

# Structuring Best Practices of Search Engine Optimization for Webpages



Riaze Miguel Ribeiro Issá  and Jose Paulo Marques dos Santos 

**Abstract** The main objective of search engine optimization (SEO) is to present the most relevant information to the user, so that his necessities may be satisfied. Therefore, it becomes imperative to understand and be familiarized with the constant developing algorithms, on what mainly concerns the variables it gives the most importance to. In this article, an in-depth explanation is presented, from a basic to a more advanced level, of the internal and external variables required for a good positioning of website pages, enabling them to escalate in the search engine results page (SERP). Such holistic approach is useful, so that the reader, using the most appropriate techniques and tools, may be one step ahead of the remaining competition.

**Keywords** Search Engine Marketing (SEM) · Search Engine Optimization (SEO) · Search Engine Results Page (SERP)

## 1 Introduction

In a digital environment that is under constant development and rapid expansion, where just one Internet connection provides unlimited access and the sharing of massive amounts of information, where there are millions of websites in which the daily increment of information is immeasurable, increasing difficulty in finding reliable and quality information, the need of creating relevant pages for user queries is of pivotal importance. The strive for a digital presence that is visible to the user has been driven by search engines that feed off this struggle which, in turn, fosters their development. The search engine algorithm will always give priority to the pages of optimized websites and, therefore, the strive for top and prominent rank positioning is achieved by websites with the best optimized pages for certain queries.

---

R. M. R. Issá · J. P. Marques dos Santos  
Universidade Europeia, Lisboa, Portugal

J. P. Marques dos Santos (✉)  
University of Maia - ISMAI, Maia, Portugal  
e-mail: [jpsantos@umaia.pt](mailto:jpsantos@umaia.pt)

This paper presents recommendable search engine optimization practices and shows that, when used properly, they make it possible for a website to achieve a relevant position in relation to its competitors. The concepts presented in this study stem from the basic to the more advanced by means of a more comprehensive approach to SEO, by defining what it is and once the basics are covered, it provides details and in-depth information on the potential strategies that can be applied inside or outside websites, aiming at improving the visibility of pages while maintaining consistency in their optimization.

## 2 Search Engine Marketing—SEM

The main objective of search engine marketing is the development of strategies that increase traffic on certain website pages, whether these are organic—SEO, or by means of ads—SEA [3]. SEM can be highly complex due to the large number of factors that influence its behaviour, whether these are the user or the search engine algorithm. The volume of searches for a given page has a very significant impact on the moment of positioning [7].

The organic search results organized by the search engine results page (SERP) are ordered by their relevance according to the user's choice of keywords. It is important to know how to distinguish organic results from paid results. Paid results are strategies developed in which the marketer pays the search engines to be displayed in ad format at the top of the page. The goal of any user is to find relevant information according to his query that can be broken into different categories: navigational, transactional or informational. The most important fact when creating a SEO digital marketing strategy for a website is to develop it with a deep understanding of the target audience. Understanding how most user searches and target market searches work can help to have a more effective impact on the user and, subsequently, to find ways to target new leads [13].

Query formulation can be carried out by using words or short phrases. Most users formulate their queries by using one to three words. When the search engine returns the results of a query on the SERP, it is possible to determine if the contents are the desired ones by observing these results or in other words, if they correspond to the expected response to the query or if on the contrary, they require users to refine their search until the desired results are obtained. Thus, when users find the content they need, this query translates into a satisfying experience because the user was able to complete the intended task [13].

### 3 Search Engine Optimization—SEO

Search engine optimization is, by definition, one of the most effective marketing strategies in gaining organic traffic by means of processes such as writing, programming and design. All these efforts help in the gradual increase of the volume, quality and visibility of any website page which users can access by means of a search engine [22]. Williams [43] states that SEO, search engine optimization, includes all the steps that a webmaster (whoever designs and maintains a website) takes to ensure the visibility of their pages on Google's results page, SERP. SEO strategies help in the visibility of digital platforms not only from a technical perspective, but also in an equally creative way that helps improve the ranking position in organic searches and achieve high rankings on SERP's [24].

The creation of strategies and their implementation in SEO are not static processes. While observing results, the webmaster must resort to his know-how and flexibility to change or update optimization techniques. This constant transformation which takes place by means of experimentation must meet previously established objectives. As such, this manipulation of a website's characteristics should be considered to improve its positioning on the SERP in relation to competitor pages [25]. Within this scope, Zilincan [47] states that SEO is characterized as the set of modifications and techniques that help search engines to collect, index and interpret the contents of a website. Zhang and Cabage [45] explain that within search engine optimization, SEO is characterized as a set of techniques, namely content marketing, link building and social sharing. A well-developed SEO strategy is critical to the success of online businesses.

According to Chotikitpat et al. [9], the search engine algorithm rules must be strictly followed in order for page contents to be properly indexed. With this traffic-boosting art inside a website, there will be new visitors and leads and, in turn, not only will increase user engagement but also brand perception. Once users see the reputation within the brand or, in other words, the authority within the segment, this will boost the probability of an increase in the volume of sales.

### 4 Google's Guidelines

In this SEO reality, not all strategies and techniques are consensual, and the scientific community is divided between what is to be considered good practices and what is not. There are webmasters who follow Google's guidelines and webmasters who use algorithm manipulation to achieve their goals.

White hat SEO include all strategies that aim to place a page in a high-ranking position by always following the guidelines, i.e. the SEO strategies approved by Google itself [43]. These are techniques that contribute to the ranking of a page [23] in the form of quality content, engagement and user-friendly navigation. According to Berman and Katona [6], White hat SEO contributes to the content of a website

effectively and positively which, in turn, translates into consumer satisfaction and consequently, makes the website more relevant from the user's and the Google robot's point of view.

On the other hand, black hat SEO encompasses all the manipulation and exploitation of Google loopholes to position pages with the risk of being penalized or banned if these practices are discovered [43]. Accordingly, Killoran [23] characterizes this type of SEO by the set of unconventional initiatives that boost traffic, namely the use of spam and irrelevant or thin content. Moreover, Aswani et al. [3] add that this way of earning more traffic has its dark side considering that in the long run, black hat SEO is ephemeral due to the constant regulation and updating of Google's algorithm mechanisms, which identify and penalize these types of strategies. According to Zhang and Cabage [45], these types of penalties result in the demotion of the website on SERP positions and result in fewer impressions, or the website may also be banned and blacklisted by Google in a worst-case scenario. Most websites that are caught by the search engine have a kind of black hat, which can include practices of keyword stuffing, duplicate content, spamming, link farming, cookie stuffing and doorway pages [20].

The grey hat SEO lies between these two extremes. Although not approved by Google, these types of SEO strategies have, unlike black hat SEO, a low probability of being penalized. Grey hat tactics have a higher risk than white hat tactics, but not as high as black hat tactics [43].

## 5 Long Tail Keywords and Keyword Research

Anderson [2], the creator of the long tail concept, states that there is a clear division between two product realities: mass market products and niche market products.

According to Enge et al. [13], there are search terms that receive more than five thousand searches per day but correspond to only 30% of all searches carried out on the Web—long tail searches, as opposed to the most common searches, short tail searches that are 70% of all Web traffic. The keywords long tail add more relevance to a given user's query because they are the most descriptive and precise and therefore confer more value to the traffic generated for the pages in question. Understanding the demand for each search is pivotal, considering that there is a need to create long tail content, i.e. niche content that is not searched for by all users, but by more specific users who are interested in the topic. Before starting any SEO project, it is necessary to establish the keywords target because the entire content and optimization strategy will stem from them. Market trends may be understood through analyses, as well as potential trends within the segment, because they make it possible to foresee the increase in demand for certain products. This way, it will be easier to find the best search terms and meet consumers' search needs. Considering that any search term that is typed into the search engine is recorded, the use of keyword search tools is vital. Although important, these tools do not directly show how valuable or important a keyword really is, as it is necessary to understand the two variables of a keyword,

its difficulty (keyword difficulty—KD) and its search volume (SV). Also, important to note is the metric scale of these two variables which in the case of keyword difficulty, goes from 0 to 100%, and in the case of search volume, which in theory can be unlimited, can start at 0 and exceed 100%.

When selecting keywords, there are aspects to take into account. A search that is carried out very randomly or broadly may cannibalize other keywords. This process is counterintuitive because it does not show an internal hierarchy on the pages. Consequently, the search engine crawler is forced to make its choice between what is more and less relevant, creating a serious optimization error. Therefore, search, hypothesis formulation and testing must be targeted at the SEO goal [13]. Within this scope, Ramlall et al. [35] state that a short tail strategy is more challenging than a long tail strategy, considering that short tail keywords have more competition due to fact that they facilitate queries as opposed to long tail keywords, which are more specific, have less search volume and queries are less common. This way, an ecosystem of billions of unique queries is built [24]. Webmasters should always carry out keyword analyses from the users' point of view so that there is a greater degree of understanding of the keywords and, in turn, insert them into the SEO strategy, thus ensuring qualified traffic to the website pages [41].

## 6 On-Page Optimization

As described by Yalçın and Köse [44], on-page optimization consists of all website improvement tactics that are carried out internally. The goal is to make sure that the entire page construction of a given website is readable to search engine indexers or in other words, the capture and transformation of HTML and CSS code into data that describe and rank this Web property to perfection. Thus, Ghulam et al. [14] state that there are a set of variables which together make up the on-page optimization of any website. As such, a correct and complete on-page optimization depends on tree structure, the selection of the name for the domain, the page name, the title metadata, the site description, the meta keywords, the organization of the URLs, the heading tags, image alt texts, the written content on the pages, website navigation, its sitemap and robots.txt status.

In the hierarchical structure of a website, pages should be organized in a top-down format with a static URL strategy that facilitates their indexing in search engines as for example, the search carried out by users. There must also be a semantic relationship with the brand and the name of the product or service for which it is targeted, hence the use of hyphens or addition symbols to separate the terms identifying the subpages.

## 6.1 *Title Tags*

The function of a page title or title tag is that of being merely indicative of the themes and sub-themes of the pages, as well as to help guide users in their search. The title tag is ranked by the search engine algorithm, so it is critical that the theme of the page correlates with the user's search context (UI) [33]. According to Enge et al. [13], to achieve this optimization, the length of the title tag must be limited and no shorter than 50 to 60 characters long. There should be no duplicate title tags and the main keyword must always be present, ideally as far to the left as possible [46].

## 6.2 *Meta descriptions*

According to Khraim [22], meta descriptions are visible to the eyes of a consumer but they should focus on the search engine and its respective indexing. By norm, the function of meta descriptions is to provide a summary of a page's contents in order to persuade the user to click on the page.

According to Enge et al. [13], considering that a meta description is created by the webmaster, their writing should be clear, objective and promotional within a minimum of 70 characters and a maximum limit of 155 characters to be able to boost website traffic. The correlation between the user's query and the keywords inserted in a meta description will not only allow the search engine to read their content, but also to rank the most relevant pages according to the UI for the user within its indexing process [46].

According to Williams [43], this variable, as well as the page title, is primarily responsible for increasing click-through-rate (CTR). The more clicks the page receives, the better its ranking position on the SERP and more importantly, the satisfaction of a user's search need is achieved. A page that is in the first position receives, on average, about 31% of all clicks generated on that page on the SERP. Next, the page in the second position receives on average 14% of all clicks, while the page in third position receives 10% of clicks. On the other hand, pages on the second page of the SERP receive about 4% of clicks and those on the third page receive only 1.6% of clicks.

It is important to understand the importance of good ranking position. The higher it is the better, considering that a drop in one or two positions is enough for a significant decrease in traffic. For example, pages between the sixth and tenth position on the first page of the SERP have a CTR of only 3.73%, which is less than the CTR of the page in fifth position which is about 5%. [43].

### **6.3 *Meta keywords***

When Google's algorithm first appeared, webmasters discovered that the search engine considered meta keywords to be very relevant, and they were one of the key ranking factors. Thus, they began to over-optimize this field by means of keyword stuffing practices to manipulate rankings on the SERP. However, Google identified this practice and started ignoring this tag although not completely because it is still used to identify spammers. It is currently considered a good practice to leave the meta keywords field empty [43].

### **6.4 *Heading Tags***

Heading tags are titles, and their function is to organize and rank the content within a page. These titles and subtitles must reflect the content of the page to inform the algorithm and grab the user's attention, because the more interesting the content, the longer the user will stay on the website. As stated by Busche [8], Heading tags serve to provide greater organization and sequence of titles and subtitles, both for the user and for the Google crawler. The H1 or Heading tag 1 is the main and most relevant title, so there can be no other H1. It is unique. The subheadings that follow range from H2, H3, H4, H5 and H6, the latter being the least important on the hierarchical scale. Heading tags play an essential role in organizing content delivered to the user [13]. Heading tags should contain the main keyword, mainly the H1, H2 and H3 as they are the second most important on-page factor [46].

### **6.5 *Images and Media Alt-Text***

According to Bailyn [4], search engines are only able to read text. During their indexing process, when the crawler identifies an image or other type of media file such as videos, this type of content is discarded.

With major technological advances leading to the consumption of other types of content other than text, alternative texts (alt-text) were created. They are basically tags that are used to describe non-text content to the crawler. For proper optimization, the alt-text should contain the keyword or descriptive phrases of the page topic [46]. In agreement, Enge et al. [13] stress that search engines only accept HTML language and refuse any other type of format present in websites. However, by placing Alt attributes, it becomes possible to name attributes such as images, audios, videos, Adobe Shockwave files or even another type of computer language such as Java, so that the algorithm can understand the content exposed in the digital property. Choudhari and Bhalla [10] indicate there is a specific way to optimize videos called video search engine optimization (VESO). This optimization takes variables such

as the title, the description, the transcribed script, the thumbnail which describes the content with an image, as well as annotations that promote linkbuilding through external links. As such, there must be thorough knowledge of the market segment and the target to whom you want to pass the message to, so it is possible to make extraction of keywords more effective because these are the ones that have a higher search volume of the target audience.

## ***6.6 Taxonomy and Ontology***

Enge et al. [13] state that taxonomy is essentially a two-dimensional hierarchical model of a website's architecture. It is essential in the presentation and arrangement of the website's themes to the consumer because the better this architecture is designed, invariably, the better the consumer experience will be. According to the same authors, ontology is a way to map the human mind. When people think about a certain topic, it is common to interrelate it with other topics, and this is important to consider when planning the categorization and creation of clusters of similar and correlated topics within the site. In this same line of research, Batsakis et al. [5] defend that it is the taxonomy that allows grouping similar topics or categories to boost the user experience by delivering more content and making it easier for them to navigate on pages.

## ***6.7 Sitemap***

The sitemap is the map of a website that contains a cluster of URLs and other files, such as extension files (PDFs, TXTs, DOCs) that make up the virtual property. According to Ziakis et al. [46], the XML file is created by the website's developer, and it is submitted to make it easier for the crawler to find all the subpages of a website. Additionally, this file notifies the search engine when any change occurs at the website level or even certain aspects at the page level. According to Visser and Weideman [42], the XML extension is used to indicate the indexing of pages to the search engine on the website so that they appear on the SERP. This same indication can also reinforce the priority of the pages you want to give more relevance to base on their traffic, considering that the higher the volume of traffic is, the higher its priority on the site should be. However, Enge et al. [13] guarantee that it is not by adding a new URL to the sitemap that its indexing is guaranteed. It is essential for the site to have its sitemap optimized as this allows for the customization of URLs and metadata if they are duplicates or hard for the crawler to collect. In addition, it is also possible to change or choose canonicals.



The sitemap must be present in robots.txt because according to Google [16], the robots.txt file is automatically read by the robots when they arrive at a given website. This file should contain “commands” for the robots, such as pages we want or do not want to index.

## ***6.8 Responsive Navigation, Page Speed and Structured Data***

When working on the hierarchical structure of a website, there are two sides to consider, the human’s and that of the robot’s. From the user’s point of view, i.e. the human side, the site must respect how the user usually consumes information, namely the menus in the header, the fact that reading is done from left to right, the webpage layout should avoid too much sensory information for users so that they do not get dispersed or frustrated by not being able to absorb everything on a particular page. Each webpage must be designed with the user in mind, so that they enjoy a good UX, with a simple and optimized navigation and no distractors. From the point of view of the crawler, it must be able to read and assimilate the most important page details. For this to be possible, the tree structure must also be optimized [43].

According to Lali et al. [24], the loading speed of a website’s pages (pagespeed) is a determining factor in the positioning of results on the SERP. Within this scope, Manhas [26] reinforces this line of thought by stating that for a good performance of pagespeed, it is necessary that the internal and external elements of the page in other programming languages—such as HTML, CSS, JavaScript and json—improve aspects related to site navigability. High-performing websites have higher percentages of user retention, engagement and conversions. For Ziakis et al. [46], loading time is an extremely important factor, considering that engines include it in their algorithm. The longer the loading time of a page, the worse its position on the SERP will be.

One element that helps the crawler in its interpretation of page information is structured data. This type of interpretation can be found on the Google results page from a previous HTML construction. This type of classification is a great support in sending information to the search engine so that it can categorize and present the data in the best way possible [24].

## ***6.9 Bounce Rate and Time on Website***

The bounce rate happens when the user clicks on one of the SERP results and then returns to the Google results page again. The faster the user returns to the results page, the greater the user’s dissatisfaction is identified by Google, penalizing the reputation of the website. On the opposite side of this experience, we find the user who navigates from page to page on a website increasing their dwell time and informing Google that the webpages are relevant and, therefore, it places them in better positions [43].

In agreement, Ziakis et al. [46] state that websites with low bounce rate inform the search engine that this is a high quality and extremely relevant website for the user.

In addition to Ghulam et al.'s [14] perspective, there are more key variables to take into account when a webmaster analyses the on-page optimization of a given website, namely: the status code of the various URLs that make up the website; the SSL security certificate; the structure of the URLs; checking the status of indexing tags and authority; the depth of the website and the correspondence between the website's language and the programming language.

## 6.10 Status Code

According to SEMrush [39, 40], HTTP status codes represent all requests sent to a Web server by search engines or platform users. Multiple webpages with 4xx and 5xx codes can negatively affect user and crawler experience, resulting in traffic drops, which in turn lead to loss of organic positions.

When analysing multiple status codes, it is essential to establish a priority in checking the code statuses and fixing existing errors. Priority should first be given to errors that affect the website, be it in the presentation of results to the user or to the search engine, and only then to less harmful codes. As such, the status codes that come up in a status code analysis more frequently are 2xx, 3xx, 4xx and 5xx.

The 2xx status codes indicate that the communication was successful and that the Web server was able to respond to the user's request [31].

The 3xx or redirect codes are an indication for a certain internal URL to redirect to another URL. This can take on seven different names: 301, 302, 303, 304, 305, 306 and 307. However, a special preference should ideally be given to the 301 redirect—permanent redirect, since it is the redirect that ensures that there is absolutely no loss of URL authority. On the other hand, 302 redirects are also quite common. This type of redirect is temporary, indicating to the search engine in question that the user is temporarily being redirected to another page, and thus there is no passing of authority. These should be done with caution, as a slight change will cause discomfort and a bad user experience, which the search engine will detect and take action, according to the severity of the situation [31].

4xx Errors indicate that there was a request to access a certain page or when access to a webpage is restricted. The most common errors are 401—Unauthorized; 403—Forbidden; 408—Request Timeout and 404—Not Found. This type of URL jeopardizes the user experience within the website and penalizes users, considering they cannot access a certain page or file through the link they click on. Additionally, internal links that send users and the search engine crawler to dead end pages imply a spending of the crawl budget (number of pages that are read by the crawler on a daily basis) with a subsequent penalization. Webpages with status code 4xx will be removed from the Google Index [31].

5xx Errors or Server Errors indicate that there are problems with the server. These problems should be immediately reported to the website developer or

hosting provider of the digital platform and resolved with the utmost urgency. The most common errors include misconfigurations, server overload or poor server performance [31].

### **6.11 *SSL Security Certificate***

According to Ziakis et al. [46], the acronym SSL stands for secure socket layer, a protocol that establishes a code between the server and the browser which allows information to be transmitted securely. This certificate ensures that the website is trustworthy, that it protects user data and prevents spam tactics. Websites with an installed SSL certificate use the https protocol instead of http. Thus, sites with an SSL certificate tend to rank better on the SERP than others.

### **6.12 *URL Structure***

According to Ahrefs [1], in the optimization of URLs, Google's algorithm gives preference to friendly URLs, that is, those with a simplified structure that allows the user to know where the website is located, as well as deliver easy-to-read information to the crawler, thus also improving its loading speed. When a new page is created, a new URL is generated and as such, knowing its structure is vital.

The URL protocol is the first element of analysis, and it is characterized by a colon, two forward slashes, followed by the host name. It can be referred to as a transfer protocol or scheme and will determine how information is transferred between the host and the user who is carrying out the search [1].

The choice of a good domain is essential when creating a website because good accessibility represents the users' gateway to the Internet. During the selection process, the choice for an exact domain strategy (EMD) as Williams [43] states, is a type of strategy that will be penalized by Google itself, due to the unfair advantage positioning it represents. On the other hand, the subdomain is the extension of the primary domain and is used to organize the different content on a website and for redesign functions. It is most often used in blogs and eCommerce stores [1].

The top-level domain (TLD) is a suffix related to the last letters after the period in the URL. There are over 1000 possibilities of TLDs which a webmaster can choose from, considering that the most used are the geographic location ones that are typically related to where the website domain was created [1].

The subfolder and slug are the organization and structuring of the website. By norm, a subfolder is the category pertaining to some broader topic that, in turn, unfolds into more specific content, the slugs. In a blog, a subfolder would be the main page which unfolds into other more detailed webpages on a part of the base topic. In the case of ecommerce, it is the category or parent page, which in turn incorporates other product webpages related to the category described [38].

An idea and good practice in the slug to take into account is to use the keyword of the page and, if there is more than one term, resort to the use of hyphens. This should be short and objective, excluding the use of special characters such as capitals, underlines and numbers [1]. In the same research line, Ziakis et al. [46], state that including the main keyword in the URL is an important positioning factor and that it presents a simplified URL structure, considering that the crawler will identify the webpage theme in a simpler way.

### ***6.13 Noindex and Nofollow Tags***

According to Google [15], the content attribute instructs the search engine on how it should crawl and index the pages. If there is no metatag blocking certain webpages in robots.txt, the crawler will interpret the information to index and follow the webpages. By doing this, it is giving itself permission to show the pages on the SERP and read all the links from them. On the other hand, pages with the noindex metatag will never be displayed on the SERP and read by the crawler. However, for the metatag to work, pages that are not intended to be indexed cannot be blocked by robots.txt, otherwise the bot will not read the information [18].

### ***6.14 Crawl Depth***

Crawl depth is extremely important for both the user and the crawler. According to SEMrush [39], search engines consider the most relevant pages to be those which, starting at the homepage, require the fewest clicks to be accessed. A good navigation structure allows access to all website pages with a minimum number of clicks. Ideally, no more than three clicks.

### ***6.15 Canonical Tag***

The canonical tag provides information to search engines about which version of the page has priority. It must appear in the sitemap and be indexed. Canonicals are extremely useful in situations where the same page is presented with different URLs (e.g., parameters, pagination) or in cases where there are slight content variations (e.g., product variants) [28].

## 6.16 Hreflang Tag

The hreflang tag is useful when there are multiple versions of pages in different languages, and it indicates to the search engine these different variations and shows the appropriate version of the page according to the language of the given region [37].

The hreflang tag must be in line with the `<html>` lang attribute which, in turn, must be configured on the website [32].

## 7 Page Rank—PR

There is one essential factor that is a strong indicator in the presentation of results. This technology is called Page Rank, PageRank or simply PR, and it is the algorithm used for Google searches that effectively positions website pages on the SERP. PageRank was named after one of Google’s founders, Larry Page. This algorithm is measured on a scale of zero to ten but in fact, no page has rank zero. The page is unrank, which is a value very close to zero. It is important to note that Google can index pages with no rank [43].

When a page links from page A to page B, page B, the destination page increases its PR in proportion to the PR of page A, the source page. If the source page, A has a PR of zero, the destination page, B receives only a tiny portion of its PR. However, if page A has a high PR, page B receives a huge leverage.

If a page has a lot of inlinks, it will receive PR from the pages that are sending those inlinks, making it a more relevant page. A simple way to understand this concept is to consider links as “votes”. The pages with the most votes are the most relevant pages from Google’s perspective [43]. The more authority and quality links a page has, the more PR it receives and, in turn, the more relevant it becomes from Google’s perspective. In other words, the page gains reputation and status. The more reputation a page has, the better it will rank on Google’s SERP. Nowadays, Page Rank is still a topic that is a subject of debate and discussion in the scientific SEO community. Taking into consideration that this is the technology used to position pages, it is essential to understand the metric and the variables that enable this process on Google [43].

Today, the Page Rank information that Google made available when SEO first appeared, namely with its first algorithms, is completely hidden from all users, webmasters included. However, multiple companies have invented tools that make it possible to simulate the search engine algorithm with the aim of coming close to the real authority of the websites.

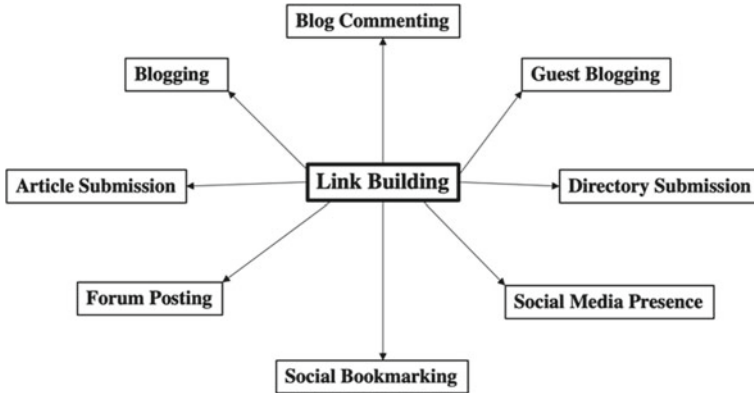
Among the most successful tools, Moz stands out. Similarly to PR, Moz has a scale that goes from 1 to 100 to evaluate both the authority of a domain (domain authority—DA) and of a given page (page authority—PA). The domain authority—DA is calculated based on several factors, of which we can highlight the number of

links to the domain root and the total number of links [29]. This score is the closest way to reality and is very useful in comparing sites when checking the reputation of a given domain. On the other hand, the DA is, as its name indicates, the authority of each web page. The set of all webpages of a given website is called domain authority, the DA [30].

## 8 Off-Page Optimization—Link Building

As described by Yalçın and Köse [44], off-page optimization includes all the website improvement tactics carried out externally. The goal of a link building strategy is to connect multiple platforms, namely brand and other platforms, to gain greater awareness and digital maturity, which in technical SEO language is better known as increasing authority. A page with more links, i.e. both internal and external hyperlinks to other websites, is the best signal to the search engine that that page is relevant to users, and therefore has a larger digital footprint, a closer-knit community, a greater link of trust and effectiveness in presenting results [19]. In Killoran's [23] view, working on link building as a way to gain authority involves knowledge of attributes, namely how to gain popularity, relevance and trust for user search. According to Enge et al. [13], the authority of a webpage is the unit of measurement of its reputation. This authority is measured by having multiple links to different websites, resorting to the hyperlink method, where it transfers authority via the different links. As a norm, there should be a relationship between the links, such as interaction sites related to the business area (core business) or social networks, so that this management can be done correctly. A higher authority is correlated with better quality content, since it is this type of content that generates greater social engagement by existing platforms, thus helping the positioning of the page on the SERP. For Deitel [11], the link building strategy is based on three pillars, namely the reciprocal link, where there is a mutual exchange of links; the natural link, in which the link is created through a content marketing strategy (for example, a blog article); and link baiting where there is a fostering of awareness in relation to a certain content.

When comparing link building and social sharing strategies, the link building strategy is undoubtedly the one that generates more positive results, such as greater traffic generation for the website, which results in a growth of authority that, in turn, allows for a better positioning on Google's results page. Social sharing proved to be a less efficient strategy because the volume of traffic was lower when compared to link building strategies [45]. A link building strategy can be approached in different ways (Fig. 1), such as article submission, directories, forum postings, presence on social platforms, blogging (which is based on content marketing), and guest blogging, which is the creation of written content pieces to be published elsewhere with the intention of placing the anchor text and receiving the backlink from the site where it was published [21]. Anchor text is a key element within content pieces which make it possible to link to other pages, either in or out of the website [27].



**Fig. 1** Link building strategies

For an understanding of domain authority, it is essential to understand how links work. There can be two types of links: internal and external. Additionally, they also have two attributes: nofollow and dofollow. Nofollow links are those which tell the search engine robot (*robots.txt*) not to follow a certain link, thus not passing authority. Unlike nofollow links, dofollow links give the robot permission to follow them, in turn passing authority to the domain [17].

## 9 Content Marketing

Search engine optimization and content marketing are two inseparable realities that need each other. SEO generates traffic and optimizes content while content marketing creates the content. According to Pažėraitė and Repovienė [34], this form of marketing exists when there is good SEO work supporting its entire structure, considering that the more traffic to a page, the more users will consume the content on that page. In the same line, Rowley [36] states that it is through good pieces of content, that positive emotions are attracted and generated in the user, with the final result being a higher degree of loyalty, satisfaction and trust. In this light, Di Gangi and Wasko [12] state that personalization of communication between users has a significant and positive impact on their social interaction during their navigation, as this engagement is defined as an individualized concern. Moreover, when users receive a type of communication that is personalized and meets their interests, greater satisfaction is inherent, with the final result being that of greater engagement. It is also necessary to consider that interaction quality must be based on the amount of information about the user, and once you have this information broken down and analysed, the greater the likelihood of success in engaging a lead. That is why it is so important to focus on the quality and not the quantity of the message, where in

the second case, the brand may become associated with negative feelings due to its high intrusiveness and lack of interpretation of needs [12].

In SEO, pages that want to rank high in search results need to be unique, authentic and have relevant content. Thus, all quality content must respond to two attributes: it must match search engine queries and it must contain links. Keyword research is essential to ensure that they are searched for by users [46].

The more quality content each page has, the more likely it is to rank correctly in Google. A recommended number would be 500–1000 words per page. Otherwise, when reading the page, the search engine will struggle with missing keywords and have a harder time positioning the pages [37].

When writing content, it is important to understand what the topic of the page is. Next, considering which keywords are best to include in the content, to meet the search carried out by users [41]. A competitor analysis should be carried out, especially those that have a good positioning in search engines, to have an idea of the type of content that generates good results.

At the time of writing, including the keyword of the page with proper sentence coherence is recommended. Keyword density is the number of times a given keyword appears in the text of a page when compared to the rest of the website text and can be calculated by the following formula:

$$\frac{\text{number of times the keyword appears}}{\text{total words}} \times 100$$

It is a key concept to consider when creating content [46].

Another important factor to take into account is the linking of each page of the website. The intention of creating internal linking is not only to make the user's visit more enjoyable, but also to facilitate the paths and indexing of the individual components of the website which consequently leads to improved page positioning [46].

## 10 Conclusion

We hope that the description of the techniques and tools that aim to improve the positioning of website pages in search engines have contributed to the understanding of the variables that require greater attention, so as to enable the search engine algorithm to carry out a better reading of all the information presented. For such a search, the greater the detail of each optimization, the greater the probability of a better ranking positioning on the SERP in relation to competitors.

The SEO approach must be a holistic one or, in other words, it must take into account several variables in order to understand which ones have the greatest impact and if they are correctly optimized, so as to achieve consistency in the presentation of the results and in turn, in their positioning.



It is essential to monitor results and constantly track page positions because the search engine algorithm is constantly being updated, and it is this monitoring that makes it possible to assess the impact of algorithm updates on all SEO work.

## References

1. Ahrefs: How to create SEO-friendly URLs (2021). <https://ahrefs.com/blog/seo-friendly-urls/>
2. Anderson, C.: *The Long Tail*. Hachette Books (2008)
3. Aswani, R., Kar, A.K., Ilavarasan, P.V., Dwivedi, Y.K.: Search engine marketing is not all gold: insights from Twitter and SEO Clerks. *Int. J. Inf. Manage.* **38**(1), 107–116 (2018). <https://doi.org/10.1016/j.ijinfomgt.2017.07.005>
4. Bailyn, E.: *Outsmarting Google*, 1st edn. Que Publishing (2011)
5. Batsakis, S., Petrakis, E.G.M., Milios, E.: Improving the performance of focused web crawlers. *Data Knowl. Eng.* **68**(10), 1001–1013 (2009). <https://doi.org/10.1016/j.datak.2009.04.002>
6. Berman, R., Katona, Z.: The role of search engine optimization in search marketing. *Mark. Sci.* **32**(4), 644–651 (2013). <https://doi.org/10.1287/mksc.2013.0783>
7. Brooks, N.: The Atlas rank report : how search engine rank impacts traffic which factors determine click volume ? Atlas Institute, 1–4 (2004). <http://94.126.173.33/ad2006/adminsc1/app/marketingtecnologico/uploads/Estudos/atlasonepoint-howsearchengineimpactstraffic.pdf>
8. Busche, L.: *Powering Content*. O'Reilly Media, Inc. (2017)
9. Chotikitpat, K., Nilsook, P., Sodsee, S.: Techniques for improving website rankings with search engine optimization (SEO). *Adv. Sci. Lett.* **21**(10), 3219–3224 (2015). <https://doi.org/10.1166/asl.2015.6503>
10. Choudhari, K., Bhalla, V.K.: Video search engine optimization using keyword and feature analysis. *Procedia Comput. Sci.* **58**, 691–697 (2015). <https://doi.org/10.1016/j.procs.2015.08.089>
11. Deitel, H.M.: *Internet & World Wide Web*, 4th edn. Pearson (2007)
12. Di Gangi, P.M., Wasko, M.: Social media engagement theory: exploring the influence of user engagement on social media usage. *J. Organ. End User Comput.* **28**(2), 53–73 (2016). <https://doi.org/10.4018/JOEUC.2016040104>
13. Enge, E., Spencer, S., Stricchiola, J.: *The Art of SEO Mastering Search Optimization*, 3rd edn. O'Reilly Media, Inc. (2015)
14. Ghulam, A., Depar, M.H., Ali, S., Rahu, S.: On-Page search engine optimization (SEO) techniques model: a use case scenario of a business. *Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET)* **5**(11), 3076–3083 (2017)
15. Google: Block Search indexing with “noindex.” (2021a). <https://developers.google.com/search/docs/advanced/crawling/block-indexing>
16. Google: Introduction to robots.txt (2021b). <https://developers.google.com/search/docs/advanced/robots/intro>
17. Google: Qualify your outbound links to Google (2021c). <https://developers.google.com/search/docs/advanced/guidelines/qualify-outbound-links>
18. Google: Robots meta tag, data-nosnippet, and X-Robots-Tag specifications (2021d). [https://developers.google.com/search/docs/advanced/robots/robots\\_meta\\_tag](https://developers.google.com/search/docs/advanced/robots/robots_meta_tag)
19. Gupta, S., Rakesh, N., Thakral, A., Chaudhary, D.K.: Search engine optimization: success factors. In: 2016 4th International Conference on Parallel, Distributed and Grid Computing, PDGC 2016, pp. 17–21 (2016). <https://doi.org/10.1109/PDGC.2016.7913146>

20. Janani., V.D., Haribaabu., V., Sajeev, R.A.: Terminating the spamming links and privacy guaranteed search logs. In: 2013 International Conference on Information Communication and Embedded Systems, ICICES 2013, pp. 394–397 (2013). <https://doi.org/10.1109/ICICES.2013.6508382>
21. Khan, M.N.A., Mahmood, A.: A distinctive approach to obtain higher page rank through search engine optimization. *Sadhana—Acad. Proc. Eng. Sci.* **43**(3) (2018). <https://doi.org/10.1007/s12046-018-0812-3>
22. Khraim, H.S.: The impact of search engine optimization on online advertisement: the case of companies using E-marketing in Jordan. *Am. J. Bus. Manage.* **4**(2), 76–84 (2015). <https://doi.org/10.11634/216796061504676>
23. Killoran, J.B.: How to use search engine optimization techniques to increase website visibility. *IEEE Trans. Prof. Commun.* **56**(1), 50–66 (2013). <https://doi.org/10.1109/TPC.2012.2237255>
24. Lali, M.I.U., Ul Mustafa, R., Saleem, K., Nawaz, M.S., Zia, T., Shahzad, B.: Finding healthcare issues with search engine queries and social network data. *Int. J. Semant. Web Inf. Syst.* **13**(1), 48–62 (2017). <https://doi.org/10.4018/IJSWIS.2017010104>
25. Malaga, R.A.: Search engine optimization—black and white hat approaches. In: *Advances in Computers*, vol. 78, no. 10, pp. 1–39. [https://doi.org/10.1016/s0065-2458\(10\)78001-3](https://doi.org/10.1016/s0065-2458(10)78001-3)
26. Manhas, J.: A study of factors affecting websites page loading speed for efficient web performance. *Int. J. Comput. Sci. Eng.* **3**, 32–35 (2013)
27. Moz: Anchor Text (2021a). <https://moz.com/learn/seo/anchor-text>
28. Moz: Canonicalization (2021b). <https://moz.com/learn/seo/canonicalization>
29. Moz: Domain Authority (2021c). <https://moz.com/learn/seo/domain-authority>
30. Moz: Page Authority (2021d). <http://moz.com/learn/seo/page-authority>
31. Mozilla: HTTP response status codes (2021a). <https://developer.mozilla.org/en-US/docs/Web/HTTP/Status>
32. Mozilla: lang—Language tag syntax (2021b). [https://developer.mozilla.org/en-US/docs/Web/HTML/Global\\_attributes/lang](https://developer.mozilla.org/en-US/docs/Web/HTML/Global_attributes/lang)
33. Noruzi, A.: A study of HTML title tag creation behavior of academic web sites. *J. Acad. Librarianship* **33**(4), 501–506 (2007). <https://doi.org/10.1016/j.acalib.2007.03.008>
34. Pažeraitė, A., Repovienė, R.: Content marketing decisions for effective internal communication. *Manage. Organ. Syst. Res.* **79**(1), 117–130 (2018). <https://doi.org/10.1515/mosr-2018-0008>
35. Ramlall, S., Sanders, D., Tewkesbury, G., Ndzi, D.: Can keyword length indicate Web Users’ readiness to purchase. *April*, 1–16
36. Rowley, J.: Understanding digital content marketing. *J. Mark. Manag.* **24**(5–6), 517–540 (2008). <https://doi.org/10.1362/026725708X325977>
37. ScreamingFrog: SEO Spider Tabs (2021). <https://www.screamingfrog.co.uk/seo-spider/user-guide/tabs/>
38. SEMrush: Subdomain vs. Subdirectory (2020). <https://www.semrush.com/blog/subdomain-vs-subdirectory/>
39. SEMrush: Site Audit Crawled Pages Report (2021a). <https://www.semrush.com/kb/543-site-audit-crawled-pages>
40. SEMrush: Site Audit Statistics (2021b). <https://pt.semrush.com/kb/544-site-audit-statistics>
41. Tsuei, H.J., Tsai, W.H., Pan, F.Te, Tzeng, G.H.: Improving search engine optimization (SEO) by using hybrid modified MCDM models. *Artif. Intell. Rev.* **53**(1) (2020). <https://doi.org/10.1007/s10462-018-9644-0>
42. Visser, E.B., Weideman, M.: An empirical study on website usability elements and how they affect search engine optimisation. *SA J. Inf. Manage.* **13**(1), 1–9 (2011). <https://doi.org/10.4102/sajim.v13i1.428>
43. Williams, A.: *SEO 2016 & Beyond*, 5th edn. CreateSpace Independent Publishing Platform (2015)
44. Yalçın, N., Köse, U.: What is search engine optimization: SEO? *Procedia—Soc. Behav. Sci.* **9**(July 2009), 487–493. <https://doi.org/10.1016/j.sbspro.2010.12.185>
45. Zhang, S., Cabage, N.: Search engine optimization: comparison of link building and social sharing. *J. Comput. Inf. Syst.* **57**(2), 148–159 (2017). <https://doi.org/10.1080/08874417.2016.1183447>

46. Ziakis, C., Vlachopoulou, M., Kyrkoudis, T., Karagkiozidou, M.: Important factors for improving Google search rank. *Future Internet* **11**(2) (2019). <https://doi.org/10.3390/fi11020032>
47. Zilincan, J.: Search engine optimization. *Future Internet* **12**(1), 506–510 (2015). <https://doi.org/10.3390/fi12010006>