can be considered the main image of the page. All other images are seen as hidden after user interaction, even if it’s a carousel. (Note: Don’t use carousels. They are a horrible, horrible user experience.)

5. H2 and Other Headlines

If you have different sections within your main content, label them with different subheadings. Some of these subheadings might include a sensemaking, short variation alternative of the targeting phrase. Do not artificially lengthen all subheadings with the targeted phrase again and again.

6. Diverse Images

The big, above or near the fold main image is the most important image of the site page. However, having different kinds of images on a page, for example, a photo, an infographic, a diagram, or an illustration, contribute to the diversity of content. Similar to the main image they should have optimized URLs, alt-text, and link to a bigger version of itself. Whenever embedding images make sure the speed goals of the website are still met.

7. Structured Content

Structured content usually comes in the form of bullet points, definition lists (infoboxes), and key value pairs. Any content than can be sensemakingly structured should be structured using proper HTML (`, <dl><dt>`), i.e.:

Title: Understanding SEO

Language: English

Paperback: 291 pages

```
1 <dl>
2   <dt>Title</dt>
3   <dd>Understanding SEO</dd>
4   <dt>Language</dt>
5   <dd>English</dd>
6   <dt>Paperback</dt>
7   <dd>291 pages</dd>
8 </dl>
```

Best Practice #13: List of links to other targeted pages

Going back to my list of best practices, every targeted page should link to other targeted

pages and be linked from at least one other targeted page. Therefore, every targeted page should have a page specific list of links to other pages on the website. Sometimes this list is part of the main content, but quite often it's part of the supplemental content (for example, on the right-hand side, below the main content). It can be five links; it can be ten links, or it can be 50 links. It can be one list, or it can be different sublists.

As always, these links should make sense. That said, add some “serendipity” to each page, and the requirement that every targeted page has a link to another targeted page will be fulfilled (note: serendipity is just a fancier word for randomness). You can give these link sections the headline “Serendipity,” “From the Archive,” “Other Stuff,” or “What other people viewed”, for example. If you have a site with five million targeted pages, and you add only three serendipity links to each of these pages (and within these, there are links and the page does not link itself) you have a 95% probability that each and every page has at least one link from another targeted page.

I strongly recommend creating sensemaking and serendipity link sections on your site. However, be aware that you don't fall towards the other extreme and only link serendipitously on your site. More important targeted pages should have more links than less important targeted pages. And make sure your list of links is an actual HTML list.

Best Practice #14: Internal Anchor Text

The internal anchor text is a targeting element. Making sure your internal anchor text is sensemaking, providing an indication of where the link will go, if clicked. Make sure your internal anchor text is a short or long sensemaking variation of the targeted phrase or at least describes the topic of the page. Or to put it more simply: never use phrases such as “click here.”

DON'T DO THIS: `CLICK HERE`, but

DO THIS: `Understanding SEO`, or

DO THIS: `SEO`, or

DO THIS: `new SEO book`.

Make sure all links are immediately recognizable as links (the trick is to underline them - blue and underlined is awesome)! This is all very good for your usage values as well.

Users love to click links. They like to click on links which align with their overall intent, and they will only click on a link if they know it's a link.

Best Practice #15: Code Quality

Google is great at reading crap HTML markup, so it's not necessary that your markup strictly fulfills the W3C HTML Specification and throws no errors with the W3C Validator.

But your markup must:

- be damn fast
- be used as HTML was intended to be used
- render in a sensemaking way for common browsers

Use headlines as headlines, use lists as lists, use tables for tables, use `` for bullet point lists, normal links as links, and so on. If an experienced developer cannot make sense of your HTML markup, then there is a good chance Google can't either. Beware of `<div>` soups (markup where everything is a div), headlines used for text and all websites which were created by WYSIWYG editors from 2000's.

And did I mention it must be damn fast? Best practice is to have nothing in the HTML file that does not get initially rendered in the browser. If you have 10,000 lines of HTML boilerplate (for example, navigation with dropdowns) and five lines of actual content, do not wonder why your site does not perform in Google.

Best Practice #16: Schema.org

Schema.org is a technical specification to enrich your page with structured content so that machines (mostly Google, who are one of the sponsors) can understand it better.

All Schema.org specs can be found at <http://schema.org/docs/schemas.html>. But wait! Do not implement them right away. Google only uses a subset of these (google [Rich Snippets] and [Structured Data Google I]). As with all technical specifications, the first thing you should do is R.T.F.M.

As you will see in the specs, Google documentation support allow structured markup for products, recipes, reviews, events, software apps, videos, articles, logos, corporate contacts, and site links search box (and this list is constantly expanding). Thankfully, the people

Schema.org was gladly deprecated in 2014, so you don't have to worry about that anymore. In the cases of products, recipes, reviews, events, software apps, videos, and articles, you get better visibility in the Google Search Results with so-called "rich snippets" and "rich cards". So for example, a recipe rich snippet would look something like this:

Epic Chocolate Cake | Chocolate Recipes | Jamie Oliver



www.jamieoliver.com/recipes/...recipes/epic-vegan-chocolate-cake/ ▼

50 mins - 470 cal

Vegan, gluten-free & dairy-free, too! ... Grease two springform **cake** tins (roughly 20cm) with margarine, then line the bottom with greaseproof paper and dust the sides with gluten-free flour. ... Sieve the flour, baking powder, xanthan gum and cocoa powder into the bowl, then fold through.

And, an event rich snippet would look something like this:

Britney Spears Tickets - Britney Spears Tour Dates on StubHub!

www.stubhub.com/britney-spears-tickets/performer/5462/ ▼

Buy and sell **Britney Spears** tickets and all other concert tickets on StubHub! Get your ... 9:00 pm The Axis at **Planet Hollywood Resort** and **Casino**, Las Vegas, NV. from ... \$104. **WedAug 17** Baby One More Time, generated further hits with "(You Drive Me) Crazy", "Sometimes", and "From the Bottom of My Broken Heart".

Wed, 17 Aug The Axis at Planet Hollywood Resort and Casino, Las Vegas, NV

Fri, 19 Aug The Axis at Planet Hollywood Resort and Casino, Las Vegas, NV

As you can see, this offers a clear advantage in search results. For the other types of structured material that do not have rich snippets, Google does something with it internally, and it contributes to your diversity of content.

My recommendation is always to implement Schema.org for all schemas where Google has announced official support. Apply it as described in Google documentation and ignore all the rest. If you are in a competitive vertical where Schema.org use is common, you must use schema.org.

As always, use schema.org in a sensemaking way. Lots of websites spam Google (and themselves) with rich snippet markup that leads to rich snippets, but was actually intended for other uses. This is called "rich snippet spam." Quite common is the star rating, which was intended for review sites of any kind, but gets used by showing

up on pages without visible reviews. Google catches up on this sort of spam (most of the time, the site just loses the rich snippets, and nothing else happens). However, misleading rich snippets lead to a bad user experience and a lower click-through rate in the long run. So, yes, you can mark up your non-recipe-article to look like a recipe, and get a beautiful image in the SERPs, but then the users will think you have a recipe, and just bounce back after they have visited you and didn't see the recipe. So just don't do it. There are so many other things you can do SEO wise, and making rich snippet spam is a very bad investment of your time and resources.

Note: as Schema.org is invisible meta data with no visible feedback if it breaks, please always (after every release) test your implementation with Googles [Structured Data Testing Tool] and regularly check "Google Search Console > Search Appearance > Structured Data" and "Google Search Console > Search Appearance > Rich Cards" for all so common "everything is broken again" errors.

CHAPTER 5: DISTRIBUTION

Distribution is the process of communicating your targeted pages to Google. In this chapter, we will examine the workings of sitemap stacking, notifying and quality assurance. The first and most important step when it comes to distribution is to create a whole site inventory sitemap.

Sitemap Stacking

In its simplest form, stacking means listing all your targeted pages into a single sitemap. A sitemap is an XML file that lists the URLs of targeted pages. It allows us to add additional information about each URL, such as when the content was last changed, and if the page contains embedded images and videos. I won't go into the finer details of this format, because you can find most of the specifics in the official specification at <http://www.sitemaps.org/protocol.html> and of course, Google's spec, [google sitemap.xml].

There are a few things you should keep in mind when sitemap stacking:

- The sitemap should be made available at the root of your site. Keep it simple, so to not run into multiple reading rights issues, and place it in <https://www.example.com/sitemap.xml> format:

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
  <url>
    <loc>https://www.example.com/a/page</loc>
    <lastmod>2017-09-22T14:05:07+00:00</lastmod>
  </url>
</urlset>
```

- The `<loc>` is the URL of the targeted pages.
- The `<lastmod>` is the date when the content was last changed. Use this if your site's architecture depicts the times that the content of your pages was last changed.
- If your pages contain embedded images or self-hosted videos, you can also add

these to the sitemap. In fact, I recommend doing this for everything self-hosted. As videos are mostly Youtube embeds, there is no need to add them in the sitemap.

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
  xmlns:image="http://www.google.com/schemas/sitemap-image/1.1"
  xmlns:video="http://www.google.com/schemas/sitemap-video/1.1">
  <url>
    <loc>https://www.example.com/a/page</loc>
    <image:image>
      <image:loc>https://www.example.com/i/book-writing.jpg</image:loc>
      <image:caption>Writing a book about SEO</image:caption>
    </image:image>
    <video:video>
      <video:content_loc>
        https://www.example.com/book-writing-video.flv
      </video:content_loc>
      <video:player_loc allow_embed="yes" autoplay="ap=1">
        https://www.example.com/video/player.swf?video=123
      </video:player_loc>
      <video:thumbnail_loc>
        https://www.example.com/thumbs/book-writing.jpg
      </video:thumbnail_loc>
      <video:title>Writing a book about SEO</video:title>
      <video:description>
        Time-lapse video capturing the course of writing a book about SEO
      </video:description>
    </video:video>
  </url>
</urlset>
```

- Make sure the sitemap is UTF-8 encoded, and the URLs within the file are entity escaped. It's a typical rookie mistake to make, even though it's explicitly stated to do so in the spec at <http://www.sitemaps.org/protocol.html#escaping>.
- Every targeted page must be stacked in a sitemap. If a targeted page is not listed in a sitemap, it is not a targeted page and should not exist or not get indexed.

There are plenty of other fields that you can add to your sitemap, for example:

- priority
- mobile
- rel-alternate
- changefreq
- image
- video

My recommendation is: Don't do it! The changefreq and priority fields are simply overhead that make no discernable difference to your site. For the mobile and rel-alternate language fields, it is an easier and testable option to place them into the HTML head.

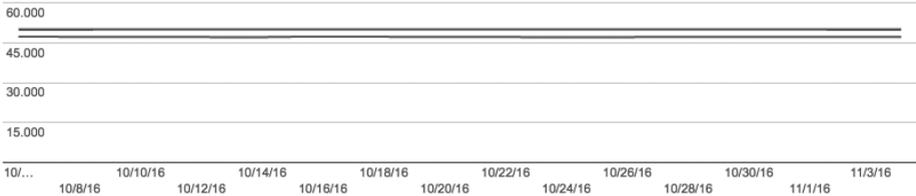
The rel-alternate mobile and rel-alternate language can be set on three different levels. They are:

- in the sitemap
- via HTTP-Headers
- within HTML Markup (the canonical)

It is important that you choose one from the list above. Do not manage this kind of information at multiple levels. I recommend using HTML markup because it's much easier to see if the correct metadata is in the markup of the HTML page. Developers naturally look at the HTML markup source of their pages, but much more rarely at the HTTP headers and, sadly, most of the time nobody looks to check if the sitemap is still correct. Therefore, it's best to keep "invisible" meta-information where people can see it in the HTML.

The simplest stacking logic is to write every targeted page in one sitemap.xml. A graph, such as the one on the next page, will show you that you have, in total, z targeted pages (as every targeted page must be listed in the sitemap) and that x% of these pages have been indexed by Google.

All content types Submitted Indexed	Web pages	Images
	50,000 Submitted 47,128 Indexed	49,871 Submitted 43,708 Indexed



If you have a more complicated site with different page types, I recommend you to stack your targeted pages into separate sitemaps. For example:

- <https://www.example.com/n-sitemap.xml>
- <https://www.example.com/l-sitemap.xml>
- <https://www.example.com/t-sitemap.xml>

The sitemap logic should mirror the namespace logic you have used for your targeted page URLs. With this setup, you will get a different indexing graph for each of your page types.

Now you can talk about communicated and indexed pages for each page type, as opposed to the first more simple setup, where you could only make whole sitewide statements. So now instead of saying, “Some of our pages are not getting indexed!” you can now say, “Pages of type ‘n’ have a lower indexing rate than pages of type ‘l’.” “What is the difference between them?”. Unfortunately, there is a limitation to sitemaps - they cannot have more than 50,000 page URLs listed. At the point when a sitemap becomes too big, you will need to split them up (with 50,000 page URLs per sitemap). It will then look something like this:

- <https://www.example.com/n-sitemap-1.xml>
- <https://www.example.com/n-sitemap-2.xml>
- <https://www.example.com/n-sitemap-3.xml>
- <https://www.example.com/n-sitemap-4.xml>

To further achieve the graph that will tell you how many page types have been submitted

to Google and how many of them have been indexed, you will need to wrap them in a sitemaps-index.xml file (a simple list that lists simple lists).

- <https://www.example.com/n-sitemap-index.xml>

```
<?xml version="1.0" encoding="UTF-8"?>
  <sitemapindex xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
    <sitemap>
      <loc>https://www.example.com/n-sitemap-1.xml</loc>
      <lastmod>2017-09-22T14:05:07+00:00</lastmod>
    </sitemap>
    <sitemap>
      <loc>http://www.example.com/n-sitemap-2.xml</loc>
      <lastmod>2017-09-22T14:05:07+00:00</lastmod>
    </sitemap>
  </sitemapindex>
```

Note: Although they look similar, the sitemap-index format is not the same as the sitemap format. Please check the official manual for the difference between these two formats - it's subtle but important.

With the sitemap-index, you will now get a graph that exceeds 50,000 submitted URLs.



You can, for experimentation reasons, stack URLs in multiple sitemaps (double submission is okay). For example:

- <https://www.example.com/all-targeted-page-with-less-than-5-items-sitemap.xml>
- <https://www.example.com/all-targeted-page-with-more-than-5-items-sitemap.xml>

Doing this will allow you to receive submitted (number of pages) and indexed values for different quality criteria of your site. If you want to narrow down why some kinds of pages may not be getting indexed, this action will be of great value. It's okay to have these pages in both the namespace-sitemap.xml files as well as in the experimental ones.

Your sitemaps must be updated automatically (if you deleted or added new pages) at least once a day. As a bare minimum, you must complete at least the simple sitemap logic (all your inventory, all your targeted pages). An actively maintained, growing site without a working distribution, is completely and utterly broken.

Sitemap Notifying & robots.txt

After your sitemap has been created, the next step is to notify Google of your sitemap.xml-files. Notifying Google is very simple. All you need to do is:

- 8) List all your sitemap-index.xml files in the robots.txt.
- 9) List all sitemap.xml files that are not part of a sitemap-index.xml, directly in the robots.txt.

robot.txt is a text file that will let Google know how to crawl your content. It is simply a file that must be placed at the root of your domain, so for example <https://www.example.com/robots.txt>.

We are going to discuss this all in more detail over the next few pages, but you should know that with this step, and after Google has crawled the robots.txt (about every 24 hours), you will have communicated your complete inventory to Google. It's not yet crawled, and, not yet indexed, but at least Google knows your stuff exists.

It's important to make sure that listing your sitemap.xml in the robots.txt is an automatic process. As soon as a new sitemap.xml exists, it must be listed in the robots.txt. When

you do it, it will probably look something like this:

```
User-Agent: *  
Disallow:  
Sitemap: https://www.example.com/n-sitemap-index.xml  
Sitemap: https://www.example.com/z-sitemap-index.xml  
Sitemap: https://www.example.com/a-sitemap.xml  
Sitemap: https://www.example.com/d-sitemap.xml
```

So what's the meaning of all of this?

“User-Agent: *

Disallow:”

This means nothing is to be dismissed on your page, and the result is that every bot is permitted to crawl everything and anything. The robots.txt is a crawling directive, not an indexing directive. So, whenever you block something, Google might index it anyway. It will not know much about the URL, but it might index it anyway.

In 99% of all cases, you want to manage indexing (which determines if it shows up in Google or not), not crawling. To manage indexing, always use the <meta name=“robots” content=“noindex”> tag, and never use the robots.txt. If you block something with the robots.txt, Google cannot see the content anymore, so even onpage <meta name=“robots” content=“noindex”> would be ignored on that page. Only in 1% of cases you want to control crawling of a certain resource. And even then, I recommend you rethink your internal application logic instead of blocking it with the robots.txt, as you might trade in a crawling issue with an indexing issue.

If you want to block other crawlers because of “reasons” (although I do not recommend it), you can do that, but always make sure that you do not block Googlebot from anything. Malicious crawlers (for example, ones looking for email addresses to spam) will crawl you anyway, regardless of what you write into the robots.txt. If it is important that content is not crawled, I recommend some other method (i.e., HTTP-Auth IP-Blocking) on the server side to prevent this.

Therefore, just do:

```
User-Agent: *  
Disallow:
```

Every other disallow statement must be seen as a bug and should be solved in another way.

Important note: Never, ever block images, CSS or JS. Google must be able to render the site the same way your users do. If you block JS, CSS or images, your site will not perform. This, therefore, is a direct violation of the first page of the Google Webmaster Guidelines:

Never!!:

```
User-Agent: *  
Disallow: /js/  
Disallow: /css/  
Disallow: /images/
```

Private Sitemaps

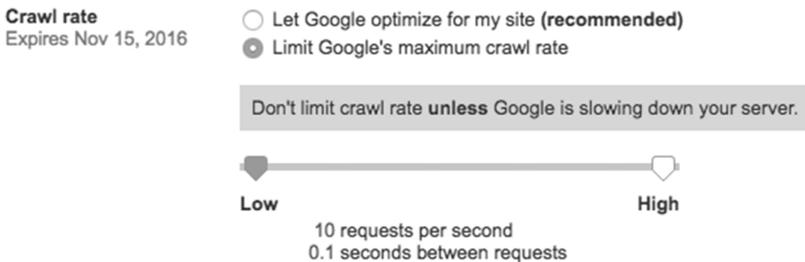
In certain cases, you may not want to have your sitemap.xml listed in the robots.txt. One of these instances could be that you want Google to crawl you, but you do not want your competition to easily infer how big your page inventory is. My opinion is this rarely makes sense, because it can be circumvented with other techniques. However, there probably is 1% of cases where this may make sense. In the 1% of cases, give your sitemap.xml difficult to guess filenames such as *https://www.example.com/a-sitemap-franz-was-here.xml*

And don't notify Google via robots.txt but via Google Search Console (where you can manually submit a sitemap.xml). This way, Google will know of them, too. The downside of this approach is that other search engines don't know about you (which doesn't really matter), but also that people tend to forget to submit a completely new sitemap.xml (if not in the next six months, then in five years) and then wonder why Google doesn't index them properly.

Getting Crawled And Indexed

Ok, so you've done everything right. You listed your complete inventory of targeted pages in the sitemap.xml and you notified Google of your sitemap(s).xmls. Google will now crawl you and hopefully, index you.

Crawling is the process of Google accessing your page and downloading its content. Your biggest enemy for this (after you have done the stacking and notifying correctly) is your server performance and arithmetic. Google assigns a certain rate to crawl your site, a so-called "crawl budget." It's based on overall content quality, estimated importance and a lot of other stuff we don't know. You can see and influence your crawl rate at Google Search Console > Site Settings.

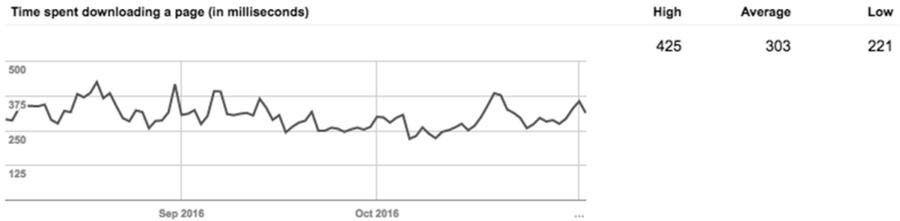


As you want to be crawled as thoroughly and as fast as possible, it's a good idea to set this slider to 10/1 (10 requests per second). If this is not possible in your view of Google Search Console, then don't worry, simply set it to the maximum for now. If you wait a few days, you'll be able to raise the slider higher. Keep doing this until you have reached 10/1.

At the same time, however, you must make sure that your servers remain fast. You must still be able to reach the golden goal speed values we discussed previously. If your site becomes too slow because of the massive crawling by Google, buy new servers (or get a better connection between your servers), or solve any other technical issues that hurt your servers' performance.

Additionally, you also need to care about the "time spent downloading a page," which

can be found in Google Search Console.



This graph must show a stable (not spiky) line, and you must deliver an average value of about < 300ms. The “time spent downloading a page” graph is not about the first byte of the page, but more like the last byte. And, if your site is slow, even if it is only temporary, then Google will reduce crawling your site.

If you do these things correctly, you will at some point get an average “pages crawled per day” of between 400,000 and 860,000.

Also, you should know that Google has something I call “crawl efficiency”. I put this value at 0.7 on a well optimized site. This means that only 70% of the pages crawled within a given timeframe are unique pages, and the rest are duplicate crawls of the same page. And 0.7 is only the value if you do everything right: if you have lots of parameters and deadweight URLs or pages, the efficiency rate can be considerably lower.

Let’s consider this further with an example. If you have a site with ten million pages and your average crawl rate is 400,000, with the average crawl efficiency being 0.7, the time taken to crawl your site is:

$$10,000,000 / (400,000 * 0.7) = 35.71 \text{ days}$$

Therefore, it would take about 36 days until you can be sure that a new or updated page has been crawled. However, if you have managed to keep the crawl rate high (~800,000), you will have halved that time. But, if you have 20 million pages, or lots of different URL variations, redirects, and parameters, it can be even longer.

So, this means, it takes time. Whenever you update a big portion of your big site it takes

Google quite some time to recognize the updates on all of your pages. In general, I recommend that it must not take more than one month with a presumed crawl efficiency of 0.7 to crawl all your pages. That's because, if it takes longer than one month to connect the impact between cause (i.e. onpage optimization) and effect (better visibility and traffic) then we (as busy human beings) don't see any cause and effect at all. This puts a theoretical limit of the maximum pages on your site to 25 million with best case values. If you have more pages than that, I recommend you split your pages into different websites with different domains.

You can also choose to not set a manual crawl rate in the Google Search Console settings. In this case, Google will assign a crawl rate to you, and this crawl rate might be higher than ten requests per second.

That said, the only cases where this happens are when you:

- a) running an industry leading site with an unbelievably fast server
- b) are hosted at a data center that has a direct glass fiber connection to Google and you have a fast server
- c) have such an unbelievably blazing fast server and a lot of high quality content.

For small sites, mulling over crawl performance and the settings recommended here hardly matter. A higher priority is making sure your servers are stable and the average time spent downloading a page is about 300 ms.

Indexing is Google's process to store your content in a way that it can be matched with the user's query. What share of our targeted pages is being indexed depends on mostly unknown factors. What we do know is that the killers for indexing are slow sites, internal duplicate content, external duplicate content, a poor QC, as well as targeted pages that are not interlinked. If you have done everything as outlined in my targeted page best practice section, you should expect:

- close to 100% indexing for all targeted pages with unique content
- close to 90% indexing for all targeted pages with a unique combination of content as the main content.

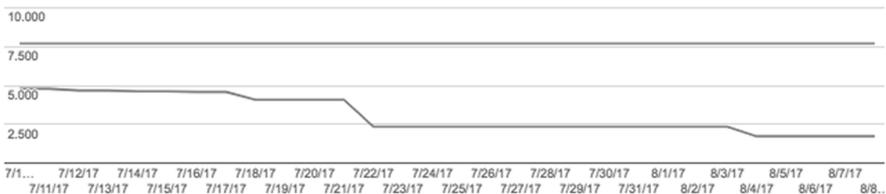
If you are below these benchmark numbers, then you are underperforming. If less than 70% of your targeted pages are indexed, your site must be considered completely

and utterly broken. If you have such a case, the reason usually falls into one of these categories:

- speed: the pages do not fulfill the web performance goals
- content: internal or external duplicate content, or no content
- interlinking: pages do not fulfill the interlinking requirements
- other: pages do not systematically fulfill one or more aspects of the targeted page requirements

Now another important note: Beware of the opening scissors!

Let me explain. If you are constantly creating and submitting more and more targeted pages to Google, you must make sure that your overall indexing rate stays at least the same, or improves. If you are creating and submitting more pages but the indexing



ratio is going down, or you submit a constant number of pages, but the number of indexed pages goes down, the “scissors are widening” as seen in the illustration above (Google Search Console > Crawl > Sitemaps > Web Pages). This means something is wrong. When you notice this happening, I recommend you stop submitting new pages and wait one week. If the scissors begin to close again (the submitted and indexed lines converge again), then you simply have a delay (which can be fixed by increasing your crawl rate). If the gap doesn’t close, you must immediately check your speed, server, distribution and targeted pages, because one or more of these factors is hurting your webproperty. Remember, if your hard work does not get indexed, then your hard work will not contribute to your traffic and is a wasted opportunity.

Updating Sitemaps

The entire site inventory (your sitemap.xml) should be updated at least once a day. There

are many cases, though, when this is not fast enough. One simple method you can use to speed up the indexing of new pages is to create a feed of new and updated material, published as an `update.rss`.

The `update.rss` is nothing other than a simple sitemap in the RSS format where you stack list all of your targeted page URLs that were created or updated in the last few hours. Then, if the `update.rss` has changed (so, if there are pages created or updated in the last few hours) you ping this `sitemap.rss` to Google. Unlike your main sitemap, this feed shows only new and changed content. To do this, you need to create an automated process that:

- (re)creates `https://www.example.com/update.rss` (must be in the root of your domain) with new and updated pages every hour
- makes a server side GET-Request to `https://www.google.com/webmasters/tools/ping?sitemap=%your url encoded url of the update.rss%` if pages were created or updated in the last few hours.

There's more about `sitemap.xml` pings at https://en.wikipedia.org/wiki/Sitemaps#Search_engine_submission. If you are in the news vertical, please google [Google News xml].

Distribution Quality Assurance

A website with a broken distribution is a broken website. And rather than having direct feedback like you get when your pages are broken, you do not have this kind of feedback when your distribution breaks. Instead, your traffic will not plunge immediately, but your growth will slowly decline over time.

To prevent this from happening, quality assurance in the form of regular manual checks of the sitemap, and examinations of the key metrics (such as the average time spent downloading a page and pages crawled per day, both available via Google Search Console), is essential.

Automated tests of the sitemap do not hurt either, so make a habit of checking if the `update.rss` really is being updated in real time, and if the `update.rss` is still being pinged, and so on.

The Startpage, Targeting, And Linking

It is important to realize that SEO does not care too much about your startpage. From an SEO perspective, the startpage is:

- a targeted page for the brand of your company
- a targeted page for one competitive targeted phrase
- a value hub.

However, it's important for your brand and the competitive targeted phrase to make sure that your title has a format like this:

```
<title>%your brand% %targeted phrase%/</title>
```

Start with your brand, and then target something you want to be found for. I recommend choosing one targeted phrase that stands for what your business is. Here are some made up ones so you can get an idea:

```
<title>CarIns Car insurances online</title>
```

```
<title>Crocodile RockBungalows on Koh Chang</title>
```

```
<title>Pills & Fun Online Pharmacy</title>
```

It's essential that your brand name is visible in the text on your startpage as well, and not only as an image. Make sure your competitive targeted phrase is also available in another visible way on the startpage, but keep in mind there is no need to repeat the targeted phrase in all targeting elements on the startpage.

When it comes to linking, you will find that almost 90% of the inlinks from other websites to your webproperty will link to the root of the domain (a link direct to *https://www.example.com* for instance). This means you must do two things:

- 1) Your startpage must be in the root of your domain. Google will always see the root of your domain as the location of your startpage, and you should show real content immediately. Do not redirect away from your startpage:**Good:** *https://www.example.com/* -> HTTP 200 must show the content of your startpage**Bad:** *https://www.example.com/* -> redirect (whetherHTTP 301 permanent (or HTTP 302 temporary) redirect to *https://www.example.com/home.html* (or something else).
- 2) You must link like crazy from your startpage to your targeted pages. To achieve this, I recommend that you have a sensmaking link section on your startpage where you statically link to your most important targeted pages. Additionally,

you should have a section that links to your newest and/or latest updates targeted pages. And, of course, a section that serendipitously links to some targeted pages.

In addition to these two things, you should focus on speed and the user experience, but generally, you can do whatever you like on the startpage. There is no need for diversity of content or anything like that here.

CHAPTER 6: DEADWEIGHT

I've touched on the topic of deadweight a few times in this book, and that's for a good reason. Deadweight pages, that is, any page that is not a targeted page or the start page, should either

- not exist
- not be indexed
- not be systematically linked
- not listed in the Sitemap.xml

However, there are some special cases where deadweight pages have a reason to exist, for instance search pages. User input triggered search pages have no QC, as the user can search for anything, meaning potentially any page might get created and might be indexed. Although search pages are deadweight, they have a user case, and it makes sense that search pages have a shareable URL (so that users can bookmark and share their searches). To deal with these pages, simply set meta robots to noindex `<meta name="robots" content="noindex">` in the HTML head. You also shouldn't systematically link to deadweight pages, as you do not ever want them to be crawled, especially if there are many or a growing number of them.

Another good example of deadweight pages that may exist is "pagination pages". Pagination pages do not have a unique dominant targeted phrase because the targeted phrase of the paginated page is always the same as the first page. As there is a reason that paginated pages exist, set them to `<meta name="robots" content="noindex">`. Additionally, on big sites with hundred of thousands to millions of pages I recommend to not even link to paginated pages systematically. Use some JS magic without direct links to lead the user through the pagination.

You should also know it is possible to set the canonical of a paginated page to the first (unpaginated) targeted list page, so for example, `https://www.example.com/c/seo?page=7 -> <link rel="canonical" href="https://www.example.com/c/seo">`, as a paginated page is just variation of the unpaginated page. For small site settings, the paginated pages to noindex is fine, setting the canonical to the unpaginated URL is just necessary on really big sites. (Huge sites should best not link to paginated pages at all,

use JS pagination instead).

It's also imperative that Google can discern between pages that exist and pages that don't. Whenever Google encounters a HTTP 200 page that has no content or has an error message, this is called a "soft 404 error". Errors have a serious impact on crawling and indexing, and if widespread, can greatly effect your whole site's traffic performance. You are also seriously hurting your crawling when you always redirect away from a page that does not exist, for example, `https://www.example.com/hudriwudri` -> HTTP 301 -> `https://www.example.com/file-not-found` -> HTTP 404. If a site does not exist, you must return a HTTP 404 directly, without dragging Google (and your user) over a chain of redirects.

However, if you move a page from one URL to another URL, then HTTP 301 Permanent Redirect is correct (for example, `https://www.example.com/hudriwudri` -> HTTP 301 `https://www.example.com/alhudriwudri-new`).

If you have deleted a page that previously existed, you should return a *HTTP 410 Gone*. This has two implications:

- Google will get rid of the URL in its index faster than with a HTTP 404
- you can make a distinction between real errors (404) and wanted deletions (HTTP 410) in GSC

Top 1,000 pages with errors

Download	MARK AS FIXED (0)	Filter	Show	25 rows	276-300 of 1,000	<	>
<input type="checkbox"/>	Priority	URL	Response Code	Detected			
<input type="checkbox"/>	980	i/4640159	410	10/24/16			
<input type="checkbox"/>	998	i/4595971	410	10/18/16			
<input type="checkbox"/>	999	i/4627128	410	10/16/16			
<input type="checkbox"/>	1000	i/4621475	410	10/11/16			
<input type="checkbox"/>	1	f/traisenpark	404	10/31/16			
<input type="checkbox"/>	2	f/tr%C3%A4umeland	404	11/2/16			
<input type="checkbox"/>	3	f/fischapark-shopping-center	404	10/25/16			
<input type="checkbox"/>	4	f/sch%C3%BCller-bau	404	10/29/16			

Having HTTP 404 occasionally on your pages is normal, and will not hurt your overall website's performance. However, if they are growing in numbers, you should "fight" them and fix the source of these errors. To identify real errors, you must be able to make a distinction between the HTTP 404 and the wanted errors HTTP 410. On any halfway big, or halfway complex website, this distinction is a must.

If you delete parts of your website (HTTP 410), then the remaining parts of your website will not suddenly perform worse. You will lose the value of the inlinks to the deleted pages and, of course, the traffic that reached those pages will be gone, but the rest of your site will keep performing. Don't be afraid to delete pages. Do not keep old pages, or pages that you do not need or want to maintain any longer, just because of SEO. If your site becomes unmanageable and unmaintainable because of the old stuff, then no search engine optimization will save you.

Edge Cases

There are some edge cases in regards to deadweight, namely pages for which you do not particularly care if they are found, but nonetheless are necessary for branded go-queries of your users, such as: [%your brand% login], [%your brand% contact]. Therefore, they must exist and be indexed. The pages I'm talking about are pages like:

- mandatory legal information, such as imprint
- contact details
- terms of service
- login
- other common service pages for your vertical.

These pages might not be optimized in the common sense, but nevertheless, should always:

- fulfill the URL rules
- be indexable (are not set to noindex)
- have an optimized title and headline
- be linked from the startpage and/or sitewide (normally via the footer)

Although these are not targeted pages, the page structure and content should still indicate what it contains, contributing to a good user experience. If you have less than

a handful of these pages, do not worry if they are in a sitemap or not. If you have more than a handful, I recommend submitting them via the sitemap.

CHAPTER 7: ONGOING SEO PROCESSES

After you have optimized your platform to consist only of

- targeted pages
- distribution
- a startpage
- correct deadweight handling

you must focus on the ongoing SEO processes.

The ongoing SEO processes are the things you will need to do continuously to see SEO success, such as creating new content, updating old content and marketing your webproperty in new and different ways. Maintaining your ongoing SEO processes is just as important as everything else we have discussed in this book because without SEO processes all your platform work will lead to nothing.

All SEO processes, as you will see below, are formulated as questions. Why is this? Because the answer to the processes is different for every company. Companies have radically different activities, both online and offline, and, even in the case that the technical setup is the same, the people, the vertical, the market and the overall context are always different. The answers to the processes must come from within the company, and cannot be blindly copied straight out of a book, even one as awesome as this. An SEO process that works for one company can not be 1:1 copied to another company. That's why all I have are questions and experience. The answers must come from you!

Once your platform is optimized for SEO, the following SEO business cases will be possible:

- if we add x% more targeted pages to our site, we gain y% more traffic
- if we update x% of our targeted pages, we gain y% more traffic
- if other sites point x% more sensemaking links to our targeted pages, we gain y% more traffic
- if we make the user experience of our site better, our traffic gets more sustainable
- if we gain more brand recognition, our traffic gets more sustainable

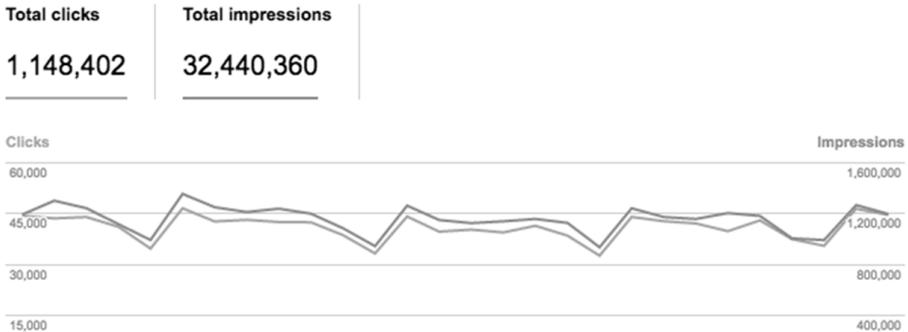
Therefore, these business cases lead to the following questions:

- How do we create x% more targeted pages?
- How do we update x% of our targeted pages?
- How do we get other sites to point x% more links to our targeted pages?
- How do we ensure that the user experience of our site gets better?
- How do we ensure that our site gets more brand recognition?

Let's go through them in detail.

Business Case #1: How do we create x% more targeted pages?

Once we have a set of targeted pages that performs this questions arises: "How do we ongoingly create more targeted pages?" The logic behind this is quite simple: If the CTR between impressions and SERP clicks stays more or less the same, and then we create more targeted pages, we get more traffic.



There are a few things to consider. An optimal case might be 100% more pages equals 100% more traffic. But that said, different targeted phrases have different search demand and varying competition in the SERPs. So, 100% more pages can lead to 150% more traffic, but can also lead to only 20% more traffic. It's not possible to calculate how much traffic x% new pages will bring you, and all the paid SEO tools which promise you this are just guessing. The one thing you can rely on though if you optimize your SEO platform correctly is: more targeted pages equals more traffic.

You must be very mindful of internal duplicate targeting: It's not about more pages! It's about more *targeted pages*. If you create more and more pages that want to get found for the same targeted phrase, you will actually lower your overall search traffic performance due to internal competition.

You must also be aware of internal duplicate content: If you are automatically creating a unique combination of content (such as lists), you must make sure that every combination is unique. It's not enough that the sorting is different. As a minimum, the utility, targeting phrase and at least one list item (but really the more, the better) of the visible list must be different.

For better visualization: A, B, C, D, ... are list items.

A page with A, B, C and another page with A, B, C are internal duplicate lists.

A, B, C and A, C, B (different sorting) are also internal duplicate lists.

A, B, C and A, B, D (one item difference) are unique lists.

A, B, C and A, B (one missing item difference) are unique lists.

Also, with more targeted pages and therefore more traffic business cases you must be careful of internal interlinking. Internal interlinking is easy with a small, more or less static number of pages. It gets hard with a big and growing number of pages. Make sure that all pages are sensemakingly interlinked, with a touch of serendipity.

SEO does not care about the (main-)navigation. The argument that you cannot create more and more targeted pages because "where will we put it in the main navigation?" is not valid. Nobody cares about your ultra-complex, hierarchical main navigation - not the user, not Google.

One easy way to create more pages is to lower the quality criteria (QC) of the targeted pages. You could adopt shorter lists (for example, job pages with just two vacancies instead of three), add less content on the pages or look for "cheaper" created content. This is a valid strategy, but be sure to realize: the lower the QC, the higher the volatility of search traffic. The pages may perform for a while, and then not. You may decide to update them, and they perform again, but when the competition moves in with higher

QC pages, you will lose traffic. It becomes a difficult cycle to keep up with to maintain steady traffic. In general, you should start with a very high QC, and only lower it slowly. As soon as you see any traffic volatility, work hard to improve the quality of your targeted pages once again.

Important: if you create more targeted pages and you don't get more traffic then something is wrong. In this situation, stop creating more pages and look for the reason (most likely a QC, internal duplicate content or internal duplicate targeting, internal interlinking or distribution issue, web performance issues, or a combination of them).

Sliding Content

One common issue with the case of creating more pages is that there is quite often content that only has limited validity. In most cases, that content is only valid over a period of time, for example, job offers, used car sales, special offers, and events.

For targeted detail pages (pages that list only one item) this leads to a loss of pages. An item existed, and then it doesn't. For targeted list pages this might also lead to a loss of pages, as the QC might not be fulfilled any longer (so for instance, a list of four listings suddenly has three entries, and the QC is > 3). But if we want sustainable, growing traffic we cannot delete and recreate pages all the time. So what to do? The answer differs for targeted list pages and targeted detail pages.

For targeted detail pages (TDP), I recommend reusing the pages. For example, perhaps an event targeted at [happy puppy pet show vienna] happened sometime in the past. You can keep the page around, provided there is a big visible disclaimer that this is an old entry (and include some further links to new events that might also be interesting). If the event happens again (maybe next year for example), you simply update the page. If you want to reuse the same URL, you must make sure there are no dates in the URL (dates in URLs are a dependency and violate URL rule number two: permanent).

If reusing pages does not make sense or is not possible in your vertical, I recommend to just delete the old entries (ideally with a HTTP 410 status code). Do not keep old pages around "just because" - they will lead to internal duplicate targeting and an

unmanageable site quite soon. Also if you can't offer a good user experience on the old targeted detail page, just delete it.

For targeted list pages, I recommend using old listings to fill up the pages. So for example, in a case where you have a QC of at least three entries on a page that shows up to 15 list items, but you only have one active listing and 10,000 old ones, you would just list the one active and 14 old entries (no pagination). You used the old ones to fill up the page with up to 15 entries. This keeps your list page in existence, with the logical consequence being that once a valid targeted list page existed, it keeps existing, and attracting traffic.

This can be seen as quite an aggressive SEO strategy. Showing old content can be valuable to the user (at least they know that there was an offer, event, job, ... in the past) but it's not easy. You must make sure that the usage metrics of these pages and the user experience is still up to your standards, and the utility of these pages to the users is still given. In any case, do not mislead the user! Make it clear via a strong visible indicator that the old entries are outdated. You might not even link the old entries to the targeted detail pages (in the instance that they no longer exist). You can offer the user other options, link to other useful pages or even link to your competition if it makes sense. But the user must not bounce back (much) more to Google than on your other pages. Overall, the "keep old stuff around" strategy is easy to do in a bad way, and hard to do well.

Business Case #2: How do we update x% of our targeted pages?

Another underlying business case of SEO is that if "we update x% of our targeted pages we will gain y% more traffic." Well maintained web pages deliver a much better user experience, with the bonus that Google is looking to deliver diversity in its SERPs by showing old established web pages, as well as pages with "fresh" up-to-date content.

Updated (up-to-date) pages sometimes perform astonishingly better. We did an experiment with one client and updated 50 targeted pages (added information, structured the page content better, deleted some outdated content) and the pages attracted 650% more traffic.

So what does “update” mean? It means adding content, but not “let’s add another paragraph of text.” The goal should be to add to the diversity of the content. If you have paragraphs of text, add a bullet point list, later add a table, then a video or infographics and so on.

Updating your content also means to delete outdated content. It means restructuring the content to be more effective (for example, if information is woven into a paragraph of text that nobody reads anyway, restructure it as a bullet point list or table). In the case of list pages, updating most of the time means simply bringing the list up-to-date. Of course, adding diverse supplemental content to these pages helps, too.

Even though it sounds simple, updating content is often a major challenge for organizations. This is especially true for companies where everything needs to be checked with legal, or where the CEO wants to have the last word about everything that appears on the website. These are operational challenges which need to be solved, as there is no way around updating targeted pages.

So, how often must you update your content? All the time - it’s an ongoing process. It obviously depends on your vertical: if you offer health information, your content must be up-to-date as anything else is a horrible and potentially dangerous user experience. Whereas, if you offer information that isn’t as critical, it can go longer without an update. Targeted list pages should also be updated in or near “real time,” so if you have a new list item that belongs on the list, update the list.

As a general rule of thumb, I recommend updating 2% of your editorial targeted pages every week, which means roughly 100% of your editorial targeted pages in one year. But please, just use this as a rule of thumb as it may be different for your vertical or business. However, there is one thing that is true for all verticals: if you want your targeted pages to perform better over time, you need to update them. After you created a page, the work is only just beginning. You then must establish a content update process within your company. Updating pages is as important as creating pages.

Business Case #3: How do we get other sites to point x% more links to our targeted pages?

With more inlinks from other valuable webproperties, we can expect our existing targeted pages to perform better overall for their targeted phrases, therefore getting more clicks from the existing impressions.

And although this is not a book about link building strategies (and I will never write one), there is one thing I have learned over time, and that is that every company must find their own strategy to attract links.

Google has set some quality standards when it comes to the topic of links. You can read these in the [Google Webmaster Guidelines]. If these guidelines are violated, Google will automatically or manually penalize the site that set the link, the site that was linked or both. Whenever you do something majorly wrong with your links, Google will send you a message via Google Search Console. This message includes samples of what you did wrong from their point of view and what you should do now to fix it. Simply do what they tell you in the message. If you do not, the penalty might result in a lack of growth, diminished search traffic performance or even complete removal from Google.

Remember: whatever you do for inlinks, you must make sure to get additional value out of these activities, as 75% of all resources in plain link building are just wasted.

However, the web is not a one way street, and good sites link to other good sites. It's good for user experience, it's good for your usage values, and it's good for Google. Often people are afraid that when they link to other websites, that they will lose the user. This idea is somewhat foolish - if the user can't go further, you have lost them already - they will bounce back to Google and remember that your website isn't a very good resource.

Therefore, link as much as you want to other web pages that you wish to recommend to your users. Do not think about SEO while doing so. That said: if you have a direct monetary or material benefit (see Google's spec [rel nofollow google]) you must set these links to rel="nofollow": `Somewhere Else`

It's important you don't obsess over nofollow, as in 99.9% of all cases your external links will benefit you and your users. And in the cases where Google thinks that you did something shady (either receiving or giving "unnatural" links), it will tell you so in the Google Search Console. Any SEO which tells you to "nofollow" all your links or "nofollow" links to "low Whatever-Rank™ websites" is an idiot and should be treated as such. "Nofollow" is just a peripheric topic for SEO. Just follow the Google spec.

Now for some advantages and disadvantages of common link building strategies:

STRATEGY

Natural Links

“Creating awesome content and people will link to you.”

Paid Links

“Paying sites (directly or via a link network) and get more links over time.”

	ADVANTAGE	DISADVANTAGE
	All resources end up where they belong on your webproperty.	People are lazy and hardly ever link to anything if not actively encouraged.
	It's easy.	It's a clear violation of the Google Webmaster Guidelines. May result in an underperforming site or complete removal from Google Search. Links are mostly poor quality and clearly identifiable as spammy, resulting in a bad brand experience. I, and Google recommend not to do it.*

STRATEGY

Link Bait

“Creating an awesome targeted page that is so good that it goes “viral” and gets you lots and lots of links.”

Outreach

“Pitching your website or targeted pages to other websites, mostly blogs.”

Guest Posts

“You posting a guest article on another website, and including a link from the article to a targeted page on your site.”

ADVANTAGE

If it works, it gets you lots of links and referral traffic (social media), as well as being a great marketing opportunity.

This is permitted within Google Webmaster Guidelines as long as no exchange of money or goods is involved. Integration in the micro-communities of the blog.

Easy, fast, very good manageability of what gets linked and how (since you write the anchor text yourself).

DISADVANTAGE

Very hard to do. 99% of all approaches you try will not work and only have a minor response. If something worked for another company does not mean it will work for you.

Time intensive. And annoying.

If done badly, it can be a violation of Google’s Webmaster Guidelines. If the anchor text is overdone or there is a clear link building purpose, it may have major negative impact on both parties involved.**

STRATEGY

PR

“Classically, telling lots of other sites how awesome you are or how awesome what you are currently doing is, and they might link to you in their resulting coverage. This is a specialist task but can be incredibly successful.”

Widgets & Badges“

Creating some manner of embeddable content, such as image-badges, iframes or javascript-embeddables that link back to you.”

ADVANTAGE

Within the Google Webmaster Guidelines. Lots of additional value from positive PR.

Good manageability of what gets linked and how (anchor text).

DISADVANTAGE

Time intensive. Most PR is boring and sucks.

If it's spammy, it's spammy. If the links are not sensemaking or not on topic with the widget, then it's a violation of Google's Webmaster Guidelines. It's best just to link your brand to the startpage. Most sites will not integrate your widgets/badges.

STRATEGY

Deep Integrations

“Part of your content or service gets deeply integrated into another webproperty, whereby the other webproperty links to you. For example, a travel website fetches (via scraping or an API) weather data from a weather website, which is displayed alongside the main content. The weather website is cited as the source, with relevant links.

ADVANTAGE

Both websites win, with the one gaining useful structured content that can be used for the diversity of content, and the other website gaining links to their targeted pages, increasing their visibility and having a return link.

DISADVANTAGE

Complicated, and lots of effort for both parties involved (for example, one site must offer an API, the other must integrate the data). You also need links from many different web properties, not lots of links from a single one. Might be considered spammy or even spam by Google if the integrations or links are nonsensemaking.

STRATEGY

Your Strategy

There are other strategies and combinations of different strategies, depending on your business. My recommendation is to go with the strengths of your business - if you have a strong sales team then don't be shy to ask them to ask (ask, not force) clients for links ("as we are partners anyway"), and even make it part of the sales team's KPIs.

ADVANTAGE

If you have creative developers, you can use them for awesome link bait single page apps, embeddable widgets or more advanced integrations. And if you can do great marketing, use this, as your marketing team will have millions of ideas. If you are changing the world for the better, or have an interesting angle for press attention, go with PR and don't forget links.

DISADVANTAGE

The wrong link strategy can break a company because the resources are not focused on your product or website, but on an external factor. In many companies, there is quite often a misallocation of resources towards "link building" using the worst aspects of the techniques mentioned in this table. In the long term, this will kill a company.

*Also on this topic, do not participate in the link provider side of this deal. It is common for sites to be approached and asked to accept links in exchange for payment (or another reciprocal arrangement). What happens is if you sell links, Google will remove your links from the equation, and they will no longer count as a vote. In the long term, your site will perform more and more poorly, and you will deliver a bad brand experience for your users.

**Only do this if it really makes sense to get covered on that particular website (one that's within your vertical and has the relevant opportunity for articles). The value of this method must be in the publication of the article and reaching the audience of the publishing site; any link back should only be an afterthought. Don't overdo the links - it's best just to link your brand to your startpage and maybe one specialized on topic targeted page.

§ Note: For iframes and javascript-embeddables the link must be outside the actual widget. For badges, the alt-text of the image substitutes the anchor text of the link. Basically, whenever you do widget or badge-bait, you need to treat it as a product with a product manager and a "sales" team that asks, begs, encourages, and helps sites to integrate the widget/badge. There is no such thing as a "viral" widget or badge - you must plan your distribution strategy together with the widget. Most websites will not

integrate your widget no matter what you do, so you can use the widget as part of your outreach strategy. Ask them if they want that awesome widget, and if not, ask them if they will at least link to your awesome website.

To end this “link growth over time” topic: Getting links is hard and mostly wasted work. It’s easy to do link building badly and spammy and therefore with a negative value. None the less, you need valuable and good references (links) from other websites. Find a way that works for you (together with your other outreach and brand building strategies) but don’t over prioritize the whole “link” part of “link building.” You need good references, not “links.”

Business Case #4: How do we ensure that the user experience of our site gets better?

It is true that “if we make the user experience of your site better, your traffic will become more sustainable.” If you have bad UX, some pages might sometimes perform for a short period of time, most of them won’t. With awesome UX most of your pages will perform most of the time for a long time. Every company must have at least one person responsible for UX. This person should not be the person responsible for SEO. UX is a big part of SEO, but UX is not SEO. UX is bigger!

The most crucial factor when it comes to user experience is speed. A slow site greatly affects the user experience, and most of the time it’s a gradual decline from a fast site, to okay, to “little bit slow”, to “completely and utterly broken slow.” As I have repeatedly said, slow sites do not perform because they deliver a terrible user experience and a bad brand experience.

Of course, user experience is not only about speed. Things such as feature bloat and ads can also have a major negative impact. If you continuously add more features to your site, then your user experience will decline. If you earn money from advertising with ads, then you can always earn more money with more ads. You might only make gradual changes at first (one more ad placement here, let’s put the ads a little further up ...), and then one month later, when you see the additional money these small changes have earned you, you will do more ... and more ... and more of those. Over time you will

effectively destroy the UX of your site, and you can't go back because any change (fewer ads or, less aggressive positioning) will have a negative impact on your bank account.

Therefore, if you want to maintain the existing traffic you have and continue to grow it, then user experience is a must. Set some usage metrics as UX KPIs, with the most simplest one being:

- bounce rate (down is good)
- time on site (up is good)
- average pageviews (up is good)

Measured via Google Analytics, make sure that they only get better, and never worse. If they do get worse (and occasionally they will), then actively counter it. Don't be afraid to drag the ads into the discussion. Ads are not out of the scope of user experience only because they earn money.

Business Case #5: How do we ensure that our site gets more brand recognition?

In the past, I have always summarized brand recognition and experience within the user experience business case. Yes, the most important branding tool you have is your website once the user has reached it. So user experience is a big part of brand experience, but it's also more.

So how can you make sure that your site continuously gains more brand recognition? You can simply go into Google Search Console > Search Traffic > Search Analytics and investigate both:

- branded searches (queries which only contain your brand or an alternate spelling of your brand (so for example, if you are AirBnB you would look for [airbnb], [airbnb], [airbb], and so on)
- generic + branded search combinations (queries which contain your brand and some organic other part for example: [airbnb berlin])

And make sure that this traffic segment is growing over time. All sites that have issues growing their traffic, even though "they have tried everything," have a diminishing

amount of branded traffic. Even “simple” things like a logo change might have a major positive impact on such underperforming sites.

In addition, think about how to reach users before they reach your site. Any kind of “nonannoying” advertisement is ok, for example:

- Offline: TV, print, billboards, radio, word of mouth, and flyers
- Online: PPC, AdWords, AdSense, retargeting, social media ads

I said “not annoying” because you must be careful to not “negatively brand” your users. Remember if they don’t actively click on you when they see you in the SERPs anymore, you have lost and failed in the business of SEO.

Ongoing Feedback

Okay, so let’s say you’ve done everything written in this book: you’ve created the perfect optimized web platform, you’ve set the right KPIs, you’ve introduced SEO processes such as ways to create new content and update old content throughout the entire company, plus everything else. *And then everything turns to shit.*

Mistakes happen, wrong decisions happen, shit happens. Always, all the time. And there is no way - no test, no setup, no experienced manager, no megastar 10x developer, no overpaid consultant - that can prevent this. And, as you cannot prevent these things, you must react to them, and to react to them, you need to know when something happens. That’s why it’s so important you constantly get feedback.

Many of the things that SEO involves are obvious, especially the content parts. You browse your web page; you see that there is no content or the content is rubbish, and you react accordingly. But, as you most likely have realized, many aspects of SEO are not so easily visible, for example: page speed, distribution, user experience, brand experience, canonicals, and the list goes on.

That’s why there are two tools which you must use daily. Otherwise, everything will go wrong. They are:

- Google Search Console

- Google Analytics or another accessible analytics tool

Small websites should look at Google Search Console at least every two days for a minimum of half an hour, and they should look daily for ten minutes at their usage values via Google Analytics. If you see something that you do not understand or you think looks worrisome, investigate deeper.

Bigger websites should have at least a part-time position for someone who looks at these tools. Really big websites should probably have their own team designated to them. These positions must be defined and their responsibilities (GSC, GA, report findings, initiate action) clear. These are the minimal viable SEO feedback setups.

An even better SEO feedback setup: once the platform is optimized, it must be easy for everyone to access the feedback of how their actions are affecting the website. Whenever somebody does something, they must receive direct feedback and know what it means, including the consequences. For example, if you have a marketing team, the marketing team must know what works and what doesn't. When did the brand searches rise? When did it drop? What gets inlinks, what doesn't? The development team must know the impact of their changes in the real world. Speed is not just the responsibility of DevOps (the server guys), it's the responsibility of everyone who touches the site. Distribution and traffic development must be set into the context of application releases. The content team must experience direct feedback how "their" content is performing, how much traffic it gets and why (search, social, internal). Feedback must be delivered where it belongs, to the people who do the actual work.

If the SEO feedback processes are implemented than a virtuous cycle can start. The content will get better (from both a SEO and user point of view), marketing will get more creative (as they will see that creativity works) and development will be more accountable and responsive. You will start to see results and be able improve on them.

CHAPTER 8: OTHER THINGS TO THINK ABOUT

We've almost covered everything you need to know about SEO - soon it will be up to you to execute all the things we have discussed in this book. But before that, there are a few other (but very important!) things you must know.

Mobile

Go responsive. Everything else is insanity. If you want to be convinced that mobile is important, then this is the wrong book and the wrong year. Mobile is a fact. The majority of all search demand for any vertical (including all B2B) is via mobile devices. There are three acceptable ways to communicate that you are a mobile ready site to Google. These are:

- responsive design
- user agent specific serving (one URL, two-page version)
- mobile redirect (two-page versions, two URLs, m. (m-dot)domains)

The big challenge when it comes to mobile sites is that the medium that was used to create the website, so that big, fat, powerful desktop computer or that fancy Macbook, is no longer the medium that the majority of users are using to consume it. That's why it's imperative that during the design process, mobile users are prioritized and tested accordingly. Any logic that attempts to create two completely different page versions (mobile and desktop) leads to underdeveloped, less maintained and often broken mobile sites. Adopting a responsive design across your whole webproperty mitigates this issue. If you wish to learn more about mobile SEO setup, then you will find everything you need to know in the Google spec [smartphone optimized sites google].

In 2016, Google announced, that they will, in the near future, render (see) the pages they index in the mobile view, and the desktop view will be just as an alternate view. Currently, as of the writing of this book in 2017, it is still the other way round. The details of how Google will deal with this major change can only be seen after they have actually switched to "mobile first indexing." Until then, relax, go responsive.

Render As Googlebot And Other Tools

Since 2014, you no longer have to guess how Google sees your site. That's because, in Google Search Console, you can choose to render web pages for Googlebot (Desktop) and Googlebot-Mobile.

Fetch as Google

See how Google renders pages from your website. [Learn more](#)

Leave URL blank to fetch the homepage. Requests may take a few minutes to process.

Desktop

FETCH

FETCH AND RENDER

Use this tool like crazy. This is how Google sees your site, and these are the views you are optimizing for. This is also the view we are using to determine the “fold,” so the content that is visible without scrolling. If Google cannot render sensemaking, good looking screenshots with this tool, then your site is broken and needs to be fixed. Now. Additionally, Google offers a “mobile-friendly test tool.” Take this test and if you fail, follow the instructions to fix your site. There is also Page Speed Insights that tells you how to make your mobile site fast. Use these tools to fix these issues, too.

App Indexing

App indexing is a technology that allows Google search results to point directly to content within native Android or iPhone apps. The challenge when it comes to App Indexing is that URLs are an alien concept to apps, and Google needs something to point to. So basically you have to twist your app to fit into the page paradigm.

You can find the spec at [Google Firebase App Indexing]. It's a horrible spec. App indexing has lost momentum in the years since it was introduced, with the overall traffic impact being less than overwhelming. So speak with your trusted native app developer if app indexing is a choice for you. They will most likely say no. Trust them.

AMP – Accelerated Mobile Pages

Accelerated Mobile Pages is an initiative by Google to produce fast (they call it “instant”) page loads on mobile. It’s a subset but at the same time an extension of HTML that can be pre-cached and pre-rendered by Google. This means, that when traffic is sent to an ‘amp’ed page, the user stays within Google (but traffic in GSC and GA is counted towards the webproperty). It’s all in the name of Google providing a fast user experience to its users, wherever they go. It’s a move away from responsive design (one HTML for all browsers, clients, and users) to a separate version of mobile users coming via Google. Google is also promoting AMP HTML as a way to become more deeply integrated into the Google mobile experience. Currently, news and recipe verticals are seeing a positive impact on their traffic because they are well integrated into the Google mobile experience.

The downside for businesses with AMP is you need to maintain (create, test, manage, quality assurance, and so on) yet another version of your pages, and you are limited in what you can do (with Javascript and HTML) within an AMP page.

From a technical point of view, you have to direct from your “normal” page a `rel=“amphtml”`, so for example, `<link rel=“amphtml” href=“https://www.example.com/a/slug/amp”>`, to the AMP HTML page. The AMP HTML page must point a canonical back to the “normal” page. The AMP HTML page must be according to the AMP HTML spec, and be validated correctly according to this spec. Also, there are some error reports available in Google Search Console about AMP.

My current standing on this topic is just to have a proper, working, unbelievably fast, long term optimizable, responsive website, and not do AMP. This recommendation might change significantly in the future depending on how the marketplace (a.k.a. Google) develops. On the other hand, if you just do not manage to create a fast website, then do AMP. That’s what it is designed for: to force publishers to have fast web pages in a way that they can not screw it up.

Multiple Domains And Languages

Another important decision for growing webproperties is how to deal with different markets and languages. My recommendations on this topic vary from company to company, and your specific business goals. However, here are some basic thoughts:

Always start with a single domain! From the sheer amount of guidance in this book, you will have realized that it's hard to have one well search engine optimized webproperty. You need to optimize the platform and the processes. Therefore, to keep things as simple as possible when you start to go multilanguage, use rel=alternate links on all pages with different language versions.

```
<link rel="canonical" href="http://www.example.com/en/seo" />
<link rel="alternate" hreflang="en" href="http://www.example.com/en/seo" />
<link rel="alternate" hreflang="de" href="http://www.example.com/de/seo" />
<link rel="alternate" hreflang="es" href="http://www.example.com/es/seo" />
<link rel="alternate" hreflang="fr" href="http://www.example.com/fr/seo" />
<link rel="alternate" hreflang="it" href="http://www.example.com/it/seo" />
```

Important note: Every targeted page must still be language consistent; the boilerplate (for example, the navigation) and main content must be the same language. Otherwise, the site will not perform. You must always make sure that the “new” languages have language namespaces (/en/, /de/, /es/ and so on). Language namespaces are not important for Google (language is determined on a page per page basis), but majorly important also for the manageability and measurability of the site:

- <https://www.example.com/en/understanding-seo>
- <https://www.example.com/de/seo-verstehen>

Make sure you submit different sitemaps for the different language versions.

The benefit of the single domain approach is that you still have only one webproperty to care about (meaning only one webproperty that you must make sure receives more and more inlinks over time). I have seen this approach work well for small sites, with up to a combined total of about 100,000 web pages (including multiple languages).

Note: If you go multilanguage, single domain, it's best practice to put all languages in a

language namespace. This is true for all but the startpage. The startpage must always be in the root. For the startpage, implement a logic like this where one language owns the root startpage:

```
<link rel="canonical" href="http://www.example.com/" />
<link rel="alternate" hreflang="en" href="http://www.example.com/" />
<link rel="alternate" hreflang="de" href="http://www.example.com/de" />
<link rel="alternate" hreflang="es" href="http://www.example.com/es" />
<link rel="alternate" hreflang="fr" href="http://www.example.com/fr" />
<link rel="alternate" hreflang="it" href="http://www.example.com/it" />
```

In this example, all other /en/ pages will be within the /en/ namespace. The startpage is the (annoying) exception, as you open yourself up to a whole area of SEO phenomenons if there is no startpage in the root of the domain.

Going Multidomain!

As soon as you are targeting different markets and languages or becoming bigger (and have more than 100,000 pages), I recommend that you go either multidomain or multi-subdomain. Adopting multiple domains allows the use of recognizable local extensions, for example, .co.uk in the UK or .de in Germany. In the case of going multidomain or multi-subdomain, use rel=alternate for market targeting over the pages of the different domains.

Multidomain

```
<link rel="canonical" href="http://www.example.com/" />
<link rel="alternate" hreflang="x-default" href="https://www.example.com/" />
<link rel="alternate" hreflang="en-gb" href="https://www.example.co.uk/" />
<link rel="alternate" hreflang="en-us" href="https://www.example.com/" />
<link rel="alternate" hreflang="de" href="https://www.example.de/" />
```

Multi-subdomain

```
<link rel="canonical" href="http://www.example.com/" />  
<link rel="alternate" hreflang="x-default" href="https://www.example.com/" />  
<link rel="alternate" hreflang="en-us" href="https://www.example.com/" />  
<link rel="alternate" hreflang="en-gb" href="https://uk.example.com/" />  
<link rel="alternate" hreflang="de" href="https://de.example.com/" />
```

That said, having a multimarket or multilanguage site is hard. Not only twice as hard, but complexity and effort squared. As soon as you go multilanguage or multimarket, the versions of the website begin to diverge. It's a whole new layer of complexity to manage. But although you have more work, you will also have more manageability with multiple domains. Attempting to do all this on one domain for a growing website is project suicide.

My conclusion on this topic is that overall, it depends. Rel=alternate is your friend. Check the Google spec and test via Google Search Console to see if your targeting strategy is working correctly.

SEO Success

If you've made it through the book, congratulations. Now you know: there is no secret to SEO. It is not a mystery or a trick, nor can it be download or sourced out to SEO agencies that don't touch your site or your business. SEO is solved. I have told you what you need to do, and why you need to do it, and hopefully put myself out of a job in doing so. All you need to do is apply it to your business. Remember, SEO is the business of getting found, and SEO is a choice. Choose wisely; the success is up to you!

----- The End -----

INDEX

Accelerated Mobile Pages	8, 141
AdSense	47, 137
AdWords	5, 47, 137
App Indexing	8, 140
A Red Line	7, 58
Brand Building	6, 46
Content Marketing	24
Crawling	112
Database Crossing Targeting	7, 67, 71
Database Driven Targeting	7, 67, 70
Deadweight	119
Development	21
Distribution	6, 7, 35, 83, 103, 116, 138
Diversity of Content	94
Domains	8, 142
Do-queries	32
Feedback	6, 8, 17, 37, 39, 50, 137, 138
Google Adsense	38
Google Adwords Keyword Planner	51
Google Analytics	17, 20, 38, 50, 51, 58, 62, 64, 136, 138
Google Autocomplete	51, 70
Googlebot	8, 109, 140
Googlebot-Mobile	140
Google Chrome	38, 60
Google Data Saver	38
Google DNS	38
Google Image Search	94
Google Keyword Planner	70
Google Page Speed Insights	60, 61
Google Suggest	51, 70
Google Toolbar	38
Google Trends	51, 70
Google Webmaster Guidelines	110, 129, 130, 131, 132
Go-queries	32
HTTPS	55, 59, 85
Image Onebox	94

Indexing	8, 114, 140
Inlinks	6, 44, 49
Key Performance Indicator (KPI)	19
Know-queries	32
Languages	8, 142
Link Building	6, 27, 44
Long Click	37
Marketing	6, 21, 24, 27, 42, 46
Meta Robots	89
Mobile	8, 139, 140, 141
Multidomain	8, 143
PageRank™	11, 12, 13
Page Speed Index	60
Private Sitemaps	7, 111
Query-Driven Targeting	7, 67, 72
Retroactive Targeting	7, 75
R.T.F.M.	6, 15, 17, 51, 89, 100
Schema.org	100, 101, 102
SEO Manager	26
SEO Point Of View	6, 25
SEO Setup Recommendation	6, 25
Sitemap Notifying	7, 108
Sitemap Stacking	7, 103
Sitemap.xml	119
Sliding Content	126
Speed	7, 27, 59, 60, 61, 138, 140
Startpage	7, 117
Targeted Page Generation Logic	7, 65
Targeted Search Intent Phrase	66
The Brand Experience Feedback Loop	6, 39
The Canonical	87
The Meta Description	91
The Quality Criteria	7, 67
The User	6, 29, 31, 32, 37
The User Experience Feedback Loop	6, 37
update.rss	116
Updating Sitemaps	7, 115
URL Best Practices	7, 65
URL Rules	7, 61, 62

WordPress 22, 24, 27
Yoast 24

