UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Università degli Studi di Padova

Dipartimenti di Studi Linguistici e Letterari

Corso di Laurea Magistrale in
Lingue Moderne per la Comunicazione e la Cooperazione Internazionale
Classe LM-38

Tesi di Laurea

International SEO

Search Engine Optimization and its application to multilingual, multimarket websites

Relatore
Prof. Elia Zanon

Laureanda
Alessia Zorzi
n° matr.2088922 / LMLCC

Anno Accademico 2023 / 2024

Table of contents

Table of figures

ABSTRACT

The proliferation of the internet has dramatically reshaped the global marketplace, necessitating an innovative and responsive approach to marketing. The modern customer has become a co-creator of the product, to whom traditional promotional techniques do not appeal as they once did. Within the realm of digital marketing, Search Engine Optimization presents itself as a responsive and strategic tool to appeal to and serve online users. SEO practices aim to enhance the visibility and engagement of online businesses, so as to intercept prospects more effectively and ultimately increase conversions and revenue. This thesis explores the complex interplay between SEO strategies and the challenges inherent in catering to diverse cultural and geographically located market segments. The present work starts with a comprehensive review of existing literature on SEO fundamentals to provide a comprehensive framework of this discipline. Through the delineation of the historical development of search engines and web design principles, the study elucidates best practices and innovative strategies for optimizing content across multilingual and multimarket landscapes from both a technical and a semantic point of view. Ultimately, the thesis addresses localization and transcreation as the cornerstone of any business shift to the international market and underscores the imperative need for adaptability required by competitiveness in today's globalized digital economy. In conclusion, this research contributes to the field of digital marketing by offering a robust framework for understanding and executing SEO in a multilingual environment, highlighting the pivotal role of cultural and linguistic considerations in achieving optimal search engine visibility.

INTRODUCTION

Ever since its invention, the Internet has gained a more and more central role in our lives. Following the digital transformation of our society, driven by technological changes and innovation, today we rely ever more often on online resources to manage our daily activities. Search engines have become trustworthy and loyal friends, to whom we can turn when in search of an answer, whatever the question. The increasingly relevant role of the web in virtually all our activities goes hand in hand with the rapid emergence of countless websites and web resources, born with the intent to meet the users' needs and serve them in the best way possible. In an era where digital presence can make or break a business, search engine Optimization (SEO) has become a cornerstone of online success. As businesses increasingly rely on the Internet to reach customers, the ability to appear prominently in search engine results has never been more critical. SEO has turned into a vital marketing tactic which blends technical expertise, creative content development, and digital networking. The globalization of the Internet further complicates this landscape, introducing the need for specialized international strategies in order to cater to a diverse and geographically dispersed audience.

This thesis explores the multifaceted world of SEO, delving into the mechanics of search engines, the art of content creation, and the intricacies of optimization—both on and off the page. It also addresses the complexities of International SEO, an area of growing importance as businesses expand their reach beyond borders. The objective is to provide a comprehensive understanding of SEO, from foundational principles to advanced strategies that can be applied in a global context.

The first chapter will be dedicated to set the basis for a fundamental understanding of how search engines, particularly Google, operate. Starting from their historical evolution, we will delve into the inner workings or these systems, illustrating how understanding them is crucial for any SEO strategy, as they dictate how web pages are evaluated and ranked in search results. Search engines serve as the gateway to the vast expanse of the internet, helping users find relevant information among billions of

web pages. At the core of their functionality lies a sophisticated set of algorithms that determine which pages are most relevant to a user's query. Particular attention will be paid to Google, the dominant player in the search engine market, which uses a highly refined algorithm that incorporates machine learning, artificial intelligence, and a vast index of web content to deliver the most relevant results. This algorithm, often referred to as PageRank, and the numerous updates it has undergone over the years. These updates, such as Panda, Penguin, and Hummingbird, have significantly shaped the landscape of SEO, emphasizing the importance of quality content, backlinks, and user experience. Additionally, this chapter will cover the critical components of search engine functionality, including crawling, indexing, and ranking, ultimately determining a web page's placement On search engine results pages (SERPs). A key aspect of this discussion will be the role of user intent in shaping search results: as search engines become more and more sensitive to the context and purpose behind user queries, SEO strategies must evolve to focus on satisfying this intent. Chapter 1 will conclude with an exploration of how these elements come together to form the basis of modern SEO practices.

Building on the notions covered in chapter one, the second chapter focuses on how to create content that resonates with both search engines and users. Keyword analysis, content creation, and copywriting are components that are essential for achieving high rankings and engaging audiences. Keyword analysis is the fundamental basis upon which lies any successful SEO strategy. This process involves identifying the terms and phrases that potential customers are using in their search queries. A thorough keyword analysis allows businesses to target the right audience by optimizing their content for these specific terms. As summarized by the Internet motto "content is king", high-quality, relevant content is what search engines aim to deliver to users, making it the most important factor in SEO. This chapter will discuss the principles of effective content creation, including the importance of originality, relevance, and comprehensiveness. It will explore how to structure content to enhance readability and user engagement, which in turn influences SEO performance. Crafting compelling headlines, meta descriptions, and text combining natural-sounding writing

and strategic keyword placement are essential practices to not only attract and engage visitors, but also gain priority to the eyes of search engines.

Optimization is where the technical and creative aspects of SEO converge, and will be covered thoroughly in the third chapter, articulated into on-page, off-page, and technical optimization, each of which plays a crucial role in enhancing a website's visibility and performance in search engine rankings. On-page optimization refers to the practices used to optimize individual web pages to rank higher and earn more relevant traffic, working on title tags, meta descriptions, headers, images, and internal links, prioritizing the role of user experience (UX) in terms of fast-loading pages, intuitive navigation, and accessible content. Off-page optimization, on the other hand, focuses on actions taken outside of the page to impact rankings, namely building high-quality backlinks from authoritative sites and PR initiatives. Finally, technical SEO is the backbone of any SEO activity. It involves optimizing the infrastructure of a website to ensure that search engines can crawl and index it effectively, and that it can be viewed and experienced by the user seamlessly.

Lastly, opportunities and challenges for SEO in light of the growing globalization trends will be tackled in chapter four. As businesses expand their operations globally, they face the challenge of reaching diverse audiences with different languages, cultures, and search behaviors. This is where International SEO comes into play, providing the strategies and techniques necessary to optimize websites for a global audience, thus ensuring that businesses can effectively reach potential customers across different countries and regions. As it will be extensively discussed, international SEO is more than just translating content into different languages; it involves a comprehensive approach that considers the unique characteristics of each target market. This chapter will begin by exploring the fundamental principles of International SEO, including the importance of understanding cultural nuances, local search behaviors, and regional market trends, highlighting the need to adapt their approach to meet the expectations and preferences of users in different countries. This is also made possible by the implementation of technical tools which help search engines understand the language and geographical targeting of a web page, such as hreflang tags and geotargeting.

Content localization is ultimately the most powerful of them all, serving a central role in successfully engaging and connecting with the local audience: cultural customization of images and messages, and the creation of region-specific content that aligns with local customs and values is essential to attract visitors as well as links, which, in turn, are extremely valuable for ranking.

Starting with the foundational principles of how search engines operate, moving through the critical aspects of content creation and optimization, and culminating in the complexities of International SEO, this work aims to provide a comprehensive framework of search engine Optimization and of its application to an international context, offering valuable insights into the strategies and techniques that drive online success in both domestic and global markets. As businesses continue to expand their online presence, the importance of mastering SEO cannot be overstated. Whether targeting a local market or reaching out to audiences across the globe, understanding and implementing effective SEO strategies is essential for achieving visibility, driving traffic, and ultimately, succeeding in the competitive online environment.

CHAPTER 1: SEARCH ENGINES AND GOOGLE

1. Search engines

search engine Optimization, as suggested by the name, revolves around the optimization of websites based on search engines. Understanding how these search engines work is therefore essential to approach this discipline. Search engines have revolutionized the way we search for information, make purchases, and connect with others. Thanks to the technological and digital advantages which took place in the last quarter of the past century, virtually all the knowledge in the world is today accessible to anybody with an Internet connection and a supported device. All of this knowledge, however, would be close to useless were it not possible to store it and efficiently organize it, so that its retrieval be feasible even without knowing precisely where it is located.

We could think of the Internet as an immense library storing all sorts of materials: texts, images, videos, sounds, that are created and accessed by billions of people daily. Much like flicking through an encyclopedia to find an answer to our question takes time and effort, doing so with such volume of information would be unthinkable. The first Internet users would rely on online directories to find information on the web. These directories, organized by categories and subcategories, allowed people to navigate to their websites of interest. In those times, however, the web was a much less densely populated territory. Today, with the volume of data available, we rely on search engines to provide us with what we are looking for.

Behind every user's search query, countless operations are performed by the search engine to deliver the appropriate results. Operations which happen in less than a second and with ever-growing accuracy, meant to discover, understand, and organize the Internet. When a user requests one of these pieces of information by typing a search query in a search bar, it is retrieved from this huge collection by means of a browser, such as Google Chrome, Mozilla Firefox, and Safari[1] Web browsers, often referred to as

1Source: Encyclopaedia Britannica, 2024.

simply browsers, are software applications[2] that function as an intermediate between the user device, that is, the electronic device, computer, smartphone, from which a person decides to access the information, and a web server, a combination of hardware and software components that work together to store information (Einoryte, 2023). When some type of information is requested by the user via entering a search query onto his device, the browser forwards this request to the server, which retrieves it from where it is stored and delivers it to the browser, which ultimately displays it as a web page that is then viewed by the user.

The role of search engines in this interchange of data is to assist the user in carrying out his research. Search engines are software applications that analyze the words and phrases entered in the search bar and, according to certain criteria, suggest a list of results of web pages. These results are, from the search engine's perspective, the most likely to meet the user's needs. For all of this to happen, however, it is first of all necessary to organize the data stored in web servers. These tasks are performed by search engines through crawling, indexing, and ranking.

1.1 Crawling, indexing, and ranking

Crawling, indexing, and ranking are complementary processes that enable search engines to analyze and subsequently organize web pages. This then allows for improvement of the experience of the user by showing him the results that are most relevant to his search. Every time a new web page is created, for it to be visible to the public it first needs to be indexed: "indexation" is the procedure through which a search engine registers a piece of content present on the Internet onto its database[3]. This allows users to view it and to access it from the web browser. In order to know which pages are to be added to its database, the search engine first goes through an operation called "crawling". Web crawlers – or spiders – are robots that crawl – or search – the Internet to discover newly published and updated content and mark it as "indexable". This

2A software application is a series of programs, procedures, and routines designed to perform specific tasks and execute commands given by the user. Source: Encyclopedia Britannica, 2024.
3Source: https://moz.com/beginners-guide-to-seohttps://moz.com/beginners-guide-to-seo

discovery process progresses via links: every time a crawler finds a page, analyzes it and later adds it to the search engine index, it then proceeds to analyze the pages that are connected to that first page. Links are, therefore, valuable tools for bots and, as we will further discuss in Chapter three, for websites' owners too, since the findability of a page heavily depends on them. The last step to make a piece of Internet content visible and accessible to the public is "ranking". As the name suggests, ranking consists in ordering the previously crawled and indexed web pages when these are displayed on the SERP, the search results page. The order in which the pages are ranked depends on their relevance to the search input given by the user. The higher a page is presented on the SERP, the higher are the chances – according to the search engine – that it will answer the user's question. In that regard, to determine the value of a page, search engines make use of algorithms, formulas or sets of rules that are able to solve specific problems or perform tasks[4]. Google, for example, considers three main categories of factors:

- internal factors, that is elements which concerns the web page' structure, such as meta tags, HTML code, information architecture;
- external factors, which relate to the page's connections to other pages and, in a broader sense, to the external word; examples of external ranking factors are the quantity and quality of acquired links and page authority;
- tech factors entail all those elements that influence the pages' performance, such as speed and loading time.

The specific criteria used are multiple and not always known to the public, and although the user-centered principle at the core is common to all systems, these parameters vary across search engines. It is to be noted, however, that crawling indeed constitutes a fundamental prerequisite for a page to be indexed and therefore ranked, but not to exist. In fact, non-indexed pages are believed to represent more than 96% of the entirety of web pages present on the Internet[5]. This great portion of the web is called the "deep web", in opposition to the "surface web", which is the openly viewable section. The deep web includes all of those pages that have not been indexed by the search

4 Source: Marriam-Webster Dictionary.
5 Source: Encyclopedia Britannica, 2024

engine but are, nonetheless, accessible to users. This is the case, for example, for pages which contain subscription content, email accounts, or internal company networks. These pages would indeed not appear on the SERP but can be reached through other pages. Only a small fraction of the deep web cannot be accessed through mainstream search engines, the "dark web". Pages in the dark web can only be viewed using specific softwares, which anonymize the identity of visitors that access them, and are used by whomever wants to communicate and trade in secret. The purposes of such use can be legitimate – for instance in the case of secret services or of journalists who operate in countries where repressive governments deem them as dissidents – but the dark web often also offers a safe environment for illicit markets, for example for illegal drugs and weapons.

2. Google

As mentioned in the previous section, there are many search engines available on the market, each with its own features and its own algorithm. Among the most popular ones we can find Bing, Yahoo!, and DuckDuckGo, but the leader in the sector is indubitably Google. It being the most commonly used search engine and given the scope of this dissertation, the present work will focus on Google as it is the one around which most SEO practices and research are centered. It is to be noted, however, that other search engines too ought to be considered when implementing SEO strategies, as potential customers might prefer the alternatives.

Founded in 1998 by Larry Page and Sergey Brin (Baker, 2021), Google serves billions of searches every day. Its original algorithm revolutionized the way search engines work and, consequently, the way users experience the Web. These factors made Google by far the most used search engine in the world, with a market share consistently above 80% at least in the past 15 years[6]

6 Source: Statista, 2023.

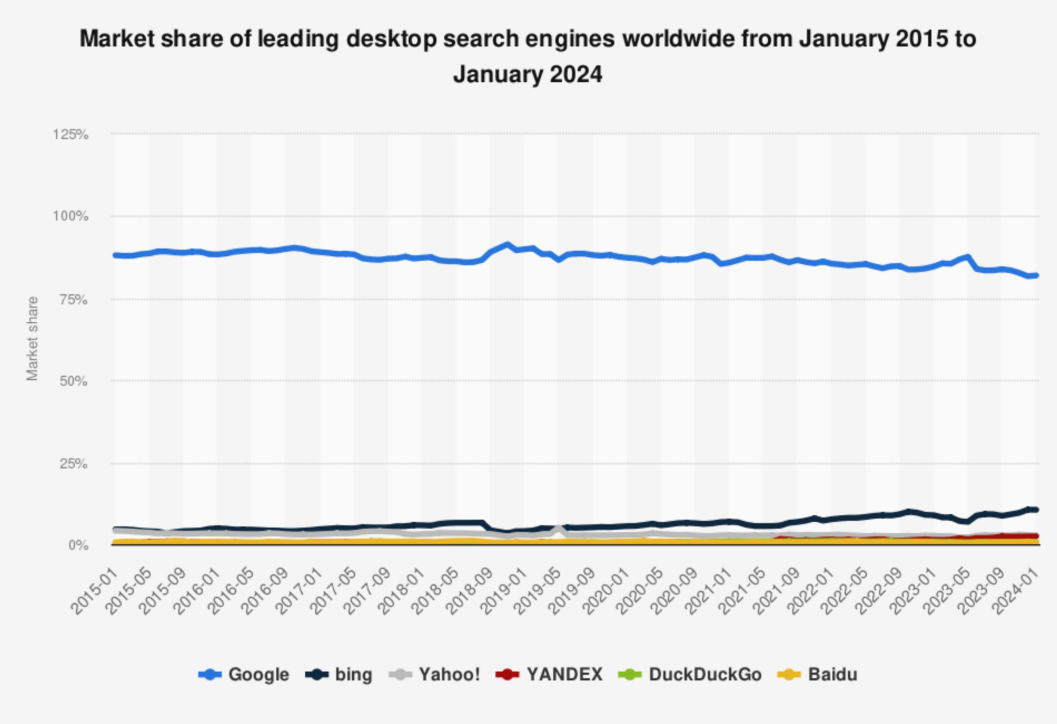


Figure 1: Market share of leading desktop search engines worldwide from January 2015 to January 2024. Source: Statista, 2024

Up to Google's launch, search engines would rank web pages based on so-called on-page elements, such as domain names and basic website structure, and on how many times the word searched by the user (in SEO, this is called a keyword) would appear on the page. This meant, for example, that repeating that same term multiple times throughout the page would allow you to rank higher in the SERP. This system was then often exploited by webmasters, who would use "keyword stuffing" – a spam tactic which consists in the overuse of keywords – to gain a better position on the SERP (Handley, 2023a).

The launch of Google came with the introduction of a new algorithm called PageRank. In comparison to its predecessors, this algorithm would rank websites considering both on-page and off-page factors – that is, factors not strictly and exclusively related to a page itself but to how the page is connected to other pages and to its context. PageRank would consider primarily the quantity and quality of pages that link to that original page (Baker, 2021). This parameter was set based on the assumption that if a web page is mentioned often by other web pages, it means that its content is

relevant and trustworthy, and especially so if these pages are important and reliable. This more sophisticated ranking method marked a step forward towards a more user-friendly outcome. It revolutionized the way ranking worked and, consequently, how SEO practitioners operated. This newly implemented ranking system too, however, soon started to be manipulated by webmasters to their own advantage. SEO practitioners began exploiting the high relevance now attributed to links by strategically working to obtain links illicitly, for example buying them or posting links in public forums or low-quality directories[7]. This would increase the PageRank score of a page and, ultimately, its ranking.

To further tackle issues like keyword stuffing, link spamming, or duplicated content – content similar or identical to content from other pages or websites (Dean, 2024b), in the years following the launch of this revolutionary ranking algorithm, Google has started implementing it with a number of new parameters and ranking systems. The algorithm is now believed to consider hundreds of factors, and has undergone various updates. These vary in extent and purpose.

2.1 Google updates

Thousands of changes to these systems are introduced yearly, but every so often a major innovation is implemented. Google does not divulge the specific criteria it uses to improve its algorithms, in order to avoid yet again any more manipulations[8]; however, the most substantial updates have been made known more or less explicitly by the company. These updates have a significant impact on the SERP and on user experience, therefore particularly relevant for SEO practitioners.

The first major change in Google's algorithm was the Florida update (2003); it targeted spam tactics popular in the previous decade, such as and primarily keyword stuffing (Montti, 2020).

Link spamming and duplicated content – other common spam techniques – have been tackled by the next update, the Jagger update (2005). This algorithm change

7 Source: Google Search Central
8 Source: Google Inside Search, 2012.

decreased the ranking score of web pages which practiced unnatural backlinking, therefore obtaining links pointing to their page upon payment and which contained copied content.

The following update also targeted unethical content-creation practices. Given Google's ultimate goal of ensuring its users be provided great user experience and valuable information, the Panda update (2011) aimed at penalizing 'content-farms', that is, pages that produce and promote low-quality content with the only intention of generating advertising revenue

The Penguin update (2012), on the other hand, focused once more on strengthening measures against manipulative link-building techniques. As we can see by now, most implementations aimed at penalizing those pages that try and exploit the loopholes in the algorithm (Handley, 2024a).

A big leap forward was marked by the introduction of artificial intelligence. With the release of the Hummingbird update (2013) and the RankBrain update (2015), Google got even closer to its visitors: first by paying closer attention to context and user location and by improving the search engine's understanding of natural language queries, that is, submitted using everyday language; this means, for example, that for the search query "chiamare studi umanistici Padova", Google will take you to the Contacts page of the School of Human Sciences of the University of Padua, even though no other word from the search query string except "Padova" is to be found on that page.

This is because the search engine considers the query as a whole rather than assessing the meaning of each individual word and is able to understand what you are searching for. This process is called semantic search. The topic of search intent will be further explained in Chapter two. Secondly, Google implemented machine learning to analyze user behavior and improve future searches. This means that the machine itself is able to apply changes to its own algorithm to better match the search intent by taking into consideration a number of metrics and constantly testing new solutions (Dean, 2024a). Finally, one last particularly relevant update was the Mobilegeddon (2015), designed in response to the increasing relevance of mobile devices. As a consequence,

Mobilegeddon rewarded mobile-friendly web pages and therefore drove SEO practitioners to take further measures to adapt their resources to this type of devices. The aspect mainly affected in terms of SEO are loading times and UX design.

These algorithms – among many others – contribute to providing users high-quality web pages by penalizing and rewarding creators based on a series of criteria, but are not definitive solutions. Google constantly works to improve its performance, implement new features to algorithms or introduce new ranking systems altogether.

2.2 Google's Search Quality Rating Guidelines

In what seems to be an ever more automated process – let us think about the growing implementation of the AI –, machines remain such and human intervention is still a valuable asset, on which Google relies to more sophistically evaluate content quality. Besides programmers in charge of writing and implementing new codes into Google's search engine, an important figure within the process of constantly improving the algorithm is that of Search Quality Rater. Raters are search quality evaluators who, from different locations around the world, evaluate the quality of search results, providing Google insightful feedback on its search engine performance. Based on these cues, Google is able to evaluate changes in the algorithm and to adjust it to better fit the users' needs.

To ensure consistency in rating, since 2005 Google has been providing a set of criteria in the Google's Search Quality Rating Guidelines (Crowe, 2021). Search Quality Raters use these instructions to "lend" their inherently human skills – unmatchable by machines – while still referring to fixed measurements. Indeed, each rater has the task of judging search results not on the basis of his personal opinions and preferences, but rather on these specific guidelines in relation to cultural standards of the specific geographic region they represent, so as to reflect the experience of the people in their rating locale[9]. By collecting content evaluation from various locations across the globe, Google is able to implement changes accordingly, both internationally and locally.

9Source: Google's Search Quality Rating Guidelines, 2024.

These guidelines are publicly available on the web and are primarily addressed to rater. Nonetheless, they provide interesting insight on Google's internal workings, especially those involved in creating content that is rated as valuable by the search engine – namely, SEO practitioners and content creators. Although these parameters do not technically classify as ranking factors, they still indirectly influence page ranking, since raters' evaluations are considered by Google to refine its algorithms.

2.3 The importance of trust: E-E-A-T

Among the many indicators included in Google's guide for raters, E-E-A-T is one of the most significant ones, which well exemplifies Google's commitment to user experience. Standing for "Experience, Expertise, Authoritativeness, and Trustworthiness", E-E-A-T guidelines are meant to reward web pages which offer authoritative and helpful content, written by experts on the topic who can draw from first-hand experience.

Introduced originally in 2014 as E-A-T ("Expertise, Authoritativeness, and Trustworthiness"), in 2022 an additional "E" (expertise) was added as fourth pillar of this model to further strengthen Google's position in favor of trustworthy content and against deceiving and misleading information (Southern, 2024). At a first glance, this addition might seem unnecessary, given that "experience" and "expertise" are oftentimes concepts that go together; in fact, it is not so. According to Google's own definition, "expertise" is "the extent to which the content creator has the necessary knowledge or skill for the topic", mainly acquire through education"; "experience", on the other hand, is "the extent to which the content creator has the necessary first-hand or life experience for the topic". While these notions certainly overlap in certain points and in certain situations – through experience, people can gain expertise, and expertise can imply experience – the relationship between the two is more complementary than exclusive. The second factor taken into account by this model is "authoritativeness". Authoritativeness refers to the overall reputation of the web page in question within its

field. An authoritative website would be recognized as a reliable source of information by experts of the industry.

"Experience", "expertise", and "authoritativeness" combined together contribute to the overall legitimacy of the web page, which, ultimately, is crowned by "trustworthiness". Trustworthiness is the most important part of Google's E-E-A-T, since a lack of trust would vanish all efforts in all the other areas. Trustworthiness applies to the authenticity of the content, but also to more practical matters, such as the security of an e-commerce website's check-out page: regardless of the experience and expertise of the authors or the authority of the page, Google does not want to deliver its users misleading information (Handley, 2024b).

2.4 Your Money or Your Life

Despite the centrality of trust in Google's mission to constantly improve itself, it is important to note that these evaluation criteria are not fixed; they vary according to the context. The amount of trust which the search engine deems as necessary to qualify a page as "high-quality" is not objective, but rather depends on the type and topics covered. This justifies the involvement of human rater to better grasp the subtleties of human-generated content. Experience, for example, gains importance in a product review that has been written by somebody who really purchased the product and that aims at honestly advising potential buyers, allowing them to make informed purchasing decisions. Expertise, on the other hand, might be more beneficial for a tech professional when researching the specifics of a product. Or again, authoritativeness is particularly relevant, for instance, for an educational website, curated by an academically qualified professor who might want to support her arguments with data from external publications. On the contrary, an entertainment website requires a much lower degree of trustworthiness, as its aim is to amuse its audience.

There are instances, however, in which trustworthiness rather literally plays a vital role. This is the case for searches related to health, finance and safety. These topics have been labeled by Google as Your Money or Your Life (YMYL). YMYL

parameters, also integrated to the Quality Rating Guidelines in 2014, concern web pages or content that can impact a person's physical, emotional, or financial well-being, and therefore require the highest levels of E-E-A-T (Handley, 2024b). When a user is searching for medical help, financial advice or a domestic violence helpline, a lot is at stake. If the search engine delivers erroneous or unclear information, this might cause serious harm; for this reason, it is of the utmost importance to ensure its accuracy and reliability.

Within the above-mentioned domains, various types of content might qualify ad YMYL topics and to various extents. It is the rater's task to determine that and provide Google with valuable feedback for the refinement of its algorithm. The company, on its part, has delineated more precise guidelines and frameworks and attributes more weight to E-E-A-T factors for queries related to YMYL topics, so as to further reward (or more heavily penalize) more attentive (or ill-intentioned) content creators (Haynes, 2022).

3. Google SERP

As explained in section 1.1, crawling, indexing, and ranking are fundamental processes finalized at presenting the users with the most appropriate results. These results are displayed on a page called the search engine Results Page (SERP). The SERP is a dynamic interface in a web browser that changes according to the search query submitted by the user to show the best solutions ordered by usefulness and relevance (Byers, 2023). Each individual result is called a snippet, and it can assume various forms, depending on the type of result. The SERP is populated with a multitude of different content formats. These can be divided into three categories: organic results; paid results; and SERP features.

Organic results

Organic results are the regular and most basic type of SERP results. They are displayed on the results page according to Google's ranking and constitute the focal point of all search engine Optimization practices. As will be discussed in detail in the

following chapter, SEO operates to boost organic results performance or web pages on the SERP by means of various strategies regarding multiple aspects of a web page (Frankel, 2023).

Organic results appear on the SERP featuring three fundamental elements:

- a URL (Uniform Resource Locator), which is the link to the page and which has the function of locating the resource on the Internet, so that the browser should retrieve information[10];
- a meta title, that is a clickable title that allows visitors to access the web page; title tags reflect the content of the page and this provides insight to both the user and the search engine, therefore concurring/contribution to the page's ranking position (Lyons, 2022);
- a meta description, which appears directly below the title and shortly describes the content of the web page; this element is not taken into consideration for ranking but can nonetheless indirectly affect the page's rank score by attracting visitors, thus boosting traffic (Lahey, 2024).

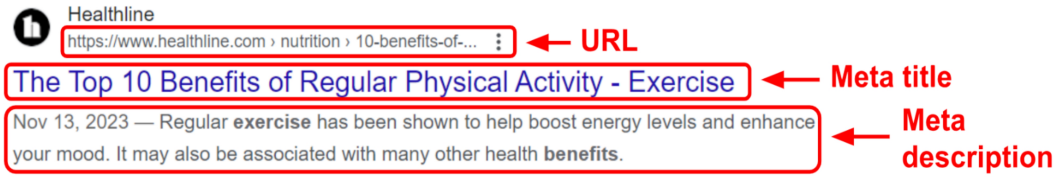


Figure 2: Example of a snippet for an organic result, with all of its components.

These elements constitute the most basic form of snippet.

Paid results

Paid results – or sponsored – also appear in the SERP featuring an URL, a title and a description, appearing in all similar to regular organic results snippets with few but substantial differences. The discriminant between the two is the transactional nature of paid results: paid results are essentially advertisements, for which a website owner has paid a fee to the platform in exchange for its website to be shown (Byers, 2023). More specifically, sponsored results are one type of search pay-per-click ads, which on Google are managed through the platform Google Ads. Paid results are therefore not

10 Source: Moz.com

concerned with page ranking – which works on what we could call a "meritocratic" basis – but are nonetheless involved in another type of competition with other websites. In what can be considered an auction, advertisers bid on a specific keyword for which they want to advertise their page and compete between one another to see it displayed in the SERP. Depending on the bid but also on such as the ad quality, ad relevance, and other factors related to the user's attributes, the ad can appear in sponsored results section, at the top of the SERP before organic results. This process is repeated every time a user performs a research and, since context and competitors might change from time to time, the outcome can change too [11].

Another element that distinguishes the two typologies is "richness". Unlike paid results, an organic result can become a so-called rich result, or rich snippet, a special type of snippet that, in addition to the regular features, can also include extra elements to further attract and better serve the user (Walsh, 2023). A rich snippet could, for example, feature ratings, images, and prices, to differentiate a page from other snippets and make it more appealing to users.

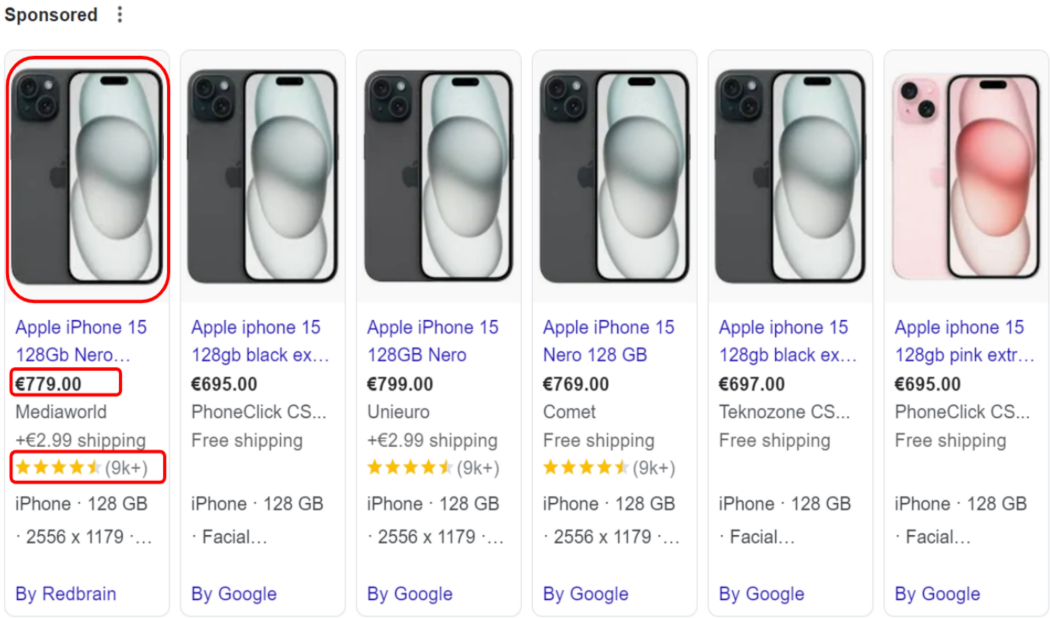


Figure 3: Example of a rich snippet for an paid result, featuring images, price, and ratings, as well as a number of additional information.

11 Source: Google Ads Help in Google Support,

In technical terms, a rich snippet is part of the third and last category of search result available on Google's SERP, SERP features.

3.1 SERP features

The last type of search results which can appear on the SERP are SERP features. SERP features are a very broad category which include all those kinds of search results that do not classify as basic organic results, therefore including rich snippets and paid results. SERP features are extremely user-oriented, as their function is to make a web page even more attractive and to better serve the user by highlighting the information that the search engine considers the most useful (Sheley, 2023). If a user is looking for a Chinese restaurant, for instance, Google might present him with a map of the area showing such businesses near him. This example of SERP feature is called a local pack, and is one of the many implemented by Google. There are over twenty different types of SERP features available, of which it is with illustrating the most popular ones (Boroda, 2022).

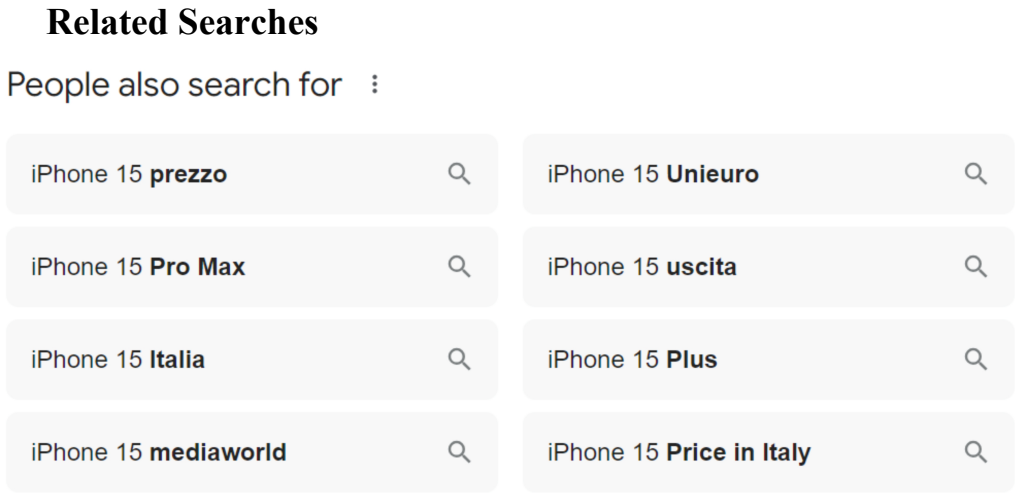


Figure 4: Example of related searches suggested by Google based on the search query "iPhone 15".

The Related Searches section is one of the most common SERP features; it is displayed at the bottom of the results page and offers a number of search suggestions

based on the search query. By clicking on one of the related searches, the user is redirected to a new SERP for the new query.

Figure 5: Example of related questions suggested by Google based on the search query "iPhone 15".

The People Also Ask feature is a box that lists a number of questions, automatically generated by Google, related to the original research and that can be expanded to directly read the answer.

Featured Snippet

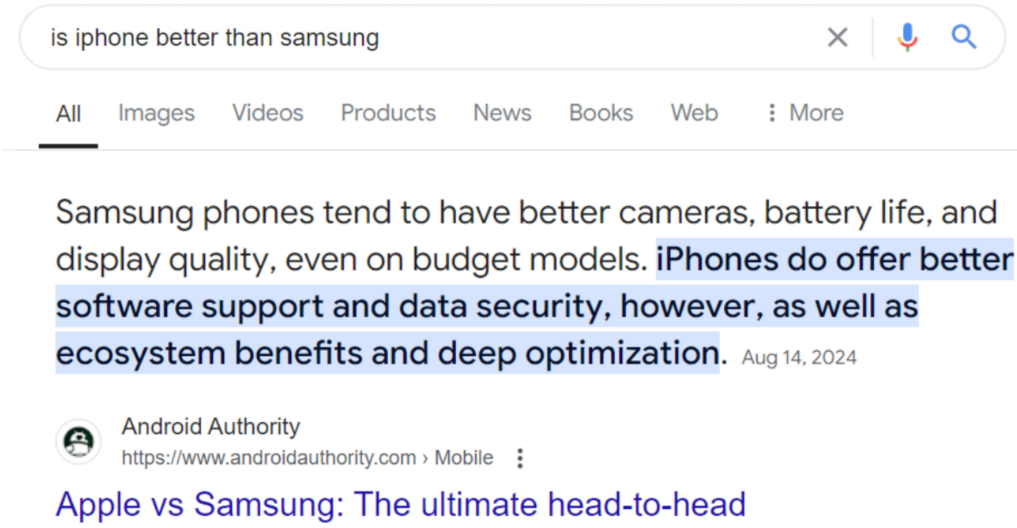


Figure 6: Example of featured snippet suggested by Google, featuring the website https://www.androidauthority.com

A Featured Snippet is a large, highlighted portion of text taken from a web page that Google has selected to provide a concise, direct answer to the query. It often appears at the very top of the organic SERP, in what is referred to as "position zero" in the rankings.

Direct Answers

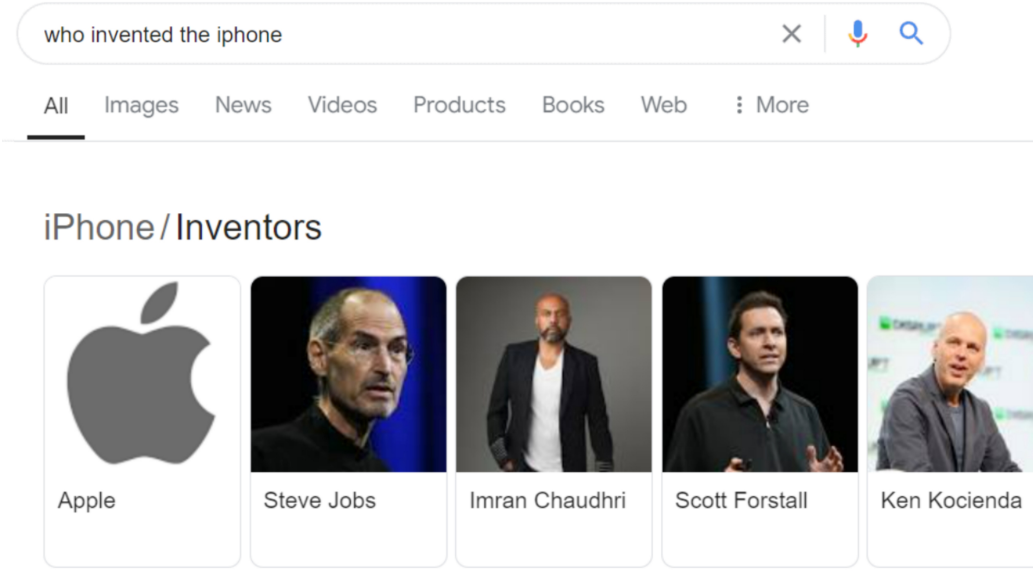


Figure 7: Example of direct answer provided by Google.

The Direct Answers section presents, as suggested, a direct answer to the search query. The query in question has to be relatively simple or notional for Google to be able to answer it, for example, “who invented the iPhone".

Sitelinks

Sitelinks are additional links which appear under a regular snippet and allow users to quickly navigate to a particular part of a website. For example, if the user searches the local cinema’ website, Google might predict that the Tickets page and the Film Schedule page will fulfill the user’s need and thus present them as sitelinks. This type of SERP feature is possible when the search engine is able to understand the website structure particularly well.

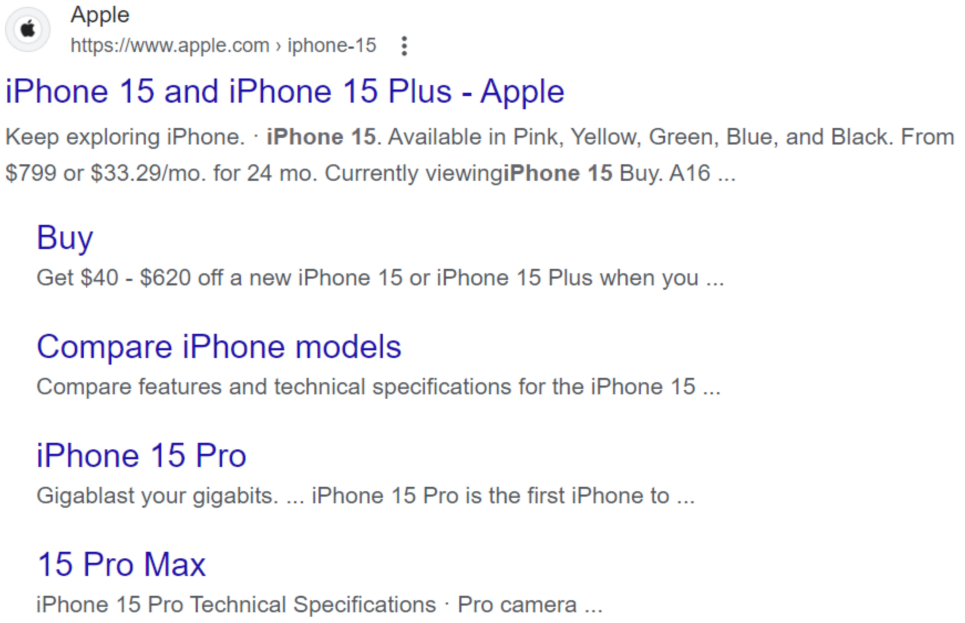


Figure 8: Example of links to the main sections of the website https://www.apple.com related to the search query.

Knowledge Panel

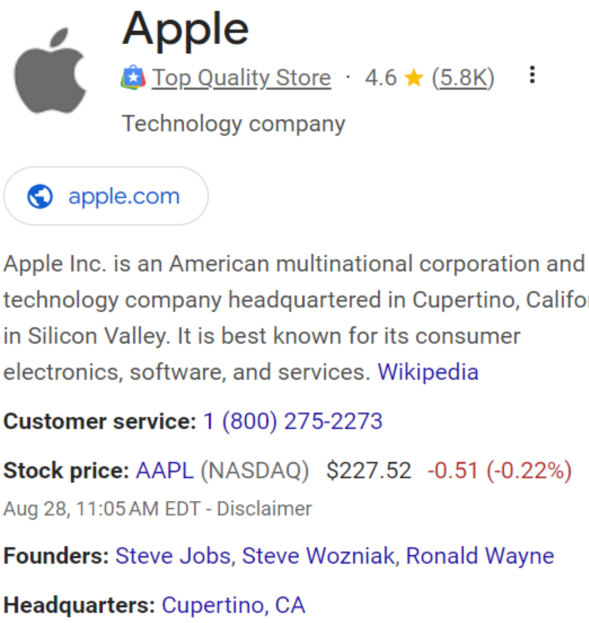


Figure 9: Example of knowledge panel for the search query "What is Apple", featuring information from Wikipedia.

A Knowledge Panel is a summary or brief overview on the researched topic, created by Google by drawing from its Knowledge Graph, a database of factual

information coming from various sources. The Panel appears on the right-hand side of the SERP when the user searches sports scores, stock prices, or weather forecasts.

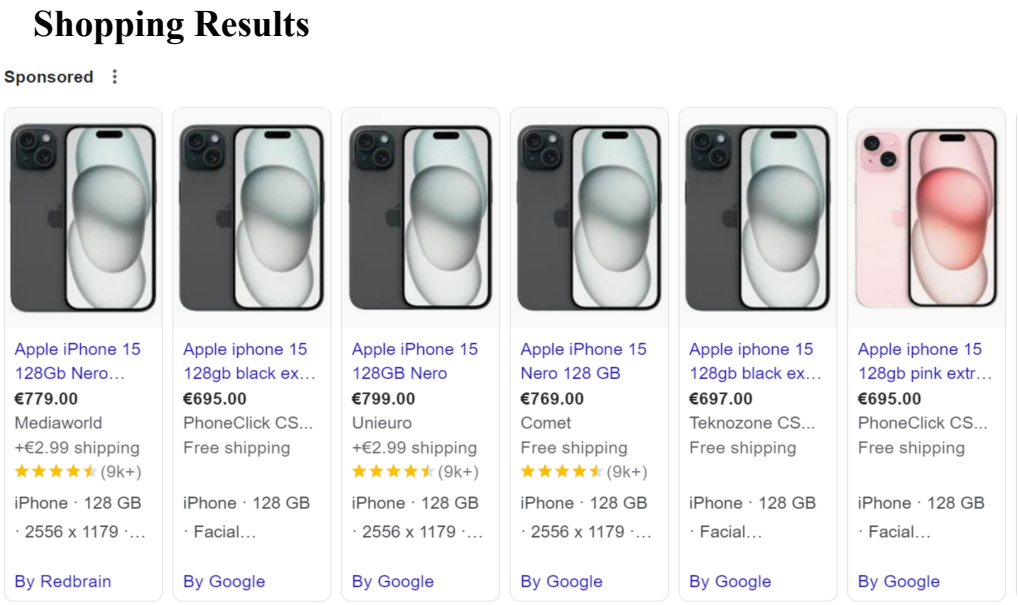


Figure 10: Example of shopping results that are, in fact, ads by various advertisers

Shopping Results appear on the SERP when Google understands that the user's intent is purchasing. The searched products are then displayed as a carousel of images which directly link to product pages on various e-commerce sites. Shopping Results can be both organic or paid, in which case they are called Google Shopping Ads and present an "Sponsored" disclaimer.

3.2 SERP features and SEO

SERP features can heavily affect how a results page is presented. The potential features are numerous and can change greatly from query to query. This introduces an additional layer of complexity to the task of SEO practitioners. As will be discussed in the following chapter, organic results are at the core of search engine Optimization strategies. SEO practitioners work to achieve the highest possible position on the SERP, and SERP features act as a double-edged sword.

One crucial factor is size: SERP features are visually bigger than regular snippets and use more space. From a SEO perspective, this can represent both an

opportunity and a threat: managing to rank organically in some SERP features means attracting more visitors, which means gaining a competitive advantage over the competitors; conversely, a competitor's SERP feature would imply that visitors would have to move further down on the results page – which they are less likely to do. At the same time, however, some SERP features – such as Direct Answers, Knowledge Panels, and Featured Snippets – can generally undermine traffic-generation initiatives: indeed, these types of results promptly answer the submitted question directly on the SERP, thus eliminating the need for users to click on standard organic results to find the solution. These instances are referred to as "zero-click searches", and account for roughly ¼ of searches[12] (Tober, 2022).

A way to increase the likelihood of being featured in a rich result is to implement schema markup. Schema markup, also known as structured data, is a type of code on a web page that enables search engines to more accurately interpret and categorize the content on your site. When it indexes a web page, Google tries to understand its content to offer it to users entering relevant queries. To do so, the search engine analyzes it and follows on-page links, and can then use this information to display the content in a rich format. Website owners can facilitate Google's job by providing explicit signals about that certain page in the HTML code of a page.

Structured data fundamentally works as labels on different pieces of information present on the page; for a cooking website, for example, it could be strategic to implement structured data on a recipe page labeling, for instance, photos of the dish, cooking time, ingredients, nutritional values. This not only helps Google understand, but also makes it easier for the algorithm to include this information in an optimized feature in search results, like rich snippets or other interactive elements, improving your page's visibility. Implementing this kind of data scheme is beneficial for SEO for a number of reasons. Although it does not directly affect ranking, it does indeed promote more accurate targeting by increasing the accuracy of query-content matching. This increase in traffic can, in turn, increase conversions, since the users who are shown this

12 This data refers to desktop search, that is search queries submitted via a browser application on a computer. The ratio for mobile devices slightly exceeds 17% (Tober, 2022).

content will be more likely to interact with it. Additionally, rich snippets can have a significant impact on organic traffic, because they generate more engagement, are more alluring and practical, and occupy more real estate on the SERP which, as we have seen, can give the website competitive advantage and differentiate it from its competitors (Harsel, 2023b).

Schema markup uses a standardized vocabulary to mark elements on the page in a way that search engines can easily process; developed by major search engines like Google, Bing, and Yahoo, this vocabulary is collected by the website schema.org, which offers a broad range of schema types. There are many types of schema markup, depending on the content of the page: some of the most common are product markup, generally used to label information about product goods such as rating, pricing, availability; organization markup, which can be pursued by companies and organizations to encourage Google to show rich snippets such as knowledge panels; local business markup, which provides useful information about a local business such as location, contacts, opening hours, ratings.

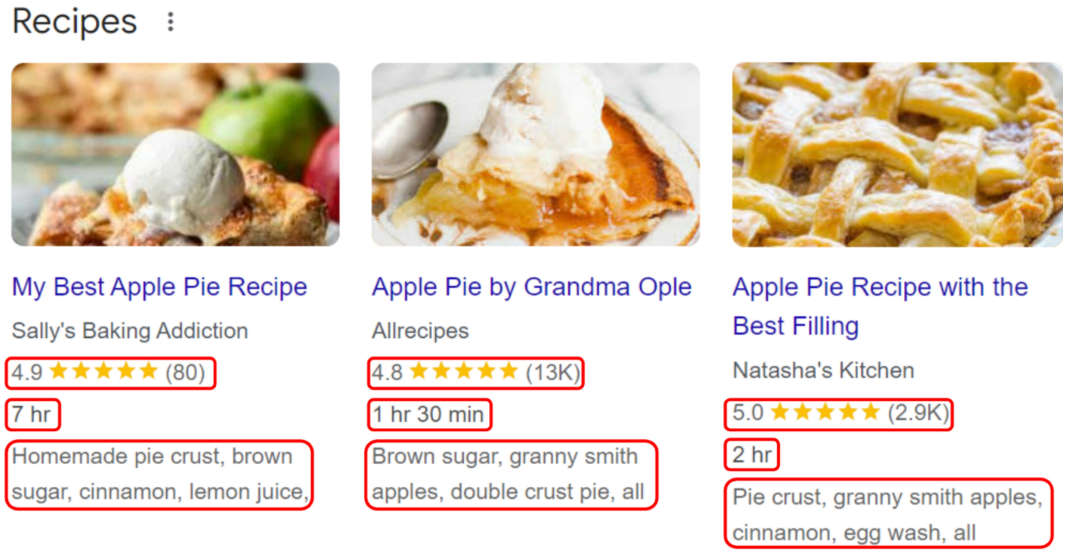


Figure 11: Example of rich result for the search query "apple pie", showing recipe markups with useful information such as rating, cooking time, and ingredients.

On Google, schema markup can be implemented using Structured Data Markup Helper, a tool provided by this search engine to easily label all individual elements on the page and automatically create the coding necessary to be added into the HTML code. Furthermore, Google also provides a testing tool called Rich Results Test to check

functionality, and ensure eligibility for this kind of snippets. It is important to note, however, that including schema markup on a page does not guarantee that it will be displayed in a rich format: sometimes, even with correctly implemented schema, Google may still present it as a standard link instead of detailed search results. Ultimately, schema markup is continuously evolving, and keeping up-to-date with the latest developments is essential. As new schema types are introduced, opportunities arise for different kinds of content to be better represented in search results (Kalyadina, 2020).

CHAPTER 2: FROM CREATION TO OPTIMIZATION

1. Search Engine Optimization

search engine Optimization is a discipline born in the second half of the 1990s and whose prominence grew following the increasing popularity of the Internet (Baker, 2021). As people started relying more and more on online resources to go about their everyday lives, researching any kind of information to make the most out of this novelty, business owners worldwide too started seizing the opportunities that the newly-born web offered to maximize their gains. Today, SEO is one the most powerful digital marketing tools: it is embraced by multinationals as well as local stores, across all kinds of industries and settings, both commercial – in which SEO is used to increment their profits, for example – as well as non-commercial – as might be a cooking blog or an institutional organization, for instance.

In such a crowded space as today's web, emerging among all options is a matter of vital importance: a web page which gets very little visibility is deemed to perish under the competition of other pages. Appearing at the top of the search results page not only increases exposure and, therefore, brand awareness – one of the fundamental goals of any marketing activity –, it also leads to increased volume of organic traffic (Frankel, 2023). Research shows that the top three organic results on Google's SERP account for more than 50% of the total clicks by users, with the first organic result attracting 27.6% of all clicks – ten times more than the result scoring tenth (Dean, 2023a).

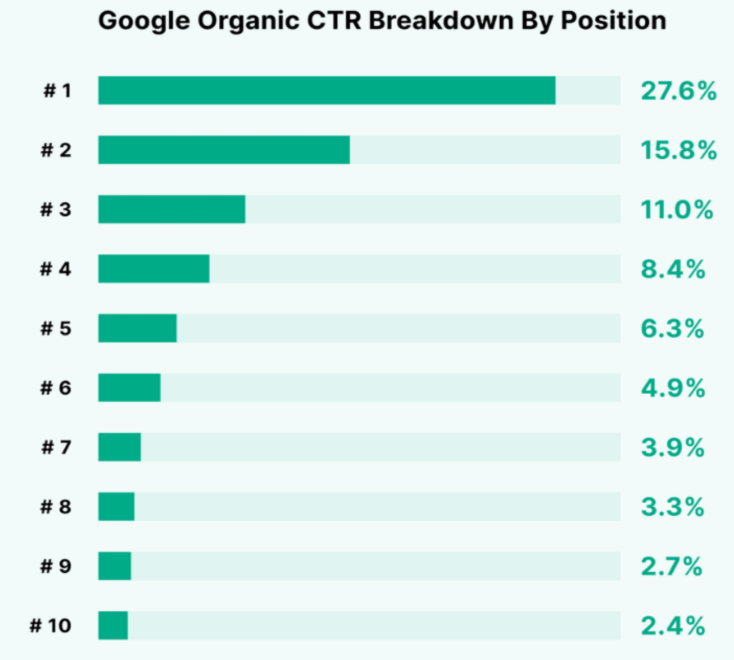


Figure 12: Click-through-rate for organic search results, divided by position.

These numbers indisputably illustrate the significant value of prime positioning. Indeed, this challenge can be tackled by implementing techniques and strategies aimed at enhancing and strengthening one's presence on the web. This concept is referred to as organic visibility, and expresses how likely it is that a user will see and visit a particular page when surfing the Internet (Cojocariu, 2023). There are hundreds of factors upon which Google relies to determine ranking – thus visibility – and working on the most relevant ones is the goal of SEO. The implementation of SEO practices enables a business to reach larger audiences by intercepting new potential leads and by serving them with the most appropriate content; this not only contributes to increasing brand awareness and promoting conversions, but it also triggers a mechanism for which better ranking leads to more traffic, which, in turn, lead to better ranking, therefore, again, to more traffic. For this reason, SEO ought to be considered as an integral part of any business strategy, which can bring a number of long-term benefits.

SEO encompasses a variety of practices surrounding every aspect of the user experience on the web browser and, later on the web page. The first, fundamental step is the keyword research process. Keyword research is the starting point of any SEO

strategy: it concerns deciding on which researched terms to concentrate one's efforts to achieve a higher-ranking position on the SERP. Keywords function as touch points between people and what they are looking for and identifying them and the purpose behind them is instrumental to provide users with content which satisfies their needs. Content creation means designing content around those targeted keywords and it is the second crucial step of any SEO implementation. Correctly interpreting the users' intent behind their queries is decisive to deliver the most relevant and useful information. Lastly, all formal aspects of a web page are adjusted to enable a more intuitive and efficient navigation. This optimization occurs from both a user experience and a search engine perspective and concerns all those elements of a page which we could consider, in a broader sense, as structural, semantically as well as technically.

Through the combination of these techniques, SEO can exploit the potential of the Internet to enhance websites' performance and allow businesses to serve a wider customer base.

1.1 SEO, SEA, and SEM

One important characteristic of SEO, which justifies all of these practices, is the fact that its goal is pursued using exclusively organic results. This means, in other words, that SEO only concerns search results that appear on the SERP because of their relevance to the search terms entered by users, as determined by the search engine's algorithms. This distinction marks an important difference between SEO and other similar online marketing strategies, namely search engine Advertising (SEA). Although both can and are employed to increase brand awareness, traffic, and conversions, they vary in cost, techniques, and effect.

SEA is a form of online marketing which is fundamentally based on paid advertising, often referred to as Pay-Per-Click (PPC), as the advertiser pays for the insertions every time a user clicks on. Paid results can appear on the SERP as a text-only Search Ad but also, in the case of products, as Shopping ads[13], and, similarly to

13 Additionally to Search ads and Shopping ads, PPC ads can also appear as other formats, such as Display ads or Video ads; these ads, however, are to be found on web pages and not on the SERP

SEO, they can also include keywords so as to make appear for relevant search queries. As mentioned in the previous chapter, however, ads compete for ranking in an auction-based mechanism which is separate from organic ranking system, and which ultimately heavily relies on economic resources: paid search campaigns can, indeed, be significantly expensive (Jones, 2024).

Alongside SERP exposure mechanisms, costs represent another important difference between SEA and SEO, which, despite still requiring some initial investment, is less dependent on financial funding and, if well designed, can lead to long-term traffic growth (Handley, 2024c). This makes SEO a generally more cost-effective technique, which can generate larger returns over time without needing continuous spending.

Another distinguishing factor is time: whereas SEA's fruits can be reaped virtually immediately after its implementation, SEO effects often require a longer time – from several weeks to months – before it can be appreciated. This is because search engines heavily rely on factors such as trustworthiness and authoritativeness as signs of relevance, which cannot be built overnight but rather take time to be achieved. Results, however, tend to be more stable over time if compared to volatile paid ads, which stop being shown – and thus generating traffic – the moment the advertiser stops funding them (Vora, 2023).

Lastly, research shows that organic results tend to be perceived as more trustworthy than ads. Indeed, users understand that non-sponsored results gain their position on the SERP "meritocratically", which makes them more reliable and truthful in their eyes. Conversely, only a small percentage of the total searches results in users clicking on paid advertisements: despite their prime position on the SERP, these appear less credible to users, who know that such content is shown to them with the clear intent of promoting sales, and are therefore more skeptical about their usefulness *(ibidem)*.

SEO and SEA employ different techniques and have different implications, yet these differences do not mean that either tactic is better than the other: they both share the same purpose and can benefit a business from different perspectives. They can therefore be thought of as complementary rather than mutually exclusive. SEO and SEA

(Wilkinson, 2023).

are both effective digital marketing tools, which can coexist in what is called SEM, search engine Marketing. SEM works both organic and paid results, encompassing both approaches. This implementation allows a website to appear in the paid as well as the organic sections of the SERP, and gain advantage over its competitors.

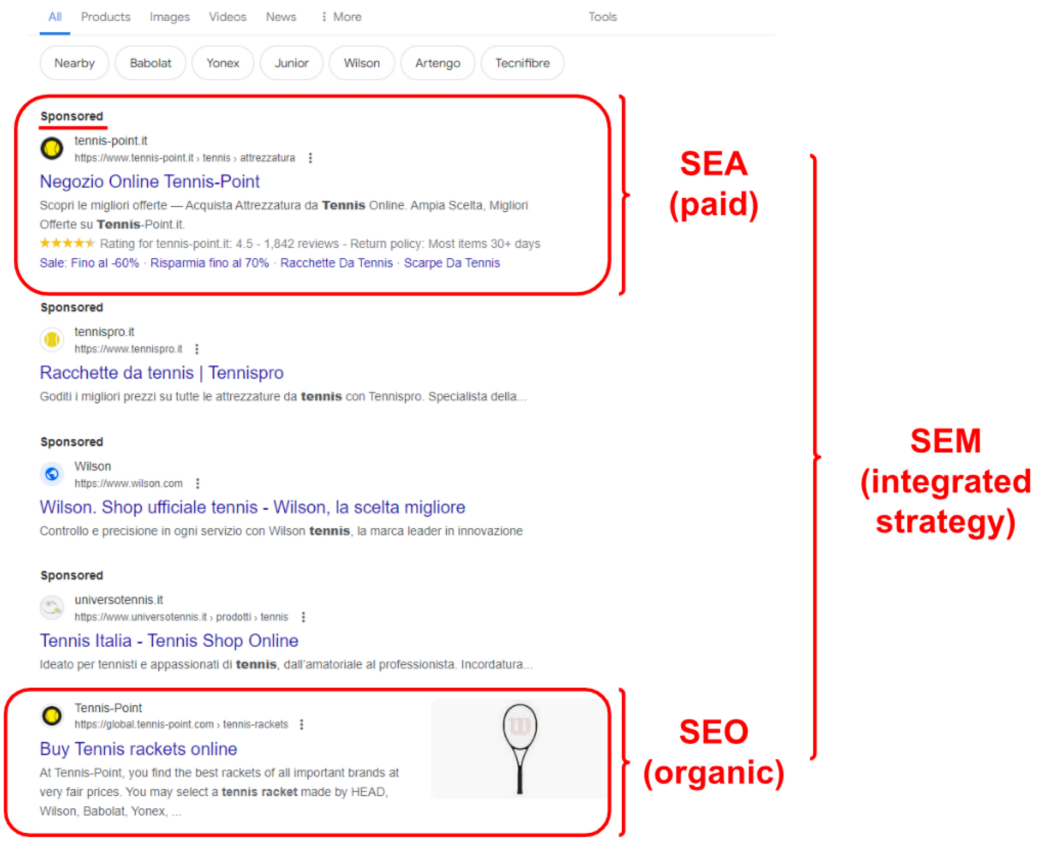


Figure 13: Organic and paid search results on the SERP for the same website.

This is firstly because, by integrating both techniques, brands can occupy more space on the SERP and, coincidentally, leaving less room to other pages' results. This increased exposure positively affects brand awareness and, because the business's prompt sales promotions are supported by its genuine helping intentions, users will be more likely to regard the results as reliable sources of information (Handley, 2024c). Additionally, these two marketing tools can be used simultaneously to achieve different ends: SEO is more suited to deliver informational content and to build reputation, since it is designed to last longer and requires longer times to generate returns; SEA, on the other hand, is primarily used to run promotional campaigns and is meant to produce

quicker results (Ali, 2023). Ideally, paid advertisement will generate traffic in the short-term, also benefiting organic ranking, and organic ranking will contribute to strengthening the image of the business in the long term, even after paid campaigns are removed.

2. Keyword Research

Keyword research is the cornerstone of any successful search engine Optimization activity. It represents the first block upon which to build one's strategy and it is instrumental to all subsequent steps. In SEO, keywords are the terms a business aims to rank for on the SERP. If a page targets the keyword "tennis rackets", for example, it means that it wants to appear in the search results of users who include those words in their search query (Handley, 2024d). The keyword research process consists in finding and selecting the terms for which one wants to rank. This selection depends, first and foremost, on what kind of products or services a certain business offers, on the users' needs in relation to these products, and on which words they use to search for it (Zerkalenkov, 2024). Aligning these three elements – content offered, content needed, and keywords used – is the key to successfully rank organically.

Before choosing which keywords to target, it is first of all necessary to analyze the web page of the business itself and identify which keywords might be relevant. These keywords are called "seed keywords" and are the starting point upon which to base the analysis (Handley, 2024e). For a sports equipment retailer that wants to optimize his tennis gear section, some examples of seed keywords might be "tennis clothing", "tennis shoes", and "tennis bag". Then, from these terms, the list might be broadened to include more candidates such as "tennis dress" and "tennis polo", "hardcourt tennis shoes" and "clay court tennis shoes", and "duffle bag" and "racket backpack" respectively. These keywords are called semantic keywords, because they are semantically related to those seed keywords and, consequently, to the topic of

tennis[14] (Handley, 2024f). To expand keywords lists, one can turn to brainstorming and creatively and spontaneously think of similar and related words; this method, however, is rather time-consuming, and significantly less effective in detecting potential keyword opportunities. Alternatively, search engines themselves can support this process by showing search suggestions on the search bar or in SERP sections such as People Also Ask or Related Searches.

A good keyword research tactic is observing what others do: this process is called competitive analysis and consists in examining similar brands within one's industry to gain insight into their offerings and strategies. This technique is broadly used in marketing research to optimize all the various aspects of its business strategy – such as, products, pricing, promotional initiatives, positioning; when applied to SEO, competitive analysis involves studying the competition on the level of keyword optimization, but also content creation and link building, as it will be discussed in the following sections (Lyons, 2023a). More specifically, competitive analysis focused on keyword use is called keyword gap analysis.

Performing a keyword gap analysis means comparing websites' ranking to detect for what keywords competitors rank, so to identify one's strengths and weaknesses and uncover possible gaps. This is because, if others rank for a certain term, it means it has potential and that it is worth targeting and including on one's keyword strategy (Beka, 2024). If a sports blog's owner, for example, discovers that all his main competitors are ranking for the keyword "women's soccer" but he is not, then it might be a good idea targeting those terms too. An important preliminary step, however, is to correctly identify one's competitors: these are similar businesses that operate in the same market and target the same segments, either offering the same product – in which case we talk about direct competition – or a different product which fulfills the same need – this is called indirect competition. This means that Harley-Davidson does not stand as a competitor for a second-hand scooter reseller, whereas an electric bikes store might well do.

14 Plural forms and variants which include stop-words such as articles and prepositions count as different keywords, therefore ought to be considered separately (Lahey, 2021).

This contrastive approach can be implemented using a number of dedicated tools, such as Semrush's Keyword Research Tools, SE Ranking's Keyword Suggestion Tool and ahrefs' Keyword Explorer, to name some. These tools are not only a source of inspiration useful to find new keywords, but also provide insight into their actual potential. Keyword analysis makes use of a number of metrics to determine the feasibility of targeting a certain keyword. If the term "sports shoes", for example, is already being targeted by very authoritative websites, ranking for it is going to be more difficult for a smaller business, thus less attractive from a SEO perspective. The metrics calculated by keyword analysis tools are various and can change from one tool to the other, but the most useful to take into account when selecting strategic keywords are search volume, keyword difficulty, and cost-per-click (Handley, 2024h). Search volume indicates the average number of searches performed for a certain keyword each month. The higher the search volume, the more often the term is included in search queries by users. Consequently, the more traffic is driven to the page thanks to the keyword in question, which should therefore be prioritized and included in the strategy (Handley, 2024g). The search volume for "sports shoes" is most probably higher than the search volume for "yellow tennis wristband", for example. High volume keywords, however, are often more difficult to pursue, for the simple reason that frequently searched terms tend to be more competitive: high volume is typically linked to high difficulty.

Keyword difficulty is a value – expressed as a percentage – that measures how difficult it is to rank organically for a certain keyword, more specifically within the first ten search results on the SERP. The higher the score, the higher the difficulty to reach the top of the results page (Collins, 2024). In this case, the keyword difficulty for "cotton tennis wristband", might be rather low, as there are probably not many pages that cover that topic, nor businesses who market that product.

Lastly, cost-per-click (CPC) measures how much an advertiser pays the advertising platform, for example, Google Ads, for a keyword targeted in an ad for each click it receives; the more popular a keyword is, the higher the competition to rank for it, thus the higher the cost (Lyons, 2024). CPC is particularly relevant for SEA, although it can also provide useful information for SEO, as it can be used to evaluate

the productivity of ads containing a specific keyword and, consequently, of the keyword itself.

Keyword analysis ultimately consists in balancing the costs and efforts necessary to rank for a certain keyword and sets the basis for all the other SEO activities, standing as arguably the most important decision making process. Selecting the right keywords means thoroughly understanding one's own audience and is essential to all further decision making processes. Indeed, creating content based on terms in line with a business offer is useless if those terms are not matched with what potential customers need and, therefore, search for. For this reason, a fundamental element of any keyword analysis consists in investigating beyond the words that appear in the users' queries and uncover the underlying purpose of their research and use this insight to offer them a page thought which they can navigate intuitively.

2.1 Search intent

Search, or user, intent is the intention behind every user's research, and it is one of the most important concepts of SEO. Search intent goes beyond what words are typed into the search bar: it means understanding the true need of what could be a potential customer. We can distinguish between four main types of search intent: navigational, informational, commercial, and transactional (Pavlik, 2022a).

Navigational intent concerns users who are searching for a specific page. They have a clear idea in mind of what they are looking for and, therefore, expect to be presented with that one page they need. "youtube login"is, for instance, an example of a search query driven by a navigational intent: the user knows exactly what page that they want – in this case, the login page for the video-streaming platform Youtube; as in this case, navigational queries often include a brand name**.**

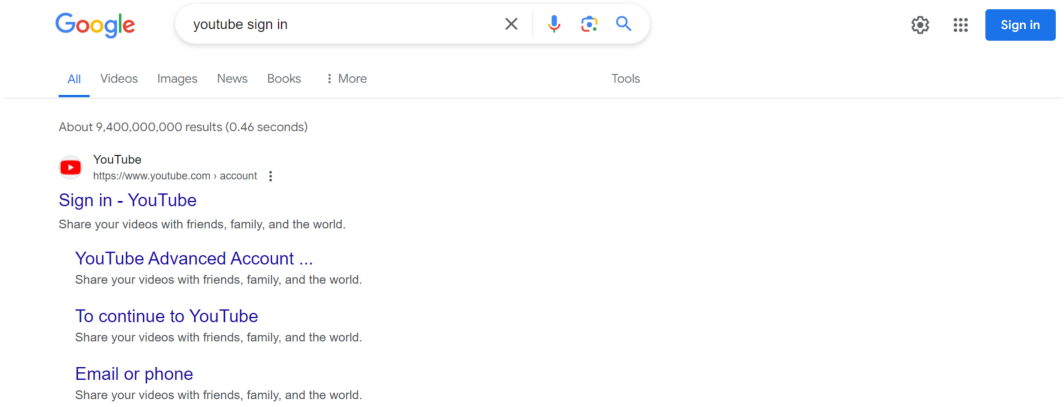


Figure 14: Example of a search query and SERP with navigational intent.

Informational intent means that the user wants to have more information about a certain topic, to learn more about a certain subject. An example would be “why is palm oil bad for you”.

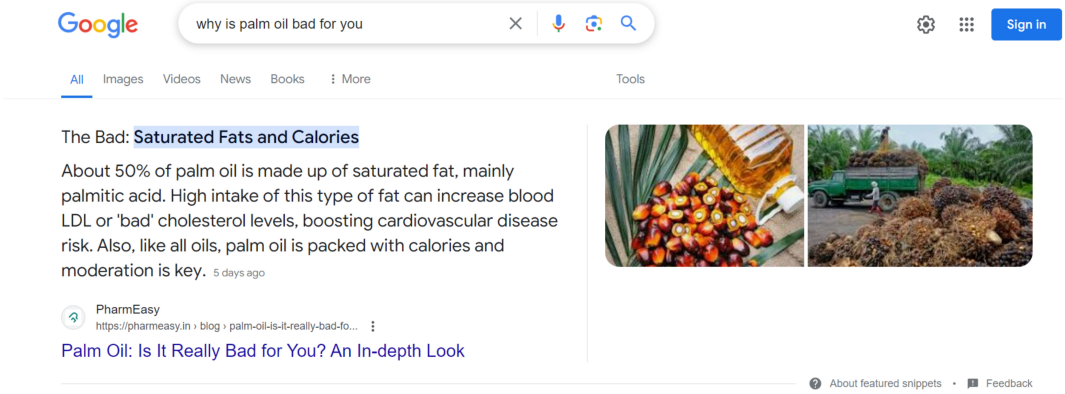


Figure 15: Example of a search query and SERP with informational intent.

Informational queries often contain words such as ”what”, “who", “when”, “how”, and “why”, as they essentially signal questions by the user (*ibidem*). These questions are often answered directly on the SERP by Google, in Direct Answer or Knowledge Panel features, but is it still strategic to try and intercept these visitors. They account for a substantial share of all search queries and represent an opportunity for businesses to show expertise and build authority and trust.

Conceptually, we could place commercial intent midway between informational and transactional intent: the user wants to gather information about a product or service before making the decision of purchasing it. Before deciding which one to subscribe to, a user might search for “cheapest online newspaper”. These types of queries often

involve reviews and advice from third-party websites, as they are typically submitted by users who want to clear their doubts before making a conversion decision.

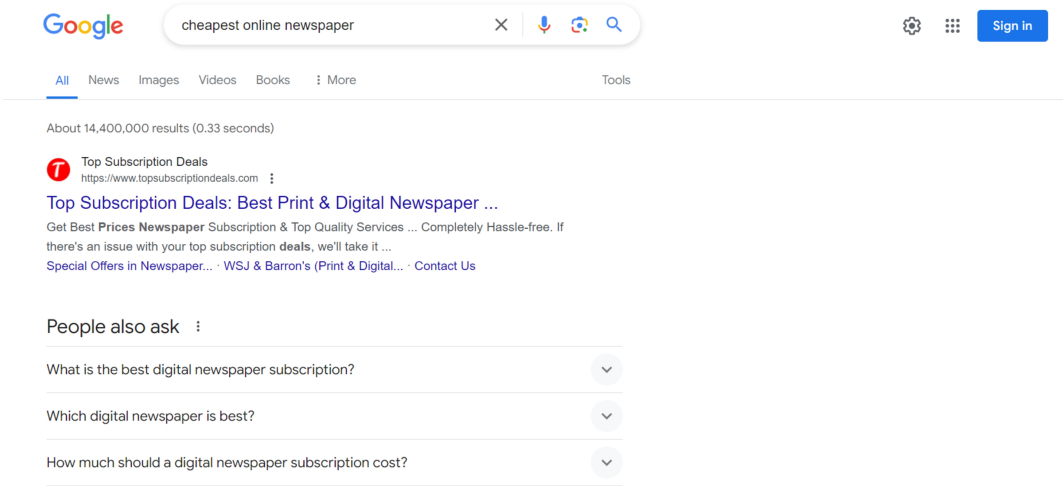


Figure 16: Example of a search query and SERP with commercial intent.

Lastly, transactional intent encompasses all those queries aimed at completing a purchase or similar actions, such as subscribing to a website or downloading a file. Users driven by transactional intent are the most valuable and more important to intercept, because they show a concrete intention of making a conversion.

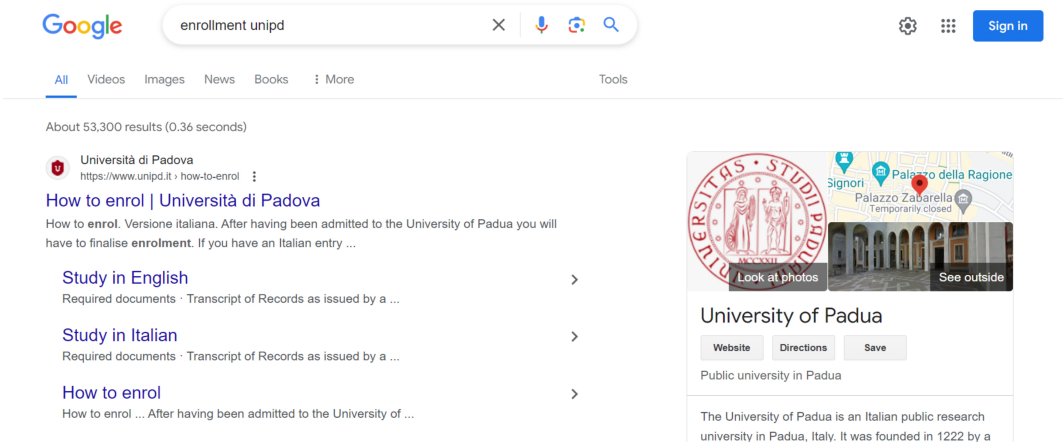


Figure 17: Example of a search query and SERP with transactional intent.

The order “information”; “navigational”, “commercial”, “and “transactional” is not accidental; it resembles the structure of the marketing funnel. The funnel – often also called AIDA model, for “Awareness”, “Interest”, “Desire”, and “Action” – is a marketing tool that provides a simplified framework for the customer journey; it reflects the journey the ideal consumer takes from getting to know the brand, in the “awareness”

phase, to becoming a customer by performing an “action”. In the intermediate steps, the prospect starts developing interest for the brand and a desire for one of its products or services. In reality, customers do not really go through all these stages linearly, yet the funnel model can be used as reference to executing different marketing strategies to maximize results (Ong, 204). SEO plays a fundamental role in optimizing this encounter with the custom, as today his journey largely takes place online: by identifying at what level of the funnel he is, it is possible to implement different approaches and techniques to encourage customers to engage. To promote interest and desire, for example, webmasters can work on their content or show reviews and opinions of other customers and convince prospects to convert by offering incentives and making the conversion process easy.

Deciphering the search intent of users helps assess their position in the funnel. The lower he is in the funnel, the narrower is his research, and the firmer is his intention is to convert (Handley, 2023b). Applying the above-mentioned categories to user’s search allows webmasters to provide content that is useful and timely. If the intent is informational, for example, pages that offer insightful readings about the research topics are more likely to rank higher, therefore a blog post might be more strategic; when the search intent is commercial, on the other hand, the user is already familiar with the product but needs some last piece of advice before making a decision, so further explanation would be unnecessary and better replaced with reviews or products comparisons (Handley, 2024h).

Search intent is at the foundation of any SEO decision, because in order to intercept new potential visitors (or customers), knowing why they are looking for something is equally as important as knowing what that something is. Understanding search intent thoroughly allows targeting the most suited keywords, tailoring page content to directly address their queries and, therefore, increase relevance and engagement, and ultimately ranking higher in search results.

2.2 Keyword strategy

Keyword research is important for SEO because it reflects what relevant audiences are searching for and enables a business to act so as to intercept them during their quest. Based on this premise, webmasters develop a keyword strategy, a plan that outlines on what keywords to concentrate SEO efforts and which helps prioritize and target them effectively (Wilkinson, 2024). Various approaches can be used to determine the most advantageous keywords. Targeting high-volume keywords might, intuitively, seem more appealing, as the returns in organic traffic are likely to be higher, but, as we have seen, keyword volume often goes hand in hand with keyword difficulty and CPC; this means that higher volume usually equals higher difficulty and higher CPC. Additionally, higher keyword volume does not necessarily mean high productivity: relatively speaking, it might be more effective to work with keywords which are more specific to the web page: this is because higher specificity in the research process is more often than not linked to a stronger conversion intention. Users who search for "running kids shoes waterproof", for example, are most probably fewer than users searching for "sports shoes", but they are potentially more likely to convert. It is therefore important that a keyword strategy be in line with the business strategy: a website owner should choose to try ranking for a less searched keyword which, however, is more relevant to the market segment he is serving and to his goals.

One type of keywords which generally score lower on the three above-mentioned keyword metrics – search volume, keyword difficulty, CPC – but which can, nonetheless, have considerable potential are called long-tail keywords. Long-tail keywords are more complex and specific strings – usually made up by three or more individual words – for which competition is much lower than shorter keywords, but which, at the same time, attract more targeted traffic (Lyons, 2023b). This is because the specificity of the user's research tells that he is already quite knowledgeable about the researched topic, and he is ready to make a conversion. For the short-tail keywords such as "tennis shoes", long-tail variants might be "tennis shoes on sale" or "tennis shoes for indoors". Long tail keywords are relatively less frequently searched in comparison to

regular keywords, but, if aggregated, they represent the vast majority of all performed searches. This means that, alternatively to targeting one single high-volume keyword, it might be more convenient to invest on multiple long-tail keywords. Their specificity is also an advantage for creating content complementary to the targeted keywords: a keyword such as “tennis court” is indeed very broad in meaning and could apply to business who build and sell tennis courts, facilities where courts can be booked to play, or blogs that describe the differences between grass, clay, and hard courts, for example. Conversely, “tennis court measurements” is a longer and more precise keyword which gives useful insight into what the user is hoping to find with his research.

Another type of potentially strategic keywords to target are question keywords. Question keywords are keywords which are very valuable when it comes to deciphering users’ intentions because they are formulated to form a question. Queries starting with interrogatives – often particles such as “what”, “how”, or “why” – are relatively easy to interpret and to address with suitable content. “What is the best concrete for a tennis court?” or “why are clay courts cheaper than grass courts” contain very specific questions, which makes the content creation process much easier, as the user is explicitly saying what he is looking for. This type of queries, however, are both an opportunity and a thread: in addition to regular competitors, webmasters must also beware of the competition represented by Google’s SERP features. Features such as Direct Answers and People Also Ask are characterized by the fact that queries are answered directly on the SERP, therefore making it optional for the users to click on the search result to access the information. As already mentioned in Chapter one, such cases are called zero-click searches. This might suggest that addressing question queries – which are easily answerable by the search engine – is not so convenient. However, it must also be considered that these features can be turned to one’s advantage; indeed, managing to be quoted by Google in these boxes might still generate traffic by attracting visitors to the website (Andrienko, 2020). This can be achieved by providing an answer to the query which is concise – so that the search engine can easily understand it and show it is the Direct Answers box, for example –, yet enticing, so that more curious users are tempted to click on the snippet.

3. Content creation

The importance of researching and strategically selecting keywords ought to then be paired with equally strategically created content. Content creation is the process of designing and delivering useful information to one's audience, with the goal of creating value for users and satisfying their needs (Perco, 2023). It is a fundamental part of any SEO strategy, and it is instrumental to building trust, authority, and reputation. Starting from high-potential keywords, webmasters can create high-quality content around those keywords. Quality is very important in content creation, as it is instrumental to achieve any respectable website's primary goal – serving the user – and is of significant relevance for ranking. This is because quality – among other factors – directly affects user experience; this, in turn, has a great impact on ranking, firstly because a good user experience spontaneously drives more traffic to the website; secondly, because Google derives useful so-called user signals from it, which are used to calculate a page's ranking score (van de Rakt, 2023). User signals are among the main ranking factors for Google and can determine whether a page will be considered relevant for a certain search query or not; in fact, content quality is arguably one of the most important of all ranking factors. A good example of user signal is bounce rate, a metric which calculates, by average, how much time a user spends on a certain page before leaving, either to further proceed on the same website or by exiting it altogether: a lower bounce rate tends to indicate that visitors are engaged by the content and, therefore stay on the page. Furthermore, through more and more sophisticated algorithms, Google can understand natural language increasingly better, analyzing semantic connections, synonyms and context – as explained in the previous chapter, the Humming update of 2013 marked a turning point in this area (Cunningham, 2024). These arguments show, in other words, that the better a piece of content is written, the better it will perform.

3.1 Types of content

When creating user-oriented content, the options are multiple. The preference of a certain format over the other mainly depends on the type of content itself and on the search intent. A video, for example, may be more suited for explaining concepts in a very practical and graphic way – for example, changing a car tire: rather than reading about it, it is much more effective to see it to understand how it works. Videos allow for more concise descriptions and are particularly suited for content such as tutorials but can also contribute to creating user engagement through storytelling and dynamic visuals; indeed, according to The State of Content Marketing 2023 Global Report by industry leader Semrush, videos are the best performing type of content. Besides their high engagement potential, they also positively influence SEO because they require the viewer to stay on the page for a longer amount of time, thus positively affecting bounce rate, an important metric useful to measure how much time visitors spend on a page (Dean, 2023a). This format, however, presents one major inconvenience: Google bots cannot properly analyze them. Because they are, by nature, non-textual files, and search engines can only analyze textual elements, they have limited searchability. A certain piece of information mentioned in the video might well intercept the user's need, but it might not show it on the SERP. For this reason, it is particularly important to accurately optimize titles, descriptions, and tags surrounding the video, integrating primary and secondary keywords.

This same issue affects another type of increasingly popular media: podcasts. Podcasts are digital audio programs to which users can listen especially when searching for educational content. This means of communication allows to foster a deeper, more personal connection with the audience, therefore building trust and creating a more interactive community around a business (Perco, 2023). Podcasts are often organized in series which cover specific formats or topics, and this makes it possible to demonstrate one's expertise on a specific subject.

An even higher degree of interactivity can be reached through social networks: social media posts are a powerful content marketing tool, as they spread on platforms

characterized by ever-growing popularity. The number of social media users currently exceeds five billion, and in only a few years it is expected to reach six billion. This underlines the importance for a business to be present on social networks. Facebook, Instagram and TikTok are currently the most crowded online social spaces[15] and offer the opportunity to directly interact with users (Silva, 2024a). Although educational and informative content can still be found on social media platforms, for example in the form of infographics[16], this type of posts generally distance themselves from more formal content and are usually employed to connect with one's audience in a more casual and spontaneous way, creating engagement and encouraging communication, which over the long term can lead to sentiments of trust and loyalty. It is important to note that, despite being effective marketing tools which can greatly benefit a business' marketing strategy, the content created for these platforms does not directly affect organic ranking: number of followers, views and shares are not signals which are taken into consideration by the algorithm for assessing ranking scores. They are, however, significantly relevant for Google's Quality Rater Guidelines, and can – though indirectly – positively impact a website's authoritativeness and increase customers' familiarity with the brand. More precisely, implementing social media content in one's SEO strategy can boost traffic via a specific kind of searches called branded searches. Branded search concerns queries that specifically contain the name of the brand or of one of its products or services, for instance "Eastpak school bags on sale" as opposed to "school bags on sale". Branded searches are necessarily performed by users that already have at least some knowledge of the brand – gained perhaps via social media marketing – and can be labeled as navigational queries, as the user is searching for a specific website. Measuring branded traffic can be an useful metric to track brand awareness and to calculate one's share of search, that is, the volume of searches related to a certain brand within its industry in comparison to its competitors (Tan, 2023).

15 According to Statista (2024), Instagram and TikTok stand behind YouTube and WhatsApp; however, while still classifying as social networks, these two platforms are not suited for publishing posts.

16 Infographics are visual representations of data points or narratives, very effective to convey complex or data-heavy information in a digestible way.

Among the many different options from which one can choose when it comes to creating content, the format which is perhaps the most relevant for SEO are blog posts. Blogs are websites – or sections within websites – where editorial content is published and represent one of the most effective resources for boosting organic traffic. Through blogging, creators can engage with their audience in a variety of ways, and opt for the one more suited for their audience and their objective. “How-To” guides, for example, are broadly used to explain a procedure in clear steps, whereas long and comprehensive articles are ideal for educational content and offer a great opportunity to demonstrate the authoritativeness of a website. Product reviews, on the other hand, are more suited to address potential interrogatives prospective customers might have. According to the State of Content Marketing Global Report (2023) by the industry leader Semrush, blog posts score among the highest formats in performance, and owe their popularity precisely to their versatility. From an SEO perspective, blogs are indeed a very strategic tool. One big difference in comparison to other formats is the fact that, because they are, in essence, all-text posts, they can be easily searched by search engines, and therefore rank more accurately. This also makes them more manageable than a video or an audio file, for example, as they can be easily edited and updated to always offer users high-quality content. In addition to versatility and high searchability, blogs allow to target multiple keywords at a time, therefore increasing visibility and widening the potential user base; they also help develop brand awareness and trust, opening the door to more informal and personal content to share with one’s audience (Hines, 2022).

These examples of formats are not exclusive to one another; in fact, implementing different types of content format across multiple channels is arguably the best option to build an effective content creating strategy. This enables the brand to reach a wider audience, consisting of different types of users with different consumption preferences and different resources. A comprehensive content strategy implies being present and active on various platforms – websites’ blogs, video sharing platforms, social networks –, but it also entails incorporating various formats in one single piece of content, and vice versa (Solntseva, 2020). Given that blog posts offer an optimal infrastructure to effectively target keywords, thus more efficiently intercept potential

visitors, these can be enriched with visual elements to better illustrate certain concepts; at the same time, blog posts' content can conveniently be converted into other formats to diversify and create a more holistic user experience, for example turning an article into a videos, or summarizing it in a social media post. The plurality of possibilities allows for a wide range of options, and combining different formats is overall beneficial to the website as it appears more appealing and generates more engagement for the visitors.

One special type of content which is particularly effective in creating engagement is User Generated Content (UGC). UGC any piece of content which revolves around a certain brand and which is created by users themselves and shared online (Besik, 2017). These kinds of contributions hold great value both for users as well as for businesses. As explained by the principle of social proof (Cialdini, 1984), which states that in social situations individuals tend to copy the behavior of other people, UGC provides users with insightful information which is particularly appealing to them because it comes directly from peers. Personal experiences of others are perceived as more relevant and trustworthy than content generated by the brand itself, especially if promotional, as it is believed to be more biased and less authentic (Bickart & Schindler, 2001: 32). On the other hand, from a SEO perspective, UGC represents a more cost-effective option than many other content marketing formats, due to the fact that, additionally to being more socially efficient, it is inherently powered by the audience. Furthermore, by creating this kind of content, users offer webmasters useful insight on what keywords are used by the audience to talk about their brand, and eventually incorporate them into the business' keyword strategy.

These contributions can emerge spontaneously, especially if users experience strong – positive as well as negative – feelings towards a brand but can also be leveraged by content creators to boost traffic and increase conversions. To exploit this social mechanism, UGC can be integrated into one's content marketing strategy by encouraging users to connect with the brand and share their experience, for example by launching initiatives such as contests but also, more basically, by creating online spaces where such exchanges can occur, like forums and comment sections. Statistics say that

99.9% of online users consult reviews before converting (Clark, 2023), effectively making e-WoM (electronic word-of-mouth) one of the most determining factors in the decision-making process. Whatever the outlined structure, it is imperative to create content people enjoy consuming, that is easy to read or to understand and that resembles natural language. Once established which keywords to target and what intent hides behind them, content creators have to understand how to actually meet users' needs and expectations. While blog longer posts, for example, indeed have their advantages, it must be kept in mind that users are not usually willing to spend too much time on a page that they do not find useful or enticing and would rather skim through it. This means that a homogeneous wall of text might discourage them from consuming the content.

A strategic layout is fundamental to quickly engage users and keep them reading. To do so, copywriters often use the so-called "inverted pyramid" model. The inverted pyramid is a popular and effective writing technique in journalism which distributes information in a structure that resembles an upside down pyramid: the most fundamental concepts, metaphorically at the base of the pyramid, are here located at the top, and the further down one goes, the more specific are the details (Toonen, 2018a). In other words, the most relevant notions – those who directly answer the user's question – are included at the top of the page, so that visitors will be presented with it as soon as they access the page; here, their attention is at its strongest, and then gradually decreases further down the page. By addressing the main topic in the first paragraph, users can quickly assess whether the following content is useful to them, and if they want to continue reading. This means that the base of the pyramid must provide a concise answer, but also be attractive enough to trigger the reader's curiosity. This section is called "above the fold" – the fold being the breaking line which marks the end of the page visible to the user before scrolling – and it usually includes a website header, which is generally the same for all the pages across the website and gathers important navigational elements but can also host banner for ads, and the heading, that is, the page title; all that fall below this line of separation is indeed called "below the fold". "Above the fold" content is crucial, as it shapes users' first impressions: according to research,

web designers have around 0.05 seconds to convince visitors Lindgaard et al., 2006: 122). It is therefore important to include all that the user might need to navigate the website – menus, search bars, tables of content, which immediately show the reader the content's structure and give him a sense of control over his page experience and can contribute to reducing bounce rate.

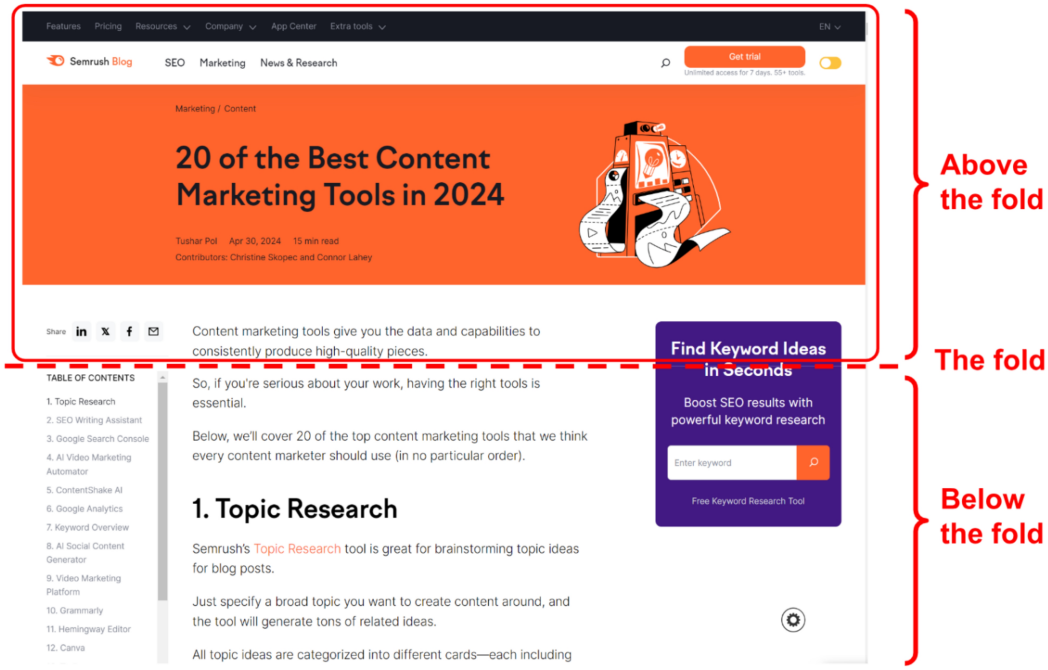


Figure 18: Example of above-the-fold and below-the-fold section on a web page. Screenshot taken from https://www.semrush.com/blog/

The topic introduced above the fold is then further elaborated in the paragraphs below, each of which should be dedicated to a subtopic. Content structure have a significant impacts on readability[17] and what is placed near the top of the page is considered as more important by Google and used to better evaluate page (van de Rakt, 2020).

A particularly relevant and strategically important type of page are landing pages. Landing pages are pages specifically designed to convert traffic into leads and are built to prompt the user to convert. For these pages, the relevant content is a clear promotion of the brand's offering, directly guiding the visitor towards making the

17 Readability refers to how easy it is for a user to understand and navigate a web page or website, and, while it is not directly considered Google's algorithm, it strongly influences, which, in turn, is a confirmed ranking factor (source: Google Search Central).

decisive step, which is often a subscription form or a purchase (Eva, 2021). Definition of the benefits, trust signals such as reviews and scores, and an enticing Call To Action are essential elements to successfully convince a prospect, which, due to the strong transactional nature of the page, should focus on transactional keywords and target users on the lower end of the marketing funnel.

Alongside layout, the search engine bases its quality evaluation on hundreds of factors; as it mentioned in Chapter one, one of the fundamental criteria is E-E-A-T, and these are precisely the principles which a content creator must follow to create relevant, high-quality material. Ideally, a piece of SEO-optimized content would provide real-life, first-hand experience as proof of the creator's reliability, and include his personal or even professional insights and opinions. For more technical content – and most definitely for YMYL pages –, proving one's expertise is imperative: few professionals can say their name speaks for itself; in most cases, the content creator ought to demonstrate he has the credentials required to provide reliable information. This can be achieved by including on the page a short description of one's academic or professional background, presenting themselves as qualified to win the trust of the user (Handley, 2024b).

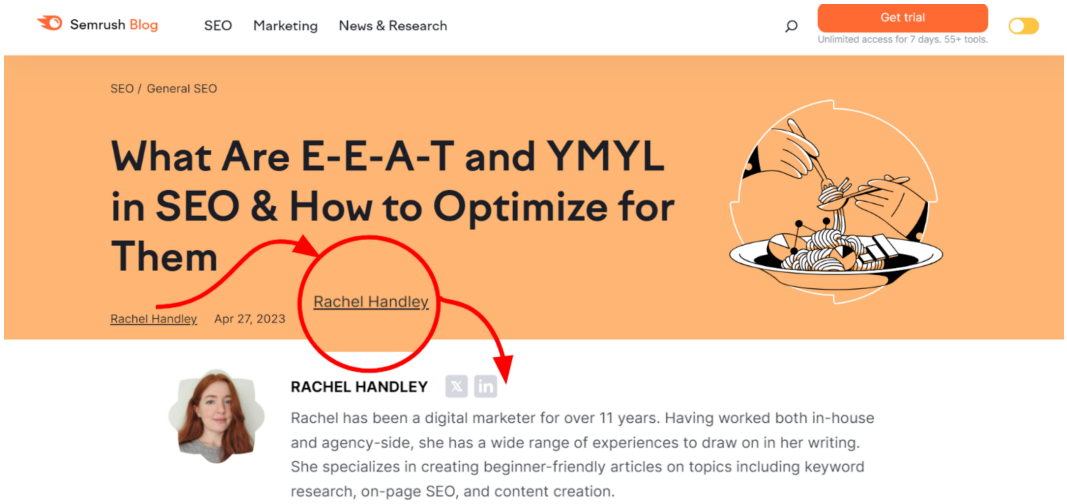


Figure 19: Example of how adding a short presentation of the author can boots credibility. Screenshot taken from https://www.semrush.com/blog/

Above all criteria, trustworthiness remains the most important, the reason being that a lack thereof undermines all efforts to successfully serve the user. The content

published must be accurate and authentic, resting on credible sources and on citation to substantiate them. Creators ought to use their content to address the users' needs and create a valuable resource for them, being transparent about their intentions and avoiding misleading or non-factual information (*ibidem*). To reiterate the relevance of creating original and user-oriented material, in 2022 Google released the Helpful Content update; this update further tackles the difficult task of systematically assessing a website's quality and meant the reward of pages where visitors have had a satisfying user experiences, that is, essentially, pages which feature content created by people and for people, to benefit users and not to gain search engine rankings.

Originality plays a fundamental role in content creation: in such a crowded online space as today's, personality, credibility, and distinctiveness are essential to offer something valuable and authentic to your audience, with which they can resonate. Being original, however, does not mean inventing "new" topics, but rather to fulfill users' needs and expectations with an original approach, writing using a unique angle (Tan & Makosiewics, 2023). A good starting point to assess what to write and what not to is observing what competitors do; in Section 2 we mentioned keyword gap analysis, which consists in comparing a website's keyword ranking to that of its competitors to uncover potential gaps. Similarly, we can talk about content gap analysis to check what other actors in the industry are doing and discover new content opportunities. This provides a clearer idea of what people search for, what kind of content they like to consume, and, consequently, enables creators to understand whether it is better to create a step-by-step tutorial, a guide full of visuals, or an entertaining podcast.

Thanks to the increasing efficiency of Artificial Intelligence, a number of tools are now available to assist web authors in the creative process. These encompass a variety of applications, spanning from writing assistants, to graphic design platforms, to video editing softwares. By leveraging these tools, creators can find inspiration, enhance their productivity, and optimize their content, reaching the intended audience in the best way possible. While human creativity, emotional intelligence, and nuanced understanding of context are essential components that cannot be replicated by AI, such tools are designed to assist creators and serve as valuable aids that empower them to

focus on more strategic and creative aspects of their work and produce higher quality content more efficiently (Burton, 2024).

3.2 Keyword and content clustering

After having identified which keywords to target, what is the motive behind them, and what kind of content is the most effective to appeal to the intended audience, the subsequent step towards content creation is to organize and distribute them across the page. Related to the process of expanding keywords lists, keyword clustering consists in grouping search terms which are similar or related to one another to facilitate the creative process and produce material which can appeal to a wider share of users. On this page, for example, we can see that both "Pilates training" and "Pilates method" are used: these are two very similar searches, for which the search engine will show similar results. It might therefore be beneficial for a website to use both terms on the same page, in order to appear on the SERP for both queries, so that users who search for either one of these keywords will be presented with the same page, and this will ultimately attract more traffic.

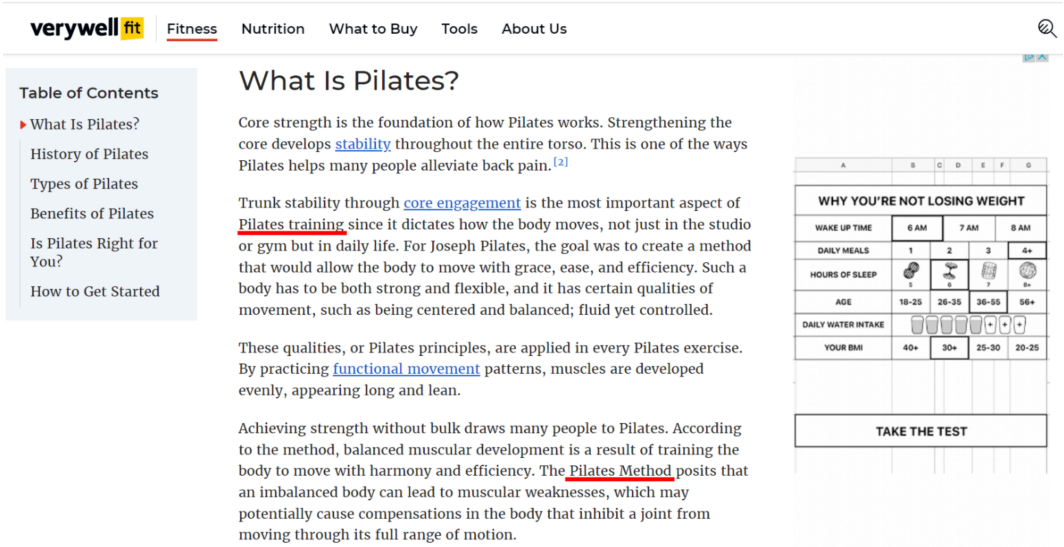


Figure 20: Example of keyword clustering in an article on a fitness blog. Screenshot taken from https://www.verywellfit.com

To develop a keyword cluster, it is first of all necessary to identify the main keyword from which to start building the group: this is called the primary keyword, and is the central term among the ones gathered, the most representative of the topic that will be covered; it is typically the one with the highest search volume and keyword difficulty, and it is included in the title and repeated multiple times throughout the page, though without exaggerating to avoid incurring in issues such as keywords stuffing. All the other related keywords collected in the same cluster can be defined as secondary keywords: they often include synonyms or long-tail alternatives of primary keywords (Handley, 2024i).

Keyword clustering is a strategic tool to strengthen a business' online presence, as it encourages creators to explore a topic from different perspectives, possibly inspiring them to create more extensive and comprehensive content. This will expand the website's perceived expertise on the topic, thus increasing engagement and brand awareness. From the table of content embedded on this page, we can see that the article covers a number of topics related to pilates – its origins, its varieties and its benefits. This is because a user searching "what is pilates", for example, would most probably be interested in knowing more about this practice; it is, therefore, clever to also address other concepts which might be of interest to him.

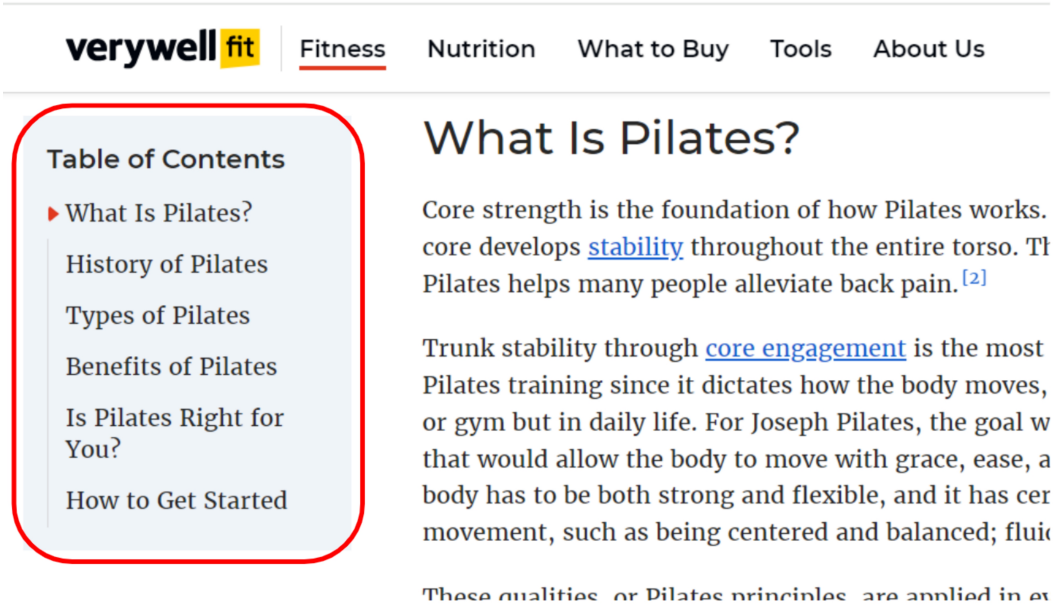


Figure 21: Example of a embedded table of content. Screenshot taken from https://www.verywellfit.com

Additionally, using a more varied terminology contributes to creating more user-friendly; in line with Google's "people-first content" principle, integrating secondary keywords means creating texts which will read more naturally and fluently, closer to how people express themselves in real life. This is rewarded by the search engine, whose constantly updated algorithm is able to recognize synonyms and draw connections between words. When a topic is particularly broad or prolific, it might be worthwhile to focus in-depth on each subtopic individually, developing across multiple pages. A topic cluster (or content cluster) is a group of thematically related web pages which belong to the same website. At the core of each cluster there is a so-called "pillar page", which covers the main, more central topic and which, indeed, functions as a cornerstone to the whole group. Around the pillar develop multiple "cluster pages", each one focusing on a different aspect (Knezevic, 2024). In the example used above, the blogger saw an opportunity to expand the topic by mentioning "stability" and "core engagement", all concepts relevant to pilates, but, at the same time, broad enough to be covered in separate pages.

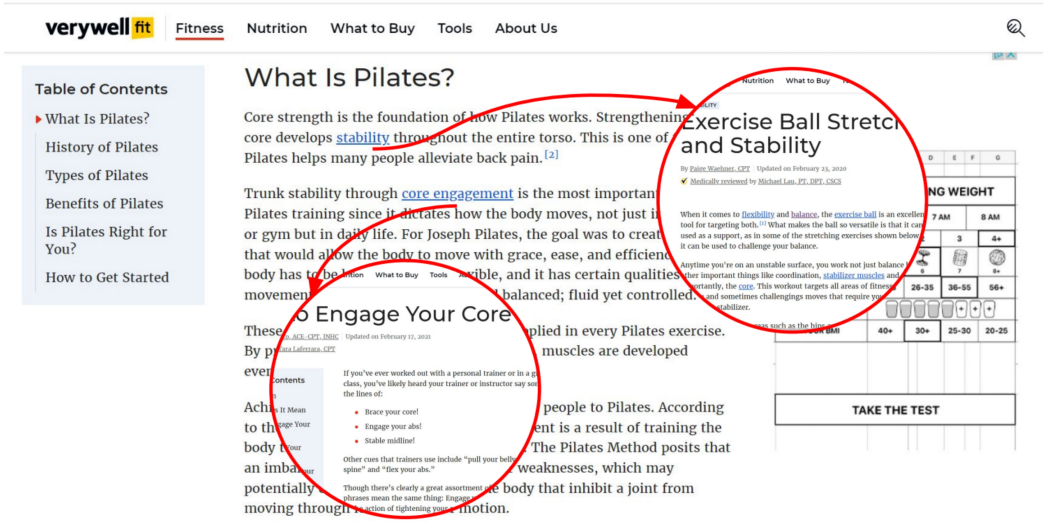


Figure 22: Example of linked embedded on a "pillar page" pointing at related "cluster pages. Screenshot taken from https://www.verywellfit.com

In this case, "What is pilates?" would be the pillar, supported by the complementary cluster pages.

The pillar and its corresponding cluster are not only related to one another topically but also structurally through internal links. Internal links are hyperlinks which connect one page of a website to another of the same website – as opposed to external linking which point to pages outside of the website (Toonen, 2023a). Their function is dually important for SEO: they help both the user and the search engine navigate the web site better. Hyperlinks redirect visitors to other pages: by presenting the user with suggestions of other content, they encourage them to further explore the topic or to engage in similar readings; this ensures a superior online experience, following the customer's journey and boosting engagement. At the same time, this internal linking system is essential for optimal indexation. Indeed, the interconnection among the various pages enables web crawlers to better understand the website and analyze its content more easily. The importance of good internal linking will be further discussed in Chapter three.

It is important to note that, when collecting potential keywords to create a topic cluster, their similarity is not measured from a semantic point of view but rather in terms of search intent rather than semantically (Handley, 2023c). This means, for example, that keywords such as "mountain biking" and "cross country running" can

coexist in the same keyword cluster for a page that covers the topic “outdoor sports". Conversely, it is also possible that the same term be targeted by different pages: since the determining factor is user intent, the pages “what is mountain biking” and “how to participate in mountain biking” attract very different audiences, despite sharing their primary keyword.

3.3 Keyword cannibalization

Besides contributing to an overall better user experience, developing a content strategy which implements keyword clustering can positively affect a website’s performance. By extensively covering a cluster, a website can present itself as a reliable source on the subject matter. Owning multiple pages, each of which addresses a specific aspect of a certain topic, can increase the website’s topical authority, that is, its degree of authoritativeness within its industry. The more authoritative a website is, the more it will be mentioned by other websites, therefore the more will its authoritativeness increase to the eyes of the algorithm too; this, in turn, will then reward the page with a higher ranking score (Sheridan, 2023).

If not done correctly, however, clustering can be actually detrimental for SEO. Expanding one’s coverage on a subject can bring all the above-mentioned benefits, but it can also lead to situations of self-sabotage. Since the goal of each page is to rank as high as possible on the SERP for their targeted keywords, when the same content – which targets the same keywords and matches the same intent – is repeated on multiple pages, these pages will eventually compete against one another for the interest – and clicks – of users. This issue is called keywords cannibalization. If one single website covers the topics “sports centers in Padua” on one page and “sport facilities in Padua” on another, for example, there might be a cannibalization issue, as both covers would try to rank for a search query such as “where to do sport in Padua”. As a result, they will ultimately penalize each other because the search engine cannot establish which of the two pursues its goal more efficiently. This issue is a great waste of SEO resources, since the effort made to optimize one page is neutralized by the effort spent on the other, and

the possibility of one single high-performing page instead of two less fruitful ones undermine potential website authority (Barysevich, 2021). The approach used to tackle this problem differs depending on the actual usefulness of the cannibalized pages. When the duplicate content is superfluous, meaning that it does not deliver any additional value to the user nor to the website, cannibalization can be solved by deleting the duplicate pages, either by mergence or redirection. Merging two similar pages can be done at editorial level, by integrating the content of the less performing one to the one that will be maintained; this approach, however, is applicable when suppressed page does not significantly contribute to the website performance in terms of traffics and when it deletion does not imply a loss of website authority. In cases where a page, despite not being worth keeping, still holds some strategic value, it is usually redirected. Redirecting visitors from the less performing page to its more valuable variant, selected based on ranking, traffic volume, authority, allows maintaining all gains brought by the former and concentrating them on the latter. This is done by communicating to the search engine that a certain page URL – the Uniform Resource Locator, which indicates where a resource is store on the Internet – is not valid anymore, and that when a user requests it, the search engine should instead retrieve another resource at a different URL[18]. From that point onward, users who will try to access the redirected page, will instead be sent to the other one (Hardwick, 2024).

There are many ways to redirect a page, which dive into the more technical part of SEO. Despite not (necessarily) being web developers, SEO practitioners still require some specialized technical knowledge to ensure the well-being of a site and that all optimization actions work smoothly. This kind of implementation, however, can easily be done through a website's Content Management System (CMS), a software platform which enables working on websites without necessarily possessing deep technical skills. CMSs are characterized by intuitive design and user-friendly interface, which allow building websites and publishing content directly through the platform rather than

18 This kind of redirect is called 301 redirects and, among other types, is the most common for SEO. They are characterized by the fact that they indicate a permanent relocation of the page.

through manual coding in HTML markup language, CSS and JavaScript (Loktionova, 2024).

Not all duplicate content requires redirecting: in fact, Google tolerates some degree of similarity across different pages, as it is considered a normal occurrence for a website about healthy living to sometimes reiterate the benefits of exercising regularly, for instance. Similarity is problematic when the same content can be accessed through different URLs: it is the case, for example, of pages that are rather literally almost identical to one another such as different regional variants, device variants, or parameterized pages used listings pages with filters or sorting options applied. Although similar, these kinds of pages cannot be treated as redundant and thus expendable, because they are functional to different uses and situations and must therefore be maintained.

Although duplicate content is actually not a penalizing factor for Google[19], these cases are highly problematic both for user experience and for search engines. On one hand, it might be confusing for visitors to have two different references for pages that they see as identical; on the other, the search engine struggles to determine which one is more valuable than the other and hesitates on which result to show on the SERP (Salsi, 2023). To suggest Google which one of the variants is the main one – and therefore which one will be shown to users –, webmasters can signal it as canonical[20]. Canonicalization means selecting one primary URL whose content is representative of a set of duplicate pages: the canonicalized page will then be given priority by Google and shown on the SERP, as well as use it as the main source to evaluate the content and quality. This so-called deduplication tool helps tackle cannibalization in situations where no other action is possible and concentrates resources on one specific page while still keeping the other ones active.

19 Duplicate pages might still be penalized by Google if it determined that the intent is deliberately manipulative; these occurrences, however, are extremely rare (source: Google).

20 In fact, Google determines autonomously the canonical version on a page, on the bases of a variety of factors; through canonicalization, webmasters can try to influence this decision by making it explicit in the page HTML code or using a CMS (Stox, 2023a).

CHAPTER 3: WEB DESIGN AND OPTIMIZATION PRACTICES

Serving the user in the best way possible is the ultimate goal of any reputable website. Understanding what users need and fulfilling this need with relevant, valuable content stands, therefore, at the core of search engine Optimization. This is, however, not enough. While keyword research and content creation set the basis for a strategic approach to promoting one's web page, it is then necessary to optimize this content in order to successfully achieve this goal. The optimization process encompasses all the various aspects of a page and occurs at various levels, each addressed by a specific branch of SEO.

On-page SEO, or on-page optimization, concerns all those elements which are located on the website or web page itself. This type of optimization aims at making the content more feasibly searchable for people as well as for search engines, and applies a series of techniques and implementations that extends beyond content and keyword targeting and includes the more formal and structural aspects of a website, namely how it is built and how it functions. This particular side of on-page SEO is usually referred to as technical SEO, a subcategory which more specifically deals with codes, tags, website performance and the such to ensure that the site runs smoothly and is easily understandable for search engines as well as users. An important characteristic of on-page and technical SEO is that it is a part of the optimization process over which webmasters have the most control over, as they directly intervene to promote their website.

By contrast, off-page SEO is related to factors which are more difficult to control, as they are external to the website. It deals, for example, with publicity, promotion and more traditional marketing techniques which more broadly affect the image and reputation of the brand, but it also stresses the importance of, again, technical elements, such as external links, which are considered by Google prime a ranking factor.

Both on-page and off-page optimization are fundamental parts of SEO, which, complementarily to one another, encompass all the relevant aspects of a website that can determine its performance and, ultimately, its value for the user.

1. HTML, CSS, JavaScript: what makes a web page

Any form of optimization relies, to a smaller or greater extent, on technical concepts; this makes it imperative to possess some degree of knowledge of web development principles. As already mentioned in Chapter two, many tech operations can often be performed using CMS platforms; however, these softwares cannot compensate for a lack of expertise, which is still an important requirement for any SEO practitioner who wants to understand what elements need to be optimized to best serve users and achieve higher ranking positions.

In Chapter one it has been discussed that crawling, indexing, and ranking are three basic processes that enable a web page to appear on the Search Results Page. Every time a page is published on the Internet, search engines such as Google use robots called web crawlers or web spiders to discover, analyze, and list the page in its index. Indexed pages are then ready to start ranking, and are shown on the SERP. Lastly, when a user clicks on one of the results, the content of that specific website is delivered to his device to be rendered and accessed. This data transfer is performed by browsers, application programs such as Google Chrome or Firefox which function as intermediaries between people and computers called web servers where all the files needed to build a website are stored; when a user clicks on a link in a snippet, the browser locates the resources of the corresponding page, requests the web server to send them and then forwards them to the user.

Just like houses on a map, these resources are findable on the Internet thanks to the address of their host web server, called the Internet Protocol (IP) address. This information is not directly visible to browsers themselves but is accessible via Domain Name Systems (DNS), which associate the domain of a website – for example, www.unipd.it – with its IP address – which could be, for instance, 172.16.254.1. This

conversion is necessary because, on one hand, domains are readable by humans but not as easily by machines, which, in turn, can understand the numerical sequences of IP addresses that are indecipherable to people.

Once all the necessary data has been located and retrieved, the user server finally assembles the web page and renders it for the user to access it. There are three main file types that, together, form functional and interactive websites: HTML, CSS, and JavaScript.

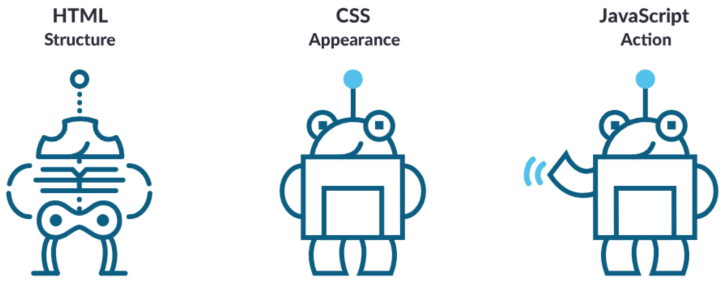


Figure 23: Model of interaction of HTML, CSS, and JavaScript. Image credits: moz.com .

HTML stands for HyperText Markup Language, and it is the standard markup language for creating Web pages. We can see HTML as the most basic part of a website code. It describes its structure, defining elements such as headings, paragraphs, and images within the text[21]. HTML elements are embedded in the code by matching opening and closing tags which define the content of the element – for example, `<h1>` and `</h1>` for the first heading of the page –, and, possibly, an attribute, a piece of text which provides additional specifications for the browser to interpret the element – such as the URL of reference for tags that define hyperlinks.

21 Source: https://www.w3schools.com/html/html_elements.asp

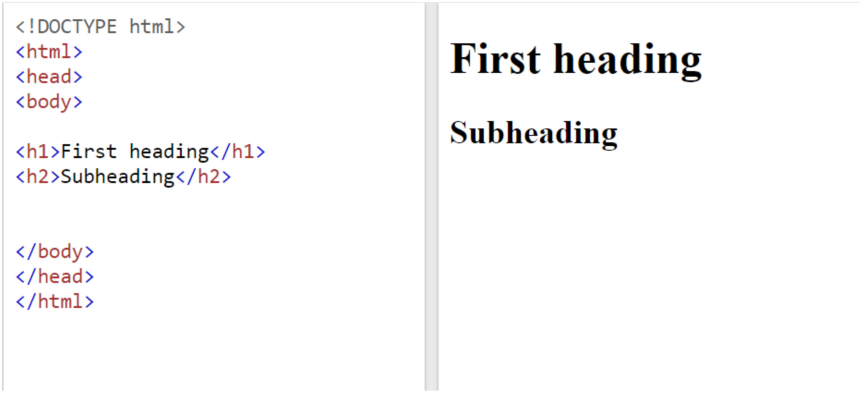


Figure 24: On the left, an HTML page as it appears in text; on the right, how it is rendered on the page.

HTML elements are those necessary to outline the structure of the web page to make it usable for users; not all of these elements, however, are visible to visitors. In fact, content such as text, images, and links are included in the `<body>` tag section, which encloses all those elements that will actually appear on the page; other important elements are metadata, such as meta title and meta description used in the snippet and keywords, which provide useful information about the HTML document itself and are placed in the `<head>` section (Zarudnyi, 2023)[22]. An important characteristic of this markup language is that it also provides semantic signals to search engines. HTML tags provide contextual information regarding the function of the various elements, which are used by the search engine to better understand the purpose of the page and the relevance of its content: for instance, if the keyword "vegan homemade pizza" is included in the page title – `<title> Easy vegan homemade pizza recipe </title>` –, Google will immediately know what the page is about. These labels are also useful to people such as developers to more easily understand what part of the code they are working, or users who use accessibility tools such as text-to-speech readers (Pavlik, 2022b).

A browser reads HTML tags to receive instructions on how to display the document, although not how to actually format it. That is, how to render web design attributes such as font, size, or color. This information is provided in another document complementary to HTML: CSS. Cascading Style Sheets (CSS) are files that indicate

22 Source: https://developer.mozilla.org/en-US/docs/Learn/Getting_started_with_the_web/HTML_basics

how the elements provided in the HTML file should be graphically presented in the browser. One single HTML page – which on its own only appears as a basic text file – can be styled into countless different layouts by using a style sheet language.

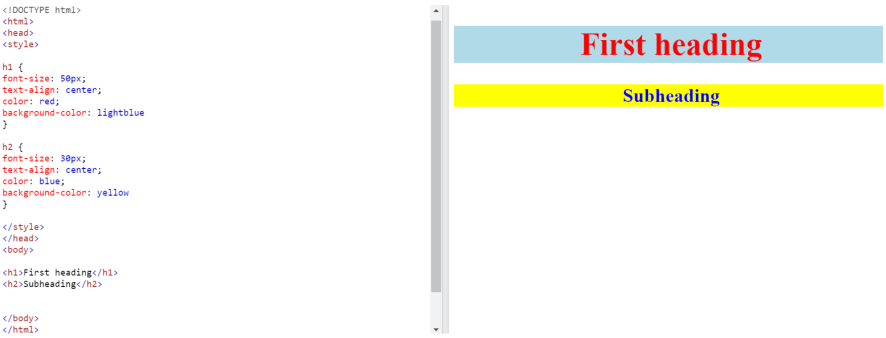


Figure 25: HTML code without formatting (on the left) and with CSS formatting (on the right).

In the example above, for instance, we can see that the `<style>` tag specifies font size, positioning, and colors of the headings. This specification is an example of metadata, and it is therefore embedded in the `<head>` section of the HTML code; this is because the formatting is simple. In reality, websites have much more complex structure and design, which can cause a CSS file to be heavy and ultimately make the retrieving and rendering process slower. To prevent this from happening and negatively affecting user experience, website generally rely on an external stylesheet that is reference in the `<head>` as a link:

```
<link rel="stylesheet" href="mystyle.css">
```

This external file is uploaded separately and contains all CSS rules. Additionally to enabling faster loading, linking stylesheets to the HTML document provides consistency throughout the domain, as it is applied across all pages on the website that would otherwise require individual formatting. The introduction of this system meant a great change in web design, as the page layouts can have a great impact on user experience; this is extremely relevant for SEO, since a better on-page experience is fundamental to boost traffic[23].

23 Source: https://www.w3schools.com/css/css_intro.asp

Lastly, to improve the on-page experience even more and to encourage users to interact, most websites nowadays make use of dynamic layouts. These animated features are implemented using another programming language that is JavaScript. JavaScript files are implemented in the `<head>` of the HTML element to make the page responsive and interactive, including features like moving elements, ads, and pop-ups, also functions such as automatically opening, closing, or refreshing windows. This array of possibilities greatly benefits user experience and SEO, although, similarly to CSS files, JS files carry a lot of information, and because of their size browsers require more time to process them. Indeed, because JavaScript content can be heavy, Googlebot indexes pages which include JavaScript coding in waves, first crawling the static HTML elements, and then dynamic features. This means that the static parts of the site are available to the user sooner than the animations. For this reason, from a SEO perspective, it is important to avoid embedding JavaScript features in SEO-critical elements such as headings or links could pose a problem, as the page would not immediately rank to the full of its potential.

2. User Experience: SEO as “Search Experience Optimization”

While HTML provides the basic structure and information of a web page, CSS and JavaScript well underline the importance of user experience, a central concept in SEO. Despite standing for “search engine optimization”, SEO is as much about people as it is about search enginespoo, as these, ultimately, exist to serve users. An enhanced user experience depends on a number of factors, which span from content quality to website speed. Strictly speaking, what has come to be known as UX constitutes a domain separate from SEO; in fact, all that is addressed by SEO aims precisely at making the online experience of visitors as valuable and satisfying as possible. Put simply, websites that design a good experience for the user, have higher chances of succeeding in attracting traffic and conversions, and increasing their ranking. In general, optimization largely concerns usability and user-friendliness, two concepts that

fundamentally support a good UX, and that are taken into account by search engines to assess ranking scores. A bad user design might drive people aways from one's website, and contribute to a sense of frustration that is detrimental for the brand.

After having caught users' attention on the SERP; a web page has to be built to facilitate navigation and conversions as much as possible; this means organizing its design in a way that marks information easily findable, easily reachable, and intuitive. A way to do so – besides placing frequently-used buttons and links on homepages or on headers and footers – is by mimicking the mind models of users to design the potential paths they could take to get where they want. Part of this also depends on customs and conventions, for which in the Western world, for instance, search bars can usually be found on the top-center or top-right side of the page, and brand logos and menus on the left, but these standards might change from country to country. In this regard we can talk about findability and learnability, central concepts in UX design which concern how discoverable a website is and the extent to which a new visitor can confidently navigate through it even without actively knowing how it work, by activating usage mechanisms trained with previous experiences (van de Rakt, 2016.

As mentioned, UX and SEO are two disciplines very much interconnected with one another. If, on one hand, a dynamic and interactive, aesthetically appealing interface very strongly influences user behavior, stimulating conversion and increasing retention, on the other, it needs to be supported by good infrastructure to ensure functionality. For this reason, alongside other, more "traditional" performance metrics such as page speed and backlinks, Google also uses so-called page experience signals as ranking factors. This broader metric builds on a number of other signals – Core Web Vitals, mobile-friendly, safe browsing measures, and ads and pop-ups intrusiveness – to reward pages that offer a superior experience to their visitors (Gabe, 2020). In the next sections, various kinds of optimization actions will be discussed, highlighting the benefits for search engines but above all for users.

3. On-page optimization

Valuable content is content that is useful and in line with what users need and search for, and it is the most important asset for a brand's online presence. To successfully serve users, web creators must offer informative, interesting, and unique content, strategically created to win their trust. Equally as important is the refining of the product: in addition to the quality and the usefulness of the content offered, a fundamental role is played by how this content is then presented. The first challenge a webmaster must overcome is convincing the users that his page is the one that, among many other candidates on the search results page, best fits their needs, and once achieved this, he has to make sure that what they see meet their expectations and they decide to stay: a page that is chaotic or too crammed will most likely push users away, even if they found the snippet enticing enough to choose to click on that specific link. To prevent this from happening, it is important to devote efforts towards readability. Readability takes into account editorial elements such as sentence length, sentence structure, word complexity, but also structural aspects, such as formatting, and the visual components of a page (van de Rakt, 2021a).

3.1 Content organization

Optimizing content is a process that concerns how the pieces of information are organized, and starts even before the user clicks on the link. The meta title – or title tag – and the meta description are elements that form snippets on the SERP, as we have introduced in Chapter one.

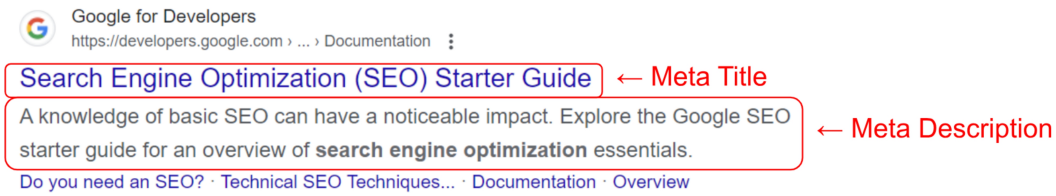


Figure 26: Example of snippet with optimized meta title and meta description.

These elements play a primary role in attracting visitors, as it is what they see first on the results page when submitting a search query; it is, therefore, important that

the title be appealing and accurate about what the page is about. A good way to ensure this is by including a page's primary keyword and recalling the page title, so that users know what to expect when clicking on the link (Lyons, 2022). The title tag used for the page in the example, for instance, would be:

```
<title>search engine Optimization (SEO) Starter Guide</title>
```

and

```
<meta name="description" content="A knowledge of basic SEO can have a noticeable impact. Explore the Google SEO starter guide for an overview of search engine optimization essentials." />
```

The meta title can coincide with the title used on the page, but not necessarily. Optimized meta tags are not only descriptive, but are also designed to fit well into their designated spaces. It might be the case that the page title is too long, therefore needs shortening to appear on the SERP. When optimizing these tags, it is important to pay attention to their length; there is not a strict, fixed character limit, yet it is suggested to keep the number under sixty for titles and around one hundred and fifteen characters for descriptions, so that each meta element will be shown in its entirety on the SERP and not abruptly truncated by Google.

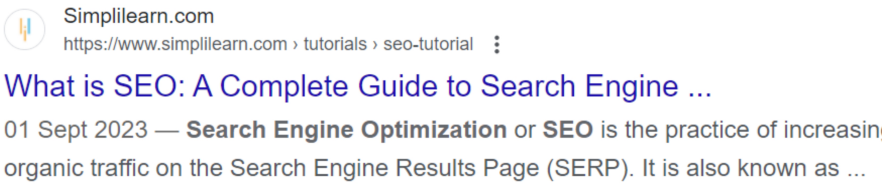


Figure 27: Example of non-optimized, truncated, meta title and meta description in a snippet on the SERP.

An optimized description, on the other hand, would not be taken directly from a fragment of page text but would rather summarize the content of the page and often contain a call to action – such as, "Explore" "Discover", "Find out more" – to encourage the user to visit the page and to explicit what intent it fulfills (Toonen, 2023b). These elements are then used by the search engine for browser tabs or link previews, which can even further facilitate navigation.

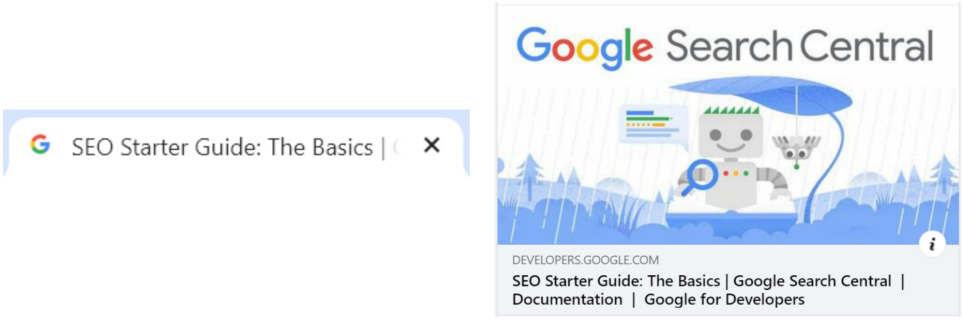


Figure 28: An optimized meta title in a browser tab and in a link preview.

In other cases, the meta title might be different from the page title because it might be more strategic to include the brand's name or particular value proposition which can appeal to users on the SERP but that is not relevant to include on the page itself. It is important to note, however, that optimizing meta tags does not guarantee that these will be shown on the results page; in fact, the final decision about how the meta title and the meta description will look depends on Google itself. It might in some cases ignore the ones chosen by the website owner and replace them with other strings of texts which it deems more appropriate. For this reason it is important to optimize them using the targeted keyword and descriptive terms, so as to make it clear to the search engine what pieces of information. Once the user has clicked on the link, it is then important to design the page content in a way that quickly catches his attention. In blog posts, an important textual feature are paragraphs and headings. Paragraphs should gradually provide more and more specific information on the covered topic; each topic must be introduced by titles and subtitles to help the user better and more intuitively understand the article layout.

The `<header>` tag is the first tag in the content structure of a page, and it is used to define an introductory section; it often includes a navigational menu, a search bar or link to social media accounts. The subsequent sections are then marked with other heading elements which use `<h>` tags. On one single page, there can be multiple `<h>` tags, which hierarchically span from 1 to 6. The H1 tag – or, `<h1>` – is therefore the main title of the page, and it is the one that indeed indicates the page's primary topic to visitors and search engines (Silva, 2024b).

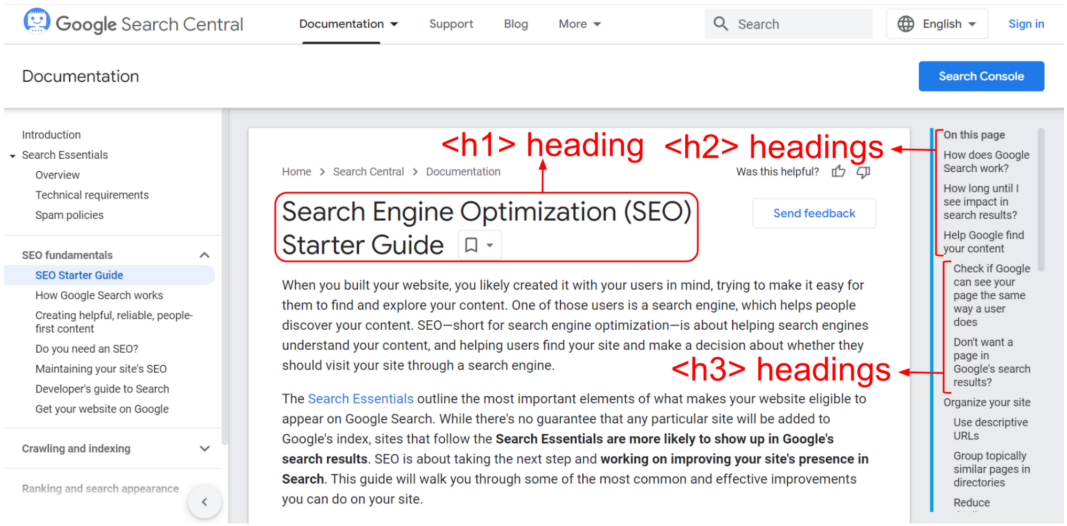


Figure 29: Example of web page with <h1>, <h2>, and <h3> headings.

As mentioned, the title tag and the H1 tag can be the same, and the latter will oftentimes be identified by Google as the meta title; however, it would be more strategic to optimize the H1 tag to produce a similar title, but more suited for the SERP.

Additionally to `<h1>`, in the example above we can see the `<h2>` tags “How does Google Search work?”, “How long until I see impact in search results?”, and “Help Google find my content”, and the `<h3>` tags “Check if Google can see your page the same way a user does” and “Don't want a page in Google's search results?”.

3.2 Visual optimization

Variety in content is the key for a more compelling and engaging user experience. Making use of different types of elements – images, videos, charts, infographics – can convey ideas and explain concepts much more clearly than only words and help the user better understand the page. Just as a text needs optimization, so do these visuals. A central issue for these content formats is that, due to their nature, they are heavy files; this means that browsers require more time to load them and render them correctly, and although on one hand, high quality can provide a better user experience, it also implies higher loading time. As will be further explained in Section 4.5, page speed is very important for SEO: slow loading can be critically detrimental for ranking, as is it considered a direct ranking factor by Google, besides being among the main causes of a high bounce rate.

The heaviest type of files are videos: videos are great means of communication which dynamically combine visual and audio input and are often used to effectively illustrate a topic. To optimize the loading time, it is convenient to take some precautions to avoid making this implementation counterproductive. A viable alternative to uploading this kind of content onto the page – which is not inherently bad for SEO, but does, indeed, add supplementary loading weight – is to embed it into the HTML markup through a link. This means, in other words, to link the resource directly from a video streaming platform such as YouTube or Vimeo and make it viewable through an embeddable player integrated into the page's code. This solution solves the issue of having a heavy element while still offering an all-round user experience (Southern, 2021). Regardless of where the original file is sourced, optimizing a video also means working on its textual components, namely the title, the description, but also hashtags; integrating accurate description and strategically placing the target keywords will benefit ranking and increase the possibilities of intercepting potential viewers (van de Rakt, 2021b). Additionally, Google itself relies on titles and description to better understand what the video is showing, since visual elements are in and of themselves not crawlable by bots.

Besides videos, images are the other big category of heavy page elements. Pictures are broadly used in content creation to complement pieces of text, and should always be used when they are relevant to what is being discussed and represent a significant contribution (Toonen, 2023b). Despite being considerably less problematic than videos, images too require some adjustments to make navigation more seamless; in order to reduce loading time, webmasters can avoid using large images and opt for lighter formats such as JPEG or WebP instead of PNG. WebP, in particular, is an extremely efficient format that compresses images with a much smaller damage in quality loss than the alternatives. As is the case for video, adding descriptive textual content to images allows machines to analyze them: this can be done by integrating an image title, a caption, and an alternative text. The difference between these two elements, which might seem similar, is that the image title is not usually visible to the user, whereas the caption appears adjacent to the image itself and is relevant to allow visitors to quickly grasp its meaning, for example when skimming a page to assess if it is worth their time (*ibidem*). Both for title and caption, the targeted keywords should be employed to facilitate the search engine's task; however, using these two elements is not always but rather optional, depending on their actual usefulness in the context. Conversely, a more significant role is covered by the third textual element, the alternative text. Alt-text tags consist in textual complementary descriptions that do not appear on the page itself, but instead make images readable by text-to-speech software. For this picture, the alt text implemented in the HTML tell the software to read it out loud as "An illustration that shows how one page is linking to other relevant resources", for example:

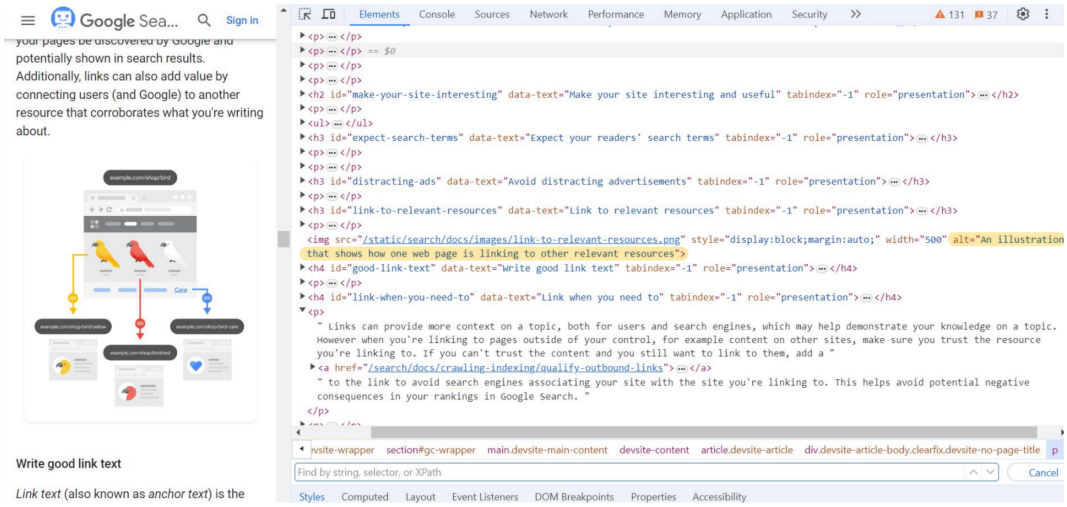


Figure 30: An illustration that shows how one page is linking to other relevant resources

Using alt texts is a strategic SEO practice because it broadens possibilities to rank for images on the SERP: indeed, Google relies on alt texts to better understand what is depicted. Coincidentally, more that 22% of all searches on Google are image searches, and images are often shown before text results; this means that optimizing this type of content gives the page more exposure and potential to boost traffic. Additionally, alt texts load more easily than an image file, which is useful in cases where, due to connection issues for example, the page cannot be rendered properly (Handley, 2023d).

The central purpose of this attribute, however, goes beyond pure SEO objectives. Alternative tags are particularly relevant in the accessibility discourse, as they make images readable for assistive technologies. These tools are of fundamental importance for users with disabilities, and adjusting a page's content to make it more supportable means being more inclusive and serving a wider audience. Remarkably, some accessibility tools are nowaday becoming more and more widespread among all user segments to perform tasks more efficiently. Let us think of search assistants such as Amazon's Alexa or Apple's Siri, which rely on voice search, or Google Lens, which uses pictures as a starting point for the search and works much more efficiently than any descriptive query. These types of web approaches are showing a rapidly increasing

trend, and implementing them into the optimization process means being up to date with the latest, increasing trend and follows users' evolving needs[24].

3.3 Accessibility

The growing worldwide attention for diversity and inclusion is calling for a reshaping of all spaces, in the real as well as the online world. For a means such as the Internet, whose strength lies in making (virtually) everything available to everybody, it is imperative – now more than ever – to develop design that can accommodate every user's needs, regardless of their capabilities[25]. Implementations such as a clearer page outline significantly improves user experience, particularly so for people with impairments who might make use of accessibility tools and who might otherwise have a much more difficult experience.

In 2023, World Wide Web Consortium (W3C), the main international standards organization for the web, released the latest version of the Web Content Accessibility Guidelines (WCAG)[26], a collection of website accessibility guidelines first published in 2008 aimed at guiding webmasters in the creation of more accessible content. These recommendations are meant primarily to help users with disabilities and aim at the standardization of accessibility practices across organizations under four basic principles: perceivable, operable, understandable and robust. Perceivable means that the page can be perceivable regardless of any sensory impairment; to facilitate perception, developers can enable color contrast options, closed-captions into video, and keep the alternative text attribute under 125 characters – the maximum amount supported by screen reading softwares.

The second and third principles – operability and understandability – complementarity imply that a website be feasibly navigable, for example through keyboard shortcuts and without time limits, and easily comprehensible, also thanks to

24 Source: Statista, 2024.
25 Source: W3.org
26 Web Content Accessibility Guidelines (WCAG) 2.1, published by World Wide Web Consortium (W3C), 21 September 2023: https://www.w3.org/TR/WCAG21/ .

titles, call to actions, standardization of navigation mechanisms – that is, for example, that the menu can always be found in the same point on the page.

Finally, W3 recommended a website be robust enough to be supported by different types of devices, namely assistive technologies such as screen readers, but also mobile devices. As is the case for other accessibility implementations, this type of optimization is not only beneficial to users with disabilities, and is in line with another important branch of SEO – mobile SEO –, which encompasses best practice to ensure that the website performs well on mobile devices and ranks higher in mobile search results. This involves various strategies and techniques designed to enhance the mobile user experience, such as improving site speed, implementing responsive design, and optimizing content for mobile viewing. With the growing number of users accessing the internet via smartphones and tablets, mobile SEO has become crucial for businesses to reach wider audiences, higher conversion rates and better customer satisfaction.

4. Technical optimization

Technical optimization is a trivial component of SEO, which ensures that a website not only is well designed, but that it is also performative. As well as offering useful and valuable content that meet the users' needs and desires, it is crucial to present this information in a way that is functional and that facilitates navigation. This side of optimization includes organizing content on a higher, infrastructural level, by connecting all the different parts in a clear and operative way using techniques such as website architecture and internal linking. These tools also contribute to supporting the correct crawling and indexation of the pages, which is an essential step at the basis of any SEO operation, as it makes the online resources available to the public and, therefore, optimizable.

4.1 Website structure and navigation

Site architecture – or information architecture – is the way in which pages are connected to one another within a website. This structure is essential to gather all the published content in an organized way, functional to usability. A well-structured website is of great help for the user, as it facilitates easy navigation and enhances user experience, and the easier it is for visitors to find what they need, the more likely they will be to engage with the brand. Additionally, site architecture is also useful for technical reasons: as we have already seen, search engines discover websites through links. Therefore, providing a clear site map which interconnects all the pages allows Google to crawl and index the site's content more easily, and to understand which page is more important than the other by looking at how they are ordered. If the search engine is provided with a clear site hierarchy, the chances that it will present prospective visitors with the most valuable result are higher (Nicole, 2024a). Conversely, if the arrangement makes little sense, the search engine might struggle to understand the purpose and the importance of the content, and hesitate when it comes to choosing which one to show on the SERP.

The best technique to delineate a straightforward hierarchy is using a so-called flat website architecture. This model enables the user to reach any page with little effort – ideally, with four clicks or less.

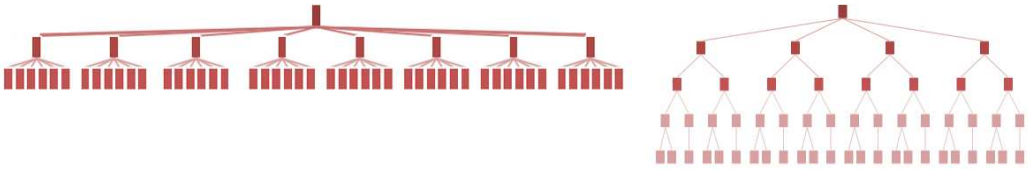


Figure 31: On the left, a flat website structure; on the right, a deep website structure. Image credits: nngroup.com .

By contrast, we can talk about deep website architecture, a model which develops in depth rather than width and therefore implies a higher click depth – that is, the number of clicks necessary to reach individual pages. Because it forces visitors to take extra steps in order to access certain pages, this structure is remarkably less user friendly, and is better avoided (*ibidem*).

The higher level is represented by the home page, which contains a limited number of lower-level pages which are called categories, or sections. Each section can then have subcategories which finally collect individual pages; this is especially true for larger websites, whose extensive offering require a more articulated frame (*ibidem*). This organizational scheme highlights the most valuable pages and groups topically related content together. Indeed, categories often reflect topic clusters, and can help manage content-related issues such as keyword cannibalization. The main objective of planning a good information architecture scheme is to guide visitors in their discovery of the website. This means helping them find what they are searching for but also showing them what the site offers. In other words, a good, user-friendly site structure should encourage visitors to perform valuable actions, and not hinder them: if customers can find what they need easily, they will be much more likely to engage with the brand and convert.

Within a site organizational structure, the home page is certainly the most relevant element. The home page functions as a navigation hub for all the website's visitors. It usually includes the primary services offered by the brand and links directly to the most important pages. This enables users to easily access them, but also to prompt them to perform certain desired actions; this is because from the home page visitors can be easily and strategically redirected to pages in line with the brand's business goal: to have more users subscribe to a membership program, for instance, embedding a link to a subscription form in a prime position directly on the homepage will naturally nudge people to click on it rather than if it were placed elsewhere in the page (Cunningham, 2022).This demonstrates how the way a site is presented contributes to determining the way people interact with it.

The fundamental asset of a site architecture is the menu. A navigational menu is the primary tool users have to understand a website. It collects all the main page categories and clearly reflects how the site is outlined (Cunnimgham, 2023). It often also includes an even more sophisticated navigational aid which is the search bar: such implementation gives visitors the opportunity to enter a query and easily and quickly find exactly what they are searching for and enjoy an optimized user experience. All

these elements are located at the beginning of the page, in the header section, and often paired with a bottom-page menu – called “footer” – which provides access to other important yet less frequently needed pages, such as “Work with us” sections or legal information.

There are different options when it comes to designing a menu, which vary depending on the size of the website and on the devices used; one of the most common types of navigation is the horizontal navigation bar. This layout partly recalls the information architecture of the site itself, keeping the main categories always visible and then, optionally, expanding in drop-down navigation structure to uncover sub-categories (Fitzgerald, 2024).

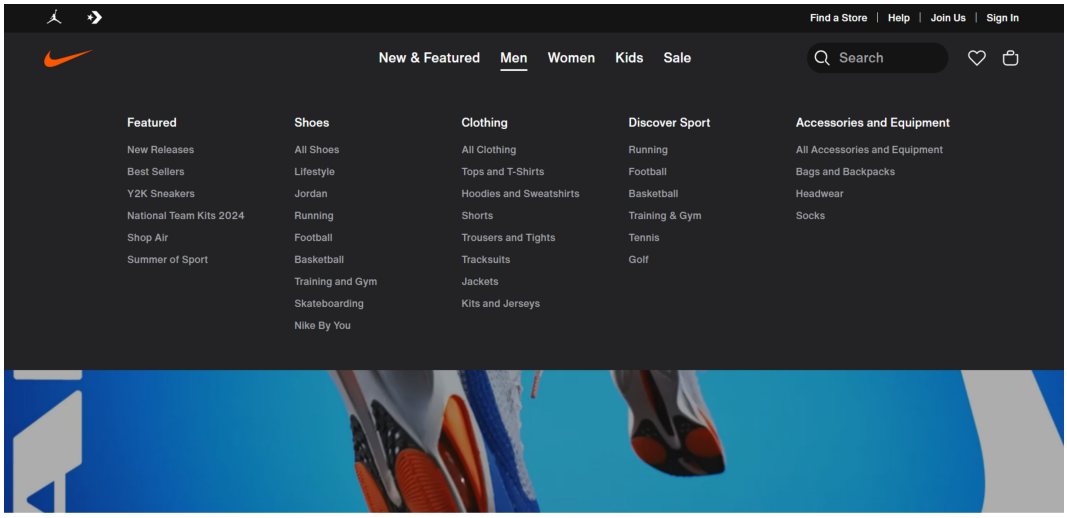


Figure 32: Example of expanding navigational menu. Screenshot taken from https://www,nike.com

Given its fundamental role, the menu bar is generally shown at the top of all pages, so that users can at any point during navigation move back to the home page or to any of the main category pages. This gives them more control over their journey, reduces frustration, and increases engagement. Indeed, these elements play a big role in ensuring a pleasant online experience. Especially when dealing with websites that have hundreds or thousands of articles, for example blogs or even more so e-commerce websites, users can feel overwhelmed and lost. To prevent this, it is useful to work on two additional navigational aids: pagination and breadcrumbs.

Pagination refers to organizing into sequentially numbered pages a series of content pieces which are all grouped into the same section of a website. It is often the case of product galleries or archives, for example. These ordered collections are then shown at the bottom of the page, and give the user the power to choose what articles to view among those available (Nicole, 2024b). The implementation of this feature, for example, enables a reader on an online forum to search for a past discussion much more easily, since he can move to another page and readily skip to an older section; this is also because the interval of numbers helps him make a more accurate estimate of where that might be by showing him how much content there is.

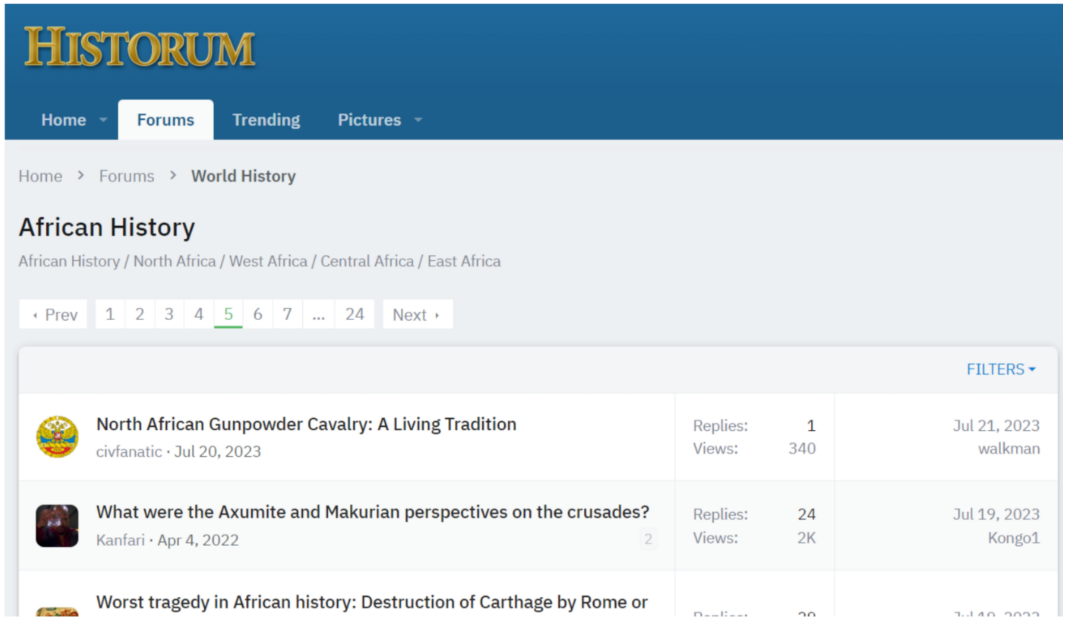


Figure 33: Example of pagination. Screenshot taken from https://www.historum.com/

From a technical perspective, pagination also contributes to reducing loading time; this is because not all the content available is shown at once, but rather only a limited part of it, which is as a consequence loaded more quickly. This aspect is another great advantage of pagination, since among the many ranking factors used by Google, page speed is one of the most influential: a faster page provides an overall better on-page experience and boosts website performance. However, pagination also comes with some disadvantages, one of which is that, if the content is presented in fragments, this implies that users may need to use multiple clicks to move from one page to the other to

explore or find what they are searching for. An increased number of clicks is not optimal for navigation: if a user is forced to click many times to explore a site, he will perceive it as a deterrent and might feel frustration to the point of abandoning the page.

An alternative to pagination which can reduce this kind of friction is infinite scrolling. By allowing this type of navigation, users can move further across a gallery simply by scrolling on one single page, without having to click to uncover more content. In this case too, page speed is preserved because the website loads its content automatically as the user proceeds further down the page. The reduced number of clicks necessary when interacting with this smoother style of navigation can increase engagement, although it does not necessarily reduce friction. In fact, the lack of reference points – provided by pagination in the form of numbers – might frustrate the visitor, who may find himself forced to scroll for a considerable amount of time before reaching the desired product (*ibidem*). From the point of view of SEO, both these options are acceptable and valid, though more or less suited depending on the context; a website that offers a limited amount of articles might benefit more from scrolling navigation, whereas a more complex one would probably use numerical pagination.

Similarly to how the menu serves a visitor who wants to go back to the homepage or to other main pages, there might be cases in which he would rather simply take a few steps back and return on to the page he viewed a few clicks before. To do so, it is possible to integrate another support tool on each page, a record of the sections of the website the user has visited throughout his journey. These directions are called breadcrumbs.

Taking the name from brothers Grimm's "Hansel and Gretel", breadcrumbs are intended to help people trace back their path on a website. They are a text-based navigation path which consist of clickable links to previously visited pages. Usually located at the top of the page, they enhance user experience by giving people more control over their navigation (Toonen, 2023d). There are different types of breadcrumbs; the type that the most literally do their job are called history-based breadcrumbs; indeed, these breadcrumbs report the exact path taken by the user, following the structure:

Home > Previous Page 1 > Previous Page 2 > Page

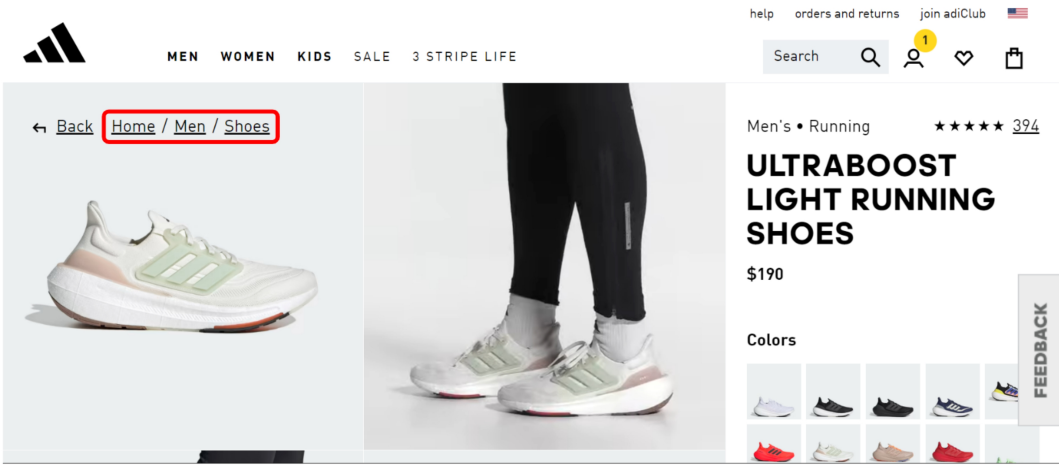


Figure 34: Example of breadcrumbs. Screenshot taken from https://www.adidas.com

Despite being arguably the most accurate reflection of the customer journey, this kind of breadcrumbs are not used very often, as their function can be easily be replaced by the "Go back" button; in fact, in most cases, breadcrumbs are used to offer people a clear indication of where they are on the website. Indeed, we could say that breadcrumbs are a user-focused representation of the website's architecture.

In the example above, for instance, visitors can clearly understand that the page that they are viewing right now – more specifically, a product page – belongs to the collection of pages called Shoes', which is a subcategory to the section 'Men', which, in turn, is a category on the homepage. This kind of breadcrumbs are called hierarchy-based breadcrumbs, and follow the structure:

Home > Main category > Subcategory > Page

Differently to history-based breadcrumbs, the string of text is in the case independent from the actual journey of each individual user: either he reaches a certain page following the route set by the site or through a link, the path shown will be the same. The advantage in this is that, if he, for example, landed on a product page for a specific sofa model, he can then click on the hyperlink "Sofas" in the breadcrumb trail and explore other models.

Lastly, attribute-based breadcrumbs further specify the selected filters, navigational elements used to narrow down a product search by only showing items that

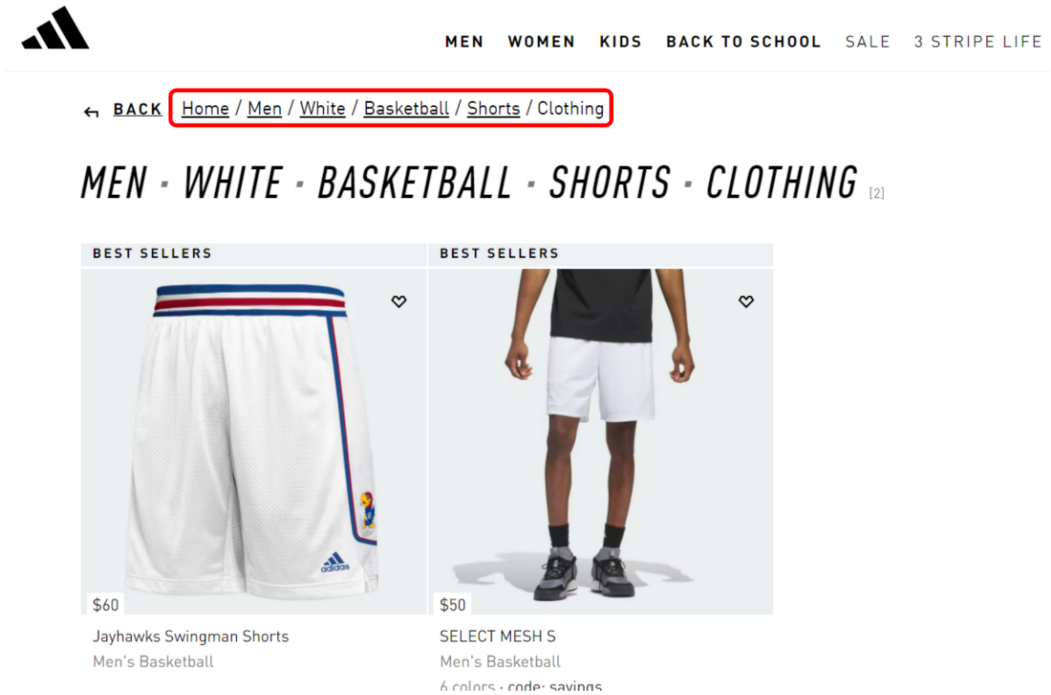


Figure 35: Example of history-based breadcrumbs. Screenshot taken from https://www.adidas.com

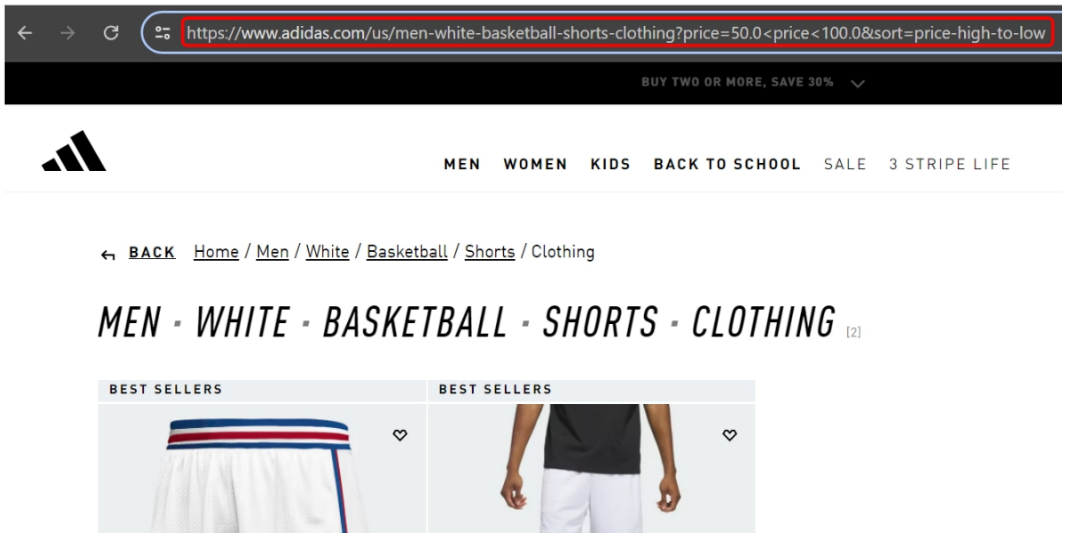
meet certain criteria. They are often used on e-commerce websites, especially if they offer a wide selection (Handley, 2023e).

Breadcrumbs are essential to make navigation smooth and effortless for users. Multi-level structured sites can be confusing, and assisting visitors can encourage them to spend more time on the website and explore other sections, benefitting brand awareness, engagement and conversions. Besides their experiential purposes, they also provide Google with more indication on the internal structure of the site, embedding helpful links to connect all the pages concerned and explicitly ranking them by importance *(ibidem)*. As it has already been mentioned, this is very useful for SEO because it facilitates crawling and accurate ranking.

4.2 URL

The URL – Uniform Resource Locator – is the address of a specific page on the Internet, used by locate it, retrieve it, and deliver it to the user[27]. It is visible in the browser's address bar and it uniquely identifies each resource.

27 The retrieval process will be discussed in detail in Section 4.5 .



An URL is made up of five distinct parts, each providing one piece of information necessary to reference a web page. In the example above, we can identify:

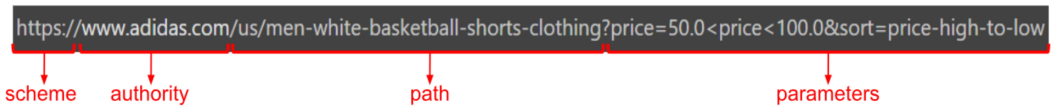


Figure 36: Example of URL.

The scheme is the first part of the URL and it indicates the transfer protocol used to access the resource, that is, the set of rules regulating communication between browser and web server. To render pages, the most common protocols are HTTP (Hypertext Transfer Protocol) or, as in this case, HTTPS (Hypertext Transfer Protocol Secure). While the former is vulnerable to interceptions and hackers, the latter is a more secure version of the traditional protocol introduced in 2000, which includes mechanisms to ensure data integrity and server safety, and communicates using encrypted data; for these reasons, websites using HTTPS are favored by search engines for better ranking, since they provide safer navigation to the user.

What follows the scheme is the authority element. The first part of this element is called subdomain, and it is generally "www.", which stand for World Wide Web[28]; it is followed by the domain – the actual name of the website – and the Top-Level Domain, or TLD, which specifies the kind of website – for example .com for

28 While common, the subdomain is not necessary for the URL to function and can be omitted.

commercial businesses or .gov for government sites – or the region it targets – .it, .us, .es, for instance.

The last two parts of a URL are page-specific, and are the path and the parameters. Subdirectories and subfolders are made explicit in the path, which mirrors the specific location of a web page within a website; in this example, the page is located in the US version of the site and in the category that gathers all clothing items that are white, for men, basketball-related, and labeled as "shorts". Furthermore, the products shown are subject to two filters: one that determines the price range – from $50 to $100 – and the order in which they are presented – from the most to the least expensive. These directives are signaled by parameters, which can introduce various kinds of sorting criteria, such as color, size, pagination (Pol, 2024).

Given the importance of URLs, their optimization is key in SEO. To maximize the chances of ranking high in the SERP, it is important, first of all, to opt for an HTTPS protocol, which is safer and, precisely for this reason, it is a ranking factor for Google as previously mentioned; TLDs, too, need to be selected mindfully, choosing what best describes the website's purpose and target[29], although they do not directly influence ranking. Besides supporting intra-server communication, URLs also function as support for users while they are visiting the site: this means that an organized, descriptive string containing appropriate keywords can significantly improve user experience. The part of the URL that changes with navigation – that is, path and parameters – is generally called URL slug, and it fulfills a function similar to breadcrumbs. URL slugs need to be descriptive, accurate and short to facilitate the search engine's job of understanding what the content of the page is, but most importantly to enhance user experience, and to be easily readable by users, who can use them as a reference to orient themselves on a website. Additionally, URLs also appear on snippets, and if their slugs are too long, they will get truncated (Pavlik, 2024a).

29 Targeting uses of TLD will be further discussed in Chapter four.

4.2 Internal linking

For a website, an efficient design is, of course, pivotal for its success. If users can easily and comfortably navigate through its pages and enjoy their time spent consuming its content, this will spontaneously boost engagement, brand awareness, and ultimately conversions. An effective information architecture, however, is not only crucial for user experience: machines, too, make use of these structures to better understand how a website is built and how its various parts are connected with one another. To do so, they rely on links.

Internal linking is one of the most important parts of technical optimization. It concerns hyperlinks that connect one page to another within the same website and function as an essential instrument to ensure effective navigation and connectivity. Internal links play a central role in enhancing user experience because they allow visitors to move from the home page to their desired product page and guide them in their discovery. The most important kind of links for on-page experience are navigational links. These links are the ones present on the menu and they point to the main pages on the website, such as primary category pages or contact pages. A similar purpose is served by footer links, which are integrated at the bottom of all pages across the website. A third type of links which, as well as being essential for navigation, also contribute to creating value for the user, are contextual links. Contextual links are hyperlinks embedded in editorial content – for example, a blog article – which connect the piece of content in which they are located with other related material. Broadly used within content clusters, they are particularly useful for user experience because they encourage visitors to further explore a specific subject and they boost engagement and topical authority (Harsel, 2023a).

These semantic interconnections can stimulate visitors to engage more with the brand, but they are also useful for search engines themselves. If, on one hand, Google can rely on information architecture to properly understand how a website is structured, on the other it uses internal linking to understand how the various elements of this structure are related on a deeper level. On a gardening website, plants and insects might

be dealt with as different categories, therefore by looking at the internal structure, Google might conclude that they are two separate subjects. However, if a blog post includes a link to another post about bee's key role in maintaining biodiversity, the search engine understands that the two topics are related and uses this information to better outline what the website is about and how it can be valuable for people. If Google has a clear understanding of it, it will match it with relevant search queries and suggest it to truly interested users (van de Rakt, 2018).

Figure 37: Example of an internal link embedded into te body o the text. Screenshot taken from https://www.lawnlove.com

Links are the fundamental discovery technique for search engines. As illustrated in Chapter one, bots crawl the Web to analyze and index as many pages as they can so that they can then be available to people. Crawlers do so spontaneously, by following both internal and external links; while the latter are relatively more difficult to manage – as will be thoroughly explained in Section 5.1 –, the former are much easier for webmasters to control and to use to support their website's performance. Consistently integrating internal links into one's content is a way to ensure that all the resources present on the website get discovered by bots. In the example above, when Google bot analyzes that gardening post, it encounters a link to a bees post, which it follows to discover and analyze that piece of content.

In addition to indexation, hyperlinks can also benefit linked pages in terms of authority. Among Google's E-E-A-T guidelines, authoritativeness is of substantial relevance in determining a page's ranking score, and a page is considered to be authoritative when it is referenced by others. Google's algorithm interprets each link as a vote of trust from one page to the other: the more times a page is linked from other

pages, the more authority it gathers in the eyes of the search engine. An important observation is that not all links are deemed as equal by the algorithm. In fact, similarly to being personally recommended by a top-level manager as opposed to a regular employee, receiving a link from a reputable page is relatively more rewarding for pages. The search engine calculates authority on the basis of the quality and quantity of links pointing to it, and while this is especially relevant for external links, it is also true that the mechanism can be exploited internally to achieve higher scores. This is done by strategically placing connections to the pages one wants to promote on pages that already have high authoritativeness (Toonen, 2023a). To boost the ranking performance of that specific post on the lifeguard of bees, for example, it is ideal to embed a link pointing to it directly from a page which receiver a lot of traffic and performs well – oftentimes, this is the home page, since it is logical the main page of a website be the one with the higher PageRank score. Consequently, that link will be more valuable than a link from another article. This practice is often used for newly published content, for example by promoting in a "Suggested readings" section, so as to quickly raise its rankability and to make Googlebot discover it sooner.

Because authority can be forwarded across multiple pages, the link value will then also spread – in gradually smaller degrees – to the ones that the linked page itself points to. If the bee's post, in turn, links to an article about the threads of climate change, then this will too benefit from that same home page authority score. This system, however, gives webmasters a lot of power to manipulate the ranking system. Indeed, as mentioned in Chapter one, the widespread use of spam linking techniques has been a major issue for search engines. Throughout the years, Google has introduced a number of updates to tackle this problem and discourage the inappropriate use of linking. One relevant measure implemented to regulate internal link building is that, if it is true that authority can be inherited, conversely, authority can also be diluted. A single page has a certain amount of authority it can gift to other pages via links, called link equity; as a consequence, the more links it contains, the lower the value of each link[30]. This means that, although it might be tempting, linking all pages from the main home

30 Source: moz.com

page will result in those links being of poor value. Furthermore, an excessive use of links is not only counterproductive – and overwhelming for the user –, but can also be a penalizing factor for Google.

In addition to the concept of link equity, Google also addresses potential misuse of these practices by evaluating a link based on its context. With the release of more and more sophisticated updates which have implemented advanced machine learning technologies, search engines are now able to understand natural language better and acknowledge more significance to relevance signals, such as semantic context. For SEO, this integration underlines even more the importance of incorporating links in a natural way both for quantity and quality. The appropriateness of a link is evaluated on two main factors: context and anchor text.

The context of a link refers to the context in which it is inserted. In a blog article, for example, the text and sentence that enclose the link are functional for the search engine to establish whether it adds value to the user's experience (Haines, 2023). If the article covers the topics of good gardening practices, then a link pointing to a page promoting running shoes will not be rewarded, as it is not appropriate. More precisely, Google pays particular attention to the exact string into which the link is embedded. This linked phrase is called anchor text. Anchor texts are clickable pieces of text that redirect the user to the linked resource. They are the visible component of a link and serve essentially as a replacement for the raw URL, which would be unintelligible to humans. The term or string of terms to which the link is anchored play an important role in the eyes of the search engine because it should accurately describe the content of the page of destination. This is, once again, to help the machine understand it and rank it more accurately (Hardwick, 202). In the example below, for instance, the links included in the text provide the user with valuable reading suggestions, which are concisely described by the anchor text.

In the HTML code of the page, this hyperlink would look like this:

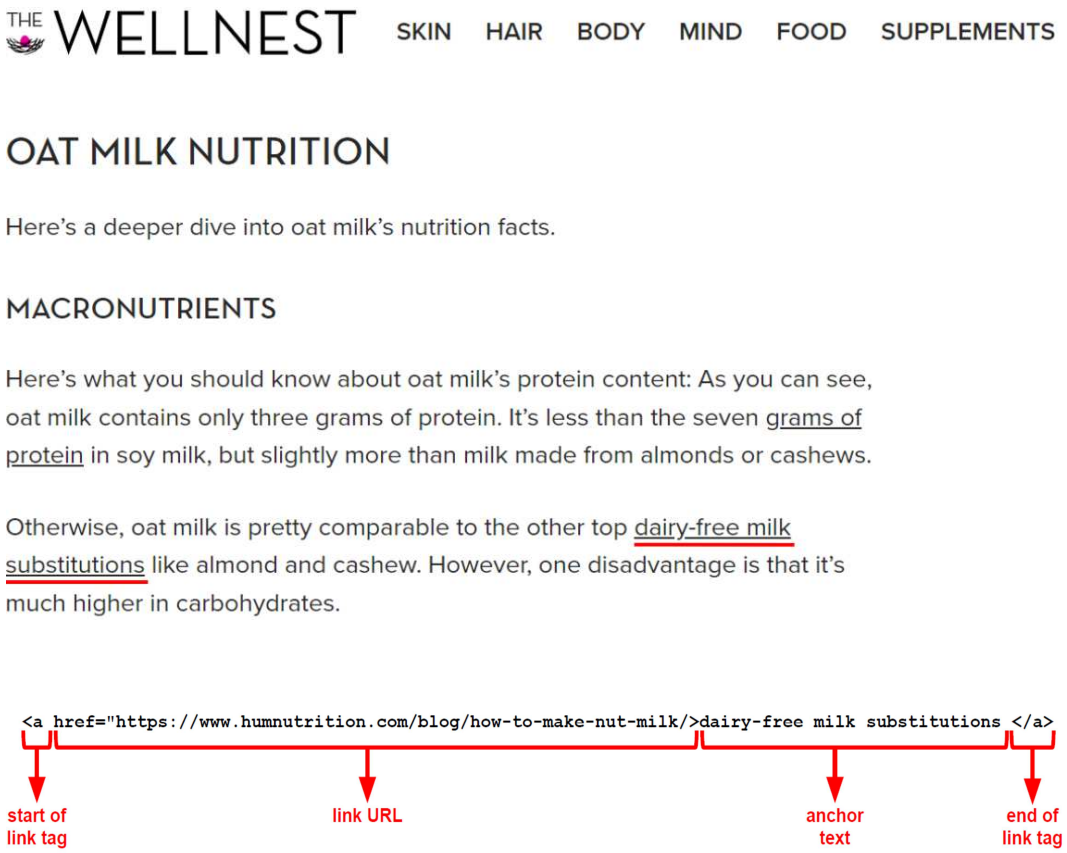


Figure 38: Example of achor text. Screenshot taken from https://www.humnutrition.com

The internal link must be integrated into a sentence naturally: the anchor text must be descriptive to help both readers and bots grasp the meaning of the linked page, and often feature keywords, brand names, and calls to action. Yet, the overuse of keyword-stuffed anchor text might sound unnatural to the reader – and to Google's sophisticated algorithm, which could deem it as an attempt to manipulate a page's ranking. Adopting an authentic and genuine style, designed for people, is fundamental to maintain credibility in the eyes of both users and search engines: hyperlinks are of great use, but must be integrated when their presence adds genuine value to the page.

If used thoughtfully, internal linking can greatly influence ranking and be highly beneficial on multiple levels, both for people and for search engines, enhancing user experience and enabling a more comprehensive and efficient indexation. There might nonetheless be cases where a page exists without being linked to other pages. These are called orphan pages, and are problematic for SEO (Larkin, 2022). This is because, in theory, if not linked to other pages, they cannot be crawled by search engines, nor

indexed, and every possibility of ranking is therefore excluded. This problem can be solved editorially, using internal linking to grant Googlebot access to the page. However, crawling and indexing are time consuming operations, and as the Web grows, it would be impossible for bots to scan every website thoroughly and efficiently. It might therefore still occur that, even if properly linked, crawlers will still require some time to index them. To prevent this from happening to very important or particularly valuable ages, webmasters can provide crawling prompts to the search engine and manipulate indexation.

4.3 Indexing tools

Indexation is the fundamental step that makes an online resource available to the public. Once crawled, a web page is stored in the Google index, ready to be retrieved when a user submits a fitting search query. The relevance of this process makes it a central prerequisite for SEO: without proper indexing, any website might remain invisible to potential visitors, regardless of its compelling content and appealing design. Crawling is an automatic operation that bots perform by spontaneously following links for this reason, it is not clear how often web crawlers visit a website. This means that, after a page is published, it might not be immediately visible online, and the time requested for bot to index it might vary from site to site.

As the Web grows, it would be impossible for bots to scan every website thoroughly. To spread its resources more equally, Google assigns a crawl budget to each website: this budget is the maximum amount of time Googlebot can spend on that single site In this limited time span, it will analyze as many pages as it can (Dean, 2024c). Crawl capacity limit and crawl demand are what determine the crawl budget. Crawl capacity limit indicates the quantity and quality of connections to Googlebot which a website's server can withstand without overloading; crawl demand is calculated on the site's size, popularity and content staleness: if a website publishes content frequently and attracts a lot of traffic, bots will read these factors as signals to visit it more frequently. The more frequently a site is crawled, the sooner new pages will be indexed

and shown on the SERP. Additionally, the less time Googlebot will need to spend on each page to analyze new material, the more pages it will be able to fit in the given time window. This offers an insight on why delivering valuable and unique content is so important, as early as in the crawling phase.

Crawl budget is an important factor to be taken into account when working on technical optimization, because it implies that, potentially, not all pages of a website will be crawled by Google, and therefore will not be rankable. Indeed, pages that exceed the limit will be ignored by the search engine. Webmasters can take a number of measures to optimally manage their websites to maximize their crawling efficiency and avoid the risk of having pages which are invisible to users.

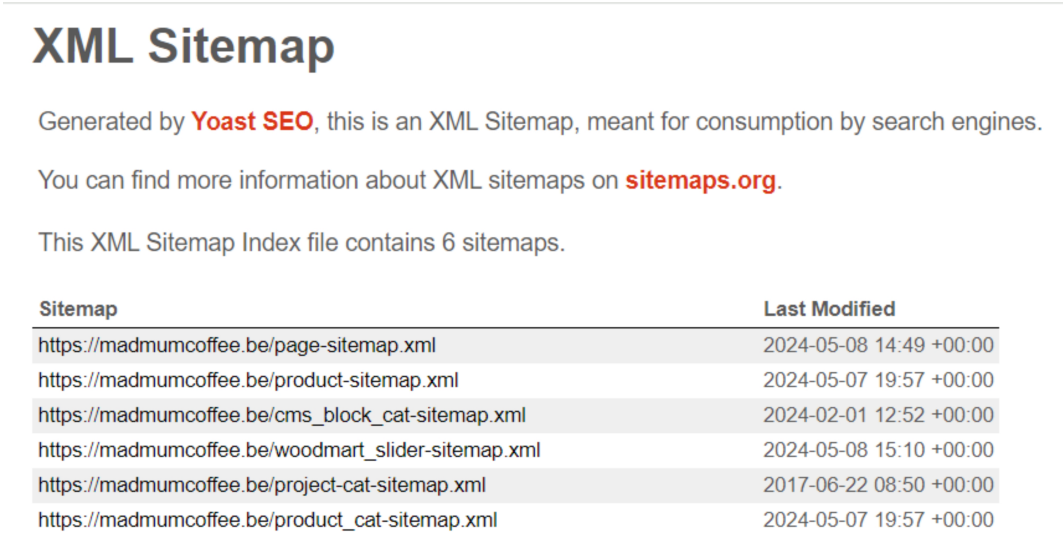
An effective way to communicate to Google the existence of a new or updated page and trigger bots to crawl it, is submitting an XML site map. A XML sitemap is a list of all the important pages of a website, gathered to signal to the search engine that they require prioritized crawling; it differs from an HTML site map for the fact that it does not describe the structure of a site to users, but rather to search engines, with the goal of prompting bots to analyze the site and make the path to indexation more efficient (Handley, 2024l).

Figure 39: Example of an XML site map.

Furthermore, it is possible to implement additional tags to specify the relative importance of the page and signal the bot to crawl it with more or less urgency than the

other URLs on the list (<priority>), to indicate the latest page update (<lastmod>) and the frequency with which it is updated (<changefreq>).

Depending on its size, a single website can have more than one XML sitemaps, so as to request Google to index entire page categories or different types of content, such as an XML sitemap for images, an XML sitemap for videos, and an XML sitemap for blog posts; this is done to help the search engine better understand these files and facilitate the indexation process. In some cases, when websites have considerable amounts of pages, this categorization is necessary, as one single map can list a limited number of URLs – not more than 50,000 URLs, or 50 MB (Toonen, 2022). Multiple XML sitemaps that refer to the same website are collected in an XML sitemap index.



XML Sitemap

Generated by Yoast SEO, this is an XML Sitemap, meant for consumption by search engines.

You can find more information about XML sitemaps on sitemaps.org.

This XML Sitemap Index file contains 6 sitemaps.

Sitemap	Last Modified
https://madmumcoffee.be/page-sitemap.xml	2024-05-08 14:49 +00:00
https://madmumcoffee.be/product-sitemap.xml	2024-05-07 19:57 +00:00
https://madmumcoffee.be/cms_block_cat-sitemap.xml	2024-02-01 12:52 +00:00
https://madmumcoffee.be/woodmart_slider-sitemap.xml	2024-05-08 15:10 +00:00
https://madmumcoffee.be/project-cat-sitemap.xml	2017-06-22 08:50 +00:00
https://madmumcoffee.be/product_cat-sitemap.xml	2024-05-07 19:57 +00:00

Figure 40: Example of an XML sitemap index.

Not all pages of the website have to be included in the list, but rather only the ones which have the potential to attract users, and which actually do deliver value. The aim of a sitemap is to ensure that the essential pages are correctly and promptly indexed, while, at the same time, avoiding wasting precious crawling budget by having Google analyze less important ones. In fact, often websites have pages which they would rather not rank for. It is the case, for example, of check-out pages of e-commerce websites, or privacy policy pages. These types of pages can not only be excluded from the sitemap, but they can also be excluded from crawling altogether.

A robots.txt is a file on a website intended to specify which pages or sections should be crawled and which should be ignored. Similarly to sitemaps, these files help website owners have more control over how search engine crawlers can access their sites: webmasters can indeed give crawling bots the instruction precisely not to crawl certain pages (de Valk, 2023). Blocking unnecessary pages with robots.txt allows Googlebot – Google's web crawler – to spend more crawl budget on pages that matter. A robots.txt file operates on two directives: the specific bot that is crawling the site, and the "uncrawlable" page; by default, any bot can crawl any link it encounters.

Every search engine has its own main web crawler – for example, Google has Googlebot, Bing has Bingbot, Yahoo has Slurp – but also a number of more specialized bots used to crawl specific file types such as images – Googlebot-Image, for instance –, or videos – Googlebot-Video; the robots.txt file can block a specific bot or, more generally, block any crawler that tries to access the page[31]. Next, it is specified what resources the blocked bot(s) are allowed or not allowed to crawl (Silva, 2024c). In the example below, for instance, Google's Googlebot and Googlebot-image are followed by a blank "Disallow" rule, meaning that any page can be crawled; conversely, all other bots are not allowed to crawl a number of pages, which in this case are mainly non-public pages such as "recommendations" and my account".

```
User-Agent: Googlebot
Disallow:

User-Agent: Googlebot-image
Disallow:

User-Agent: *
Disallow: /free_trial/checkout
Disallow: /free_trial/recommendations
Disallow: /logout
Disallow: /my_account
Disallow: /checkout
Disallow: /recommendations
Disallow: /survey/t
```

Figure 41: Example of a robots.txt file.

Sitemaps and robots.txt files function together to guide crawlers in the discovery of a website; however, these tools' main objective is to enable website owners to better

31 This rule is set by using the asterisk symbol as a wildcard.

manage crawl budget rather than preventing indexing altogether. Indeed, while Googlebot will not explore a URL marked with “disallow” in the robots.txt file, it could still access it by accessing it through a link pointing to it from another website: as result, the page in question will still be visible and clickable from the SERP, despite being included in the robots.txt list. This scenario can be particularly undesirable when dealing with sensitive data not meant for public access, such as personal customer information pages, but also for content that is not valuable for users nor search engines. Meta robots are tags which can be embedded in the HTML code of a page to instruct web crawlers on how to crawl, index, and display its content. They are located at the beginning of the page, in the so-called `<head>` section, a space in the HTML code that contains metadata about the document itself, such as headings (Weiler, 2024).

Meta tags support a number of robots that fulfill different functions; most of them affect the aspect of snippets on the SERP, determining parameters such as length – `"max-snippet"` –, visuals previews – `"max-image-preview"` and `"max-video-preview"`, and whether it can be translated to be shown for queries in other languages – `"notranslate"`. The most relevant ones for crawling and indexing are `"noindex"` and `"nofollow"`. The `"noindex"` tag, as the name suggests, gives crawlers the instructions not to index that specific page, thus preventing it from appearing on the SERP.

```
<meta name="robots" content="noindex">
```

This command is often used to hide content from search engines that is private and user-specific, such as login pages, but also content that would more generally be considered as invaluable by the algorithm, for example category pages, “Terms and Conditions” pages, “Thank you” pages e-commerce websites use after a customer completed a conversion, internal search result pages that appear when a user submits a query through the menu’s search bar, or low-quality pages which cannot be deleted from the site, such as technical and administrative content (Alderson, 2021). The meta robots tag **`"noindex"`** and robots.txt serve purposes that sometimes overlap, as both tools ultimately aim at preventing bots from accessing certain content. However, since

meta robots tags are embedded directly on the page itself[32], they are ideal for managing individual pages, whereas robots.txt are generally used to intervene on larger parts of a website.

Conversely, the `"nofollow"` tag allows indexation for the page in question, but blocks crawling for all the URLs pointing to other pages. This operation prevents link equity to be spread and is often used to invalidate links which might hurt the page, on a "reputational" level – when the link points to a website the page does not want to be associated with, for example on moral grounds – or for specific SEO reasons. In cases where not all the links cited on the page are to be disabled, but rather only a few of them, the command is implemented directly in the anchor text tag:

```
<a href="https://example.com/" rel="nofollow">Click
here</a>
```

The `rel="nofollow"` attribute has two variants that apply to paid endorsements and UGC respectively, which are `rel="sponsored"` and `rel="ugc"`**.** The sponsored variant is particularly important to explicitly communicate to the search engine the paid nature of the link, and avoid being penalized for spam techniques, whereas UGC – such as comments or reviews – is generally marked as `"nofollow"` because it tends to be perceived as unnatural linking schemes by the search engine (Go, 2024).

4.4 Page speed optimization

All of the practices that have been discussed above do – more or less directly – relate to page speed, one of the most critical factors for a good user experience. Any potential friction between the user and the page he is trying to access might undermine the efforts to create an engaging site and seamless navigation. Generally, users tend to be intolerant of waiting to get what they want, especially when they know that there are many other competitors ready to offer the answer they are looking for. The feelings of

32For this reason, in order to make the "noindex" tag effective, it is necessary that the specific page that contains it is not blocked by the robots.txt file, as the search engine has to be able to access that page to find the tag (de Valk, 2023).

frustration caused by an unresponsive page can lead to a higher bounce rate and have very negative consequences for traffic and conversion potential.

Page speed refers to the speed of the loading process of the elements that make up the page; it indicates how much time is necessary for it to become fully functional and interactive for its users (Camarena, 2023). Slow load times severely disrupt the overall page experience. Google – which prioritizes users in all of its performance criteria – knows this, and will reward pages that require shorter loading times over those which are slower. While relevance still remains the most SEO important factor – as it is logical, since the main objective is to provide the user with useful and accurate information, above all –, this makes load speed significantly influential for ranking. Speed does not equate quality, but in a highly competitive environment as today's web, it can be decisive. When the search engine has to choose among various pieces of high-quality content, user experience factors often make the difference in which one will rank higher. Creating valuable material remains central, but so is optimizing all those elements that might reduce the responsiveness of the website.

Page speed is, in fact, not a metric in and of itself; it is a broader concept which entails a number of factors – actual metrics that combine to provide real-world measurements of a page's performance efficiency. The most important ones among these metrics are referred to by Google as Core Web Vitals, and include Largest Content Painful (LCP), Cumulative Layout Shift (CLS) and Interaction to Next Paint (INP). These three metrics concern load speed, visual stability, and responsiveness respectively, and are measured based on real-word data from website visitors (Stox, 2023b).

LCP measures the amount of time required to load the largest element on the page, which is usually an image. The higher the value, the more it takes for that piece of content to be visible to the user. According to Google, a good page experience requires a loading time of 2.5 seconds or less; this value is considered by the search engine as a critical threshold, below which any micro-optimizing does not make much difference; this makes it an imperative goal to which webmaster should aim (Mustapic, 2023).

CLS measures the visual stability of the page while it is loading. When the browsers start receiving the content requested to the web server to provide it to the user, the rendering process does not happen all at once: it occurs gradually, and. as it is being loaded, it might happen that the various elements shift in place across the page. This can be detrimental for user experience, as it makes navigation more difficult and less enjoyable. CLS is calculated taking into account the shifting that occur within the first five second from the moment the user first accesses the page, which is the time window in which the most displacements happen; it considers only layout shifts that are inherently part of initial page, and not the ones that result from an action of the user. The dimensions evaluated are the size of the shifting element and the entity of the movement; thus, the lower the CLS score, the more stable the page. Ideally, this value should not exceed 0.1 . (Stox, 2023c)

Lastly, INP measures the degree of responsiveness of the page, that is, how quickly it responds to user interactions. It does so by reporting on the reaction time to clicks, taps, and keyboard interactions throughout a page's lifespan, that is, throughout the entire navigation session. This marks a difference with the predecessor of INP, FID (First Interaction Delay), which only measured the latency time of each interaction individually, and was for this reason recently replaced[33]. A good INP score should be below 200 milliseconds (Stox, 2024).

Figure 42: Google's Core Web Vitals. Image credits: https://web.dev/articles/vitals

To evaluate a page's efficiency based on these criteria, Google offers a performance tool, PageSpeed Insights, to analyze the loading process of a page; using both actual and simulated data, it provides an overall score ranging from 0 to 100 and suggests actionable opportunities for optimization and recommendations for

33Interaction to Next Paint replaced First Input Delay in March 2024. Source: Google

improvement. LCP, FID, and CLS are interconnected and have reciprocal influence on one another: they largely depend on the same factors, which mainly concern the efficiency of communication between the browser and the web server hosting the website, and the size of the page elements that the browser needs to render (Pandya, 2024).

A number of actions can be taken to optimize the various processes involved in the exchange of files between the browser and the web server. In Section 1 it has been explained that, to deliver a page to the user, browsers retrieve and render the page components by requesting them to the web server that stores these resources. This request occurs after the website domain – readable to humans – has been converted into its IP address – readable to machines – though a DNS and the files have been located on the Internet.

Intuitively, a first step toward the optimization or communication between these two so-called agents – the browser and the host server – is to ensure a fast connection by selecting a good hosting provider and a good DNS provider; this means choosing to store one's website on server that has enough storage room, that can handle larger traffic streams, that has a higher bandwidth[34]. Particularly important is the data center location: the host servers that store the website's files are physical servers based across the globe; selecting a host that is located near to the target audience and the server means that the data sent will have to cover a shorted distance, therefore allow for faster the loading time, which, as previously explained, is a decisive factor for page speed. If the host server of an e-commerce website that sells to customers in Belgium is located in Canada, the connection will be less efficient than if it were located in Greece (Hallebeek, 2022). Additionally, most hosting services support a function called keep-alive, which too can positively affect speed. Normally, for every piece of data it needs, the browser establishes a connection with the web server[35], makes the request, receives

34 The term "bandwidth" refers to the amount of data that the server can transfer to the browsers upon request in a certain time period; the larger the amount, the faster the exchange, thus the loading of content.

35 To communicate, the two agents use a language called HTTPS (HyperText Transfer Protocol Secure), a communication protocol which encrypts the information exchanged during the connection and prevent external, unauthorized access, which is particularly important for sensitive data such as username, email address, password, and payment details. HTTPS became the mainstream

the file, and closes the connection; this procedure is then repeated for every piece of data that has to be retrieved to render the page. Given the large number of elements that needs to be transferred, the keep-alive function can be enabled to instruct the server to send all the data in one single connection window rather than opening and closing the connection for each individual resource, thus requiring less time to complete the operation (Sarfraz, 2020). Similarly, DNS providers should also be chosen carefully so as to reduce the IP address lookup time.

To further optimize page speed it is also possible to delay loading for certain pieces of content. Lazy loading is a technique that defers the loading of non-critical resources such as images and videos until they are needed by the user, which can be particularly useful for improving initial page load time, reducing initial page weight, and minimizing system resource usage. By implementing lazy loading, developers can significantly decrease the number of HTTP requests made during the initial page load, since media elements won't be fetched until the user scrolls down the webpage. This not only enhances overall performance but also maintains user experience, as users will seamlessly view images and videos when they come into view without noticing any difference compared to regular loading. However, it's important to note that while lazy loading optimizes several performance metrics, it can adversely affect the Largest Contentful Paint (LCP) score if the largest element is an image, as deferring its load can delay the rendering of this key element and negatively impact the LCP score (Mustapic, 2023).

Besides optimizing the retrieval process, the content itself too can be made more easily transferable. In Section 3.2, it has been mentioned that files such as images and videos tend to be very heavy, for which it is convenient for SEO to compress them using more compass file formats – such as JPEG and WebP – or to embed them on the page by sourcing them from an external platform, such as YouTube. These implementations are crucial, especially if the page includes a lot of visuals, but the

communication protocol in 2019, effectively replacing the preceding HTTP (HyperText Transfer Protocol). This switch was strongly encouraged by search engines, including Google who adopted it as a ranking factor in 2014, thus making it an essential implementation for SEO in order to avoid penalization (Pecánek, 2020).

compression can also concern codes. CSS and JavaScript files in particular are very performative, but do indeed put additional weight on a page, and therefore affect its speed. Through a process called minification, it is possible to reduce the size of a code by removing unnecessary or redundant data within the strings of text that form it – such as white spaces, comments, long tags – without altering their functionality, so as to reduce the rendering time (Pandya 2024).

Another technique useful to reduce loading time is caching. A cache is a copy of the data that the browsers requests to render the website; this set of resources is downloaded and stored after the first visit to the page, and then reused when the user accesses that same website again in the future. This retrieval process makes the next rendering much faster, as the content does not need to be downloaded again (Alderson, 2020). There are two types of cache, browser cache and server cache: browser cache is stored directly in the browser application itself, therefore on the user's device. Webmaster can manipulate this process through metadata embedded in the HTML code meant to instruct browsers, for example signaling which elements to store and which not to and for how long[36]. Conversely, with server caching, downloadable data is stored in servers which work as intermediaries between the host server and the browser. These storage servers are part of what is called a Content Delivery Network. A CDN is an interconnected network of servers geographically distributed across the world which work as data storage centers where browsers can retrieve website files instead of directly downloading them from the web host. The advantage in this lies in the reduced distance, which, as mentioned earlier, affect the loading speed and therefore the page experience: the requested file are cached in multiple server located in various servers called Points of Presence (PoPs) and when a search query is submitted, the browsers connect to the closest one available to request the necessary resources. Also called proxy servers due to their relative proximity to the user, PoPs enable faster transfer and therefore reduce the latency between the request and the response. This parameter can

36 Certain types of content, such as internal documents, sensitive documentation, or ads are generally not enabled for caching, for privacy reasons or to avoid unnecessary storing (Harnish, 2020).

be measured though Time to first byte (TTFB), another Web Vital[37] which calculates the time it takes for the browser to receive the first byte of data from the server (Pandya, 2024). In addition to improved load speed, CDNs are also worth implementing because they provide a more secure connection: since the same resources are hosted on multiple servers instead of just one, if one of them is not responsive, the browser will connect to the next and still manage to deliver the content to the user. This is a significant aid for example in case of a sudden increase in traffic, which would result in an overload for the server and therefore impact the website's performance (Toonen, 2018b).

5. Off-page optimization

Gaining visibility is a goal which, in order to be achieved, requires efforts in all areas of relevance of a website. This includes, as we have seen, creating high quality content and formatting it in such a way that it is easily understandable by both people and search engines, organizing it using a logical structure that is easy to navigate and optimally interconnected, and optimizing the technical performance to reduce potential attrition. All these actions fall under the direct control of website owners, who, according to their ability and to the ability of their competitors, can manipulate their sites' performance and appear as more or less valuable to Google. A website's value, however, does not only consist of keywords, sitemaps, or codes. As highlighted by Google's E-E-A-T guidelines, authoritativeness is of the utmost importance in assessing whether a resource is reliable – thus worthy of being suggested to users – or not.

Reputation is not something one can easily build: winning the trust of others takes time, and largely depends on factors that are not as easily controllable as implementing a tag or resizing a picture. Off-page SEO encompasses all those traffic-boosting techniques which take place externally to the website itself and more broadly concern the image of the brand. Some of these techniques directly affect ranking, as their output is taken into account by Google's algorithm, whereas others work as

37 Web Vitals are a bigger set of quality signals to which Core Web Vitals belong; all of these metrics are useful and indicative of a website performance, although only the Core vitals are actually used by Google as ranking signals (source: Google).

promotional activities and benefit a site more indirectly – but nonetheless still significantly, generating interest in the brand and affecting Google's Quality Raters' evaluation[38].

5.1 Link building

One of the main objectives of off-page optimization consists in enhancing one's own reputation by showing Google that other websites find that content so reliable to the point of implementing it on their own pages. To do so, webmasters make use of links. As it has been discussed in Section 4.3, links play a central role in helping the search engine understand a page better, and the more links point to a page, the more important that page is within the information architecture of a website. Similarly, backlinks – clickable links from one website to another – help assess the importance of a certain page by passing on link equity, not within one same domain as internal links do but rather across the web. Google interprets a backlink as a sort of vote through which a website vouches for another website. This means that receiving many links will increase the authority score of a page, and even more so if the mention comes from a website that is authoritative in its industry. An authoritative website itself receives, indeed, a lot of reputable links (Soulo, 2023).

The set of practices SEO masters adopt to encourage other sites to cite theirs, is called link building. These strategies serve the purpose of boosting ranking score by building one the the most powerful ranking signals: authority. There are various possible link building activities, which span from more to less ethical and punishable. Manually adding links to other people's pages affected by a security flaw, or paying other webmasters to get linked on their site, are, for example, among the illicit techniques webmasters can use. Because building brand awareness and reputation is a long, difficult task, it might be tempting to take these kinds of shortcuts. Although Google cannot easily detect them, these kinds of tactics are deemed as manipulation by the search engine, and represent a risk of being penalized.

38 Quality Raters' so-called "Reputation Research" includes looking at online reviews, recommendations from experts, mentions on authority websites (Dean, 2023b).

Spontaneously obtaining links is the purest form of link building – and arguably the ultimate goal of any online business. If somebody freely decides to mention a site, it means that the brand is strong enough to spark genuine and explicit interest, and while it is always essential to offer valuable content, earning links requires material that is truly noteworthy, and that can inspire creators to implement it on their own pages. SEO masters can use a number of methods to stimulate brand awareness, reputation, and authority to ultimately gain mentions, which do not necessarily involve SEO activities. In some cases, the quality of the content is so high or the authoritativeness of the brand so strong to earn the website links organically. Launching a new product, for example, naturally attracts attention, thus links. Precisely for this potential, this is often referred to as "link bait". Link baiting is the creation of extremely linkable assets that will most likely be mentioned by other sources due to its high value – such as infographics, free tools, also insightful articles, original research, provocative opinions – while at the same time are difficult for competitors to replicate (Dearie, 2023). Captivating and useful content is extremely effective because they serve both the audience of the websites that publish them but also that of the website linking to it, and it is an easily scalable technique, meaning that one single piece of content can then be shared by more and more editors, and reach more and more people.

As previously mentioned, however, convincing others to spontaneously mention a website is all but easy and straightforward, and it is but a small share of domains that can rely on earned links to increase traffic. Whereas in some cases, the content itself is so valuable that it essentially promotes itself, much more frequently and realistically, web creators need to find a way to promote their content and their brand and proactively seek endorsements. The concept of outreach consists in pitching one's content to convince an editor to use it on his page, with the goal of increasing awareness and traffic. A good pitch highlights how such collaboration would benefit the endorsers too, and the success of it largely depends on the choice of the recipient: when contacting potential prospects, it is important to analyze and match the industry, the target, and the quality of their business. For a sportswear retailer, it would be pointless to reach out to a marketing blog, for instance, as the audience of the latter would not be interested in an

article or a service of the former; conversely, such material would fit well on the website of a gym franchise, provided that its site is authoritative enough to offer qualitative backlinks (Lahey, 2023). To find valuable potential contacts, SEO practitioners can make use of a number of tools – Semrush's Backlink Analytics, Moz's Link Explorer, Ahrefs' Link Intersect, to name a few – which detect possible candidates on the basis of a competitors analysis: by inspecting backlink profiles across various websites, not only is it possible to assess one's performance compared to the other actors in the sectors, but it also provides useful information to understand their strategy, and target the website they themselves have targeted. Indeed, if a certain website agreed to cite a car rental company, this sets a precedent for other similar companies to successfully pitch their brand and it is more likely that, if the content offered is compelling, the site will accept the proposal. This analysis is useful to conduct a much more efficient and effective outreach campaign, pursuing links which are both easy to replicate and relevant (Chetvertak, 2024).

Given that link baiting is the ideal link building technique – as spontaneous as it is hard to achieve –, SEO practitioner often reach out to prospective endorsers by essentially outclassing competitors, that mean, by taking a piece of content, crate a better version of ot and pitching it to those website who linked the original in the first place: this is the so-called skyscraper technique. The core idea behind this concept is that people – and therefore, websites – want to see the best content they possibly can find. Based on this assumption, web creators try to find existing popular content and to improve it by offering something superior – metaphorically, an even taller skyscraper. The ultimate goal of this is to replicate the backlinks by letting the pages which mention that their content is better (Ong, 2023). In this case, too, analyzing the competition is essential to see what is already being offered and to understand how to improve it, for example making the resource more comprehensive, more in-depth, or to present it more effectively or more compellingly. If a competitor is offering a guide to the then best vegan restaurants in Bruxelles which is dated 2021, then it might be strategic to create an updated 15-item list, enriched with captivating visuals or a more personal touch. A similar concept lies behind what is called "resource page link building", that is, when

webmasters try to get featured on a page that lists useful resources within a certain industry, for example on directories such as the Yellow Pages; in this case too, the appeal for the editor is to offer fresh and up-to-date content to its audience (Hardwick, 2023).

Extremely relevant in link building is guest blogging, which, despite being relatively more challenging, is one of most widespread tactics used by creators. This practice consists in writing and publishing a blog post on somebody else's website, and receiving, in turn, credit – and links – for contributing. In this case, convincing the host site to endorse the piece of content requires more hard work, as it implies a bigger "intrusion" in his online space; however, this also offers him the opportunity to leverage the expertise and the experience of guests, which may be specialists in their niche. For the author, the chance of writing for another blog means gaining reputation and authority, giving more exposure to his brand and attracting more targeted traffic (Ong & Hardwick, 2023). Although being considered a part of link building, strictly speaking guest blogging does not actually bring any real backlinks to the website. This is because, since the author may receive monetary compensation from the host blog – especially in the case of very reputable and authoritative authors –, and paying for links is considered a violation of Google's guidelines, sponsored links such as guest posts are not taken into account when it comes to link equity. In the page code, guest posts links are marked with a **`rel="sponsored"`** attribute, because their endorsement is not impartial, and therefore they are not counted as links by the search engine. In fact, the search engine should be able to intercept this kind of content on its own, and automatically apply the tag; nevertheless, it is advisable to manually insert it in order to avoid incurring any penalties. This implementation does not affect user experience, as the link still works as usual for users, the only difference being that it is not taken into account for SEO purposes. This does not mean, however, that it is not relevant for SEO; in fact, despite not transferring any authority, `"nofollow"` links still drive referral traffic and contribute to creating a more diverse and natural backlink profile. This is particularly important for search engines as it demonstrates a more organic growth

pattern: a website with only "follow" links[39] might appear manipulated and trigger some kind penalization by Google (Pavlik, 2024b). Furthermore, disclosing a paid placement. is becoming a legal requirement in more and more countries for transparency reasons towards the reader.

It can sometimes happen that, despite mentioning it, a website fails to properly link to the source; as a consequence, while the brand *is* mentioned, that mention does not bring any link equity nor referral traffic to the website. This poses a problem for that particular site, but also, on the other hand, creates an opportunity for others to easily gain backlinks. In the first case, if website owners find themselves victim of improper linking, they can request that their mentions be fixed – in case the mistake is on the side of the endorser – or redirected – if the link itself does not work anymore, which is the case for broken pages. Broken pages, or 404 pages[40] can occur for a number of reasons that make the requested resource not retrievable, for example if the website owner has deleted the resource, thus it is not available on the server anymore, or if the URL was changed; this is particularly undesirable, because it means that interested users who clicked on the link cannot visit the page; this hurts user experience, potentially damages brand image, and invalidates all efforts to create and promote the content, since a broken link cannot rank on the SERP (Tan, 2024). When broken pages are spotted by competitors, this gives them the possibility of capitalizing on this error. Broken link building is a tactic which takes advantage of broken pages pointing to a others' sites by offering endorsers an alternative – and most importantly, functioning - link to mention; as a matter of fact, most websites tend to accept and add the proposed new link, since they want to provide their audience with valuable content, and want to remove non-functioning hyperlinks which can definitely undermine their intention. It is crucial, however, to offer a piece of content – either already existing or custom-created – which fits well in the vacuum left by the broken link: this does not mean copying the original,

39 The term "follow links" does not technically exist, and is only used in this context in contrast to "nofollow links" for the sake of clarity.

40 The name comes from the typical "404 Not Found" error status for pages that cannot be found due to the fact that content has been removed or moved to another URL.

but rather creating a close but improved replacement[41]. Broken linking building can have a higher success rate than other link building tactics, as it is a fundamentally mutually beneficial relationship: it helps solve an issue by offering something in return. This is why it is also important to apply this approach internally too, and ensure all the current backlinks are up to date and working (Dean, 2023c).

So far, the great value of links has been extensively explained how valuable links are for a website's authority score, and how the more mentions, the higher the ranking. Links, however, can also be detrimental for a website's reputation: being mentioned by low-quality pages – that is, spam pages or not relevant to the context – can result in a penalization by Google, which interprets it as a manipulative tactic, despite the fact that the linked page is not actually involved in it: any page, can, in fact, mentioned whomever it wants. Illicit backlinking actions are usually effectively ignored by the search engine, but if large in volume, low-quality backlinks can trigger a penalization. To prevent this, webmasters can signal such links to Google to help it determine which ones are to consider, thus effectively neutralizing them. In Section 4.4 we have discussed the `rel="nofollow"` function, which is a valid way to solve this issue; additionally, Google provides a specific tool called Google Disavow Tool that reads a .txt file collecting all the domains[42] producing harmful backlinks. Disavowing links is a fairly extreme action, which is generally not necessary as the search engine is equipped with sophisticated enough algorithms that are able to spot low-quality URLs; in fact, Google itself advised against it, as it can significantly impact ranking, and should be implemented unless the link has caused serious disruption. A low-authority site is not necessarily a bad link: presumably, it will not contribute meaningfully equity-wise, but it is unlikely that it will cause any harm. Every link, regardless of how little its value is, is seen as a vote of confidence by Google, and should not be wasted (Davydov, 2023).

41 Although the originally linked page might not exist anymore, it can still be viewed on internet databases such as Internet Archive's Wayback Machine; this is useful to get an idea of what kind of content to create when trying to offer a replacement for the broken link.

42 When disavowing links, it is more efficient to disable entire domains rather than individual URls, since all backlinks coming from the same website are most likely all low-quality links (Grubbs, 2022).

5.2 Digital PR

Off-page optimization encompasses all those practices that more loosely relate to online promotion. Similarly to traditional Public Relations, in the digital environment brands use Digital PR to increase brand awareness and reputation. It is a broad concept that is often used interchangeably with backlink building, but that, in fact, encompasses link building strategies and other online marketing techniques and combines to earn media coverage. Although many PR campaigns do lead to acquiring backlinks, the unlinked placement still has valuable brand awareness benefits, and can lead to third-parties to publish articles that do link. Launching a new product or creating linkable assets can indubitably generate interest around a brand, but in order to attract significant actions, it is often necessary to encourage this sentiment with self-promotion. Curiosity around the brand can the be supported by promotional actions such as advertising, which can produce quick – though short-termed – results and effectively boost traffic, but also sponsoring events, appearing in videos or podcasts or creating entertainment material, for instance; this does not directly lead to an increase in clicks and visits, but contributes to build overall prestige and prominence (Ong, 2024). We could say that, if link building is meant to make Google think that a website can be trusted, PR are meant to make users think that too.

Similarly to what has been previously said about outreach, a PR pitch involves building relationships with journalists, creating a network of contacts, and getting them to write about one's brand, for example about the launch of a new product, a promotion or a specific achievement. This might, however, be difficult to arrange, because it takes something innovative and relevant to grab the spotlight, and it is not always easy to know what the media will be interested in. An alternative to this is by responding to journalists' call: reactive PR is a practice which involves offering one's expertise about a subject to journalists who request it and will use it as source. Reactive PR relies on dedicated platforms which connect experts and authors to facilitate these collaborations – one of the most renowned ones is HARO, Help a Reporter Out, although it is getting replaced by other, less saturated alternatives (Abouobaia, 2022)

Among the many tactics used in digital PR, one particularly relevant asset are online communities: being present on social media is nowadays pivotal to establish one's space online, where the brand can interact directly with its audience. Since links on social media are not earned but rather self-promotional, paid sponsorships, or user-generated, they are counted as "no follow" by the search engine; nonetheless, implementing social media content in one's SEO strategy will promote the brand and contribute to generating more traffic. More precisely, despite falling outside of SEO, Social Media and Influencer Marketing can contribute to organically boost traffic via a specific kind of searches called branded searches. Branded search concerns queries that specifically contain the name of the brand or of one of its products or services, for instance "Nike tennis shoes" or "Air Max shoes" as opposed to "sport shoes" (Tan, 2023). These searches are necessarily performed by users that have already gained at least some knowledge of the brand, and are an important marketing achievement because they not only drive traffic directly to the website, they also reflect a strong intention on the user's end, which know exactly what he is search for and is most likely ready to convert[43].

43 Branded search can be considered a subtype of navigational queries, given the high level of specificity and the strong intent.

CHAPTER 4: INTERNATIONAL SEO AND WEB LOCALIZATION

The growing importance of the Internet has drastically changed the landscape in which companies operate in a number of ways, effectively revolutionizing how business is conducted. It has provided marketers with endless new opportunities to reach their audiences, one of which is, indeed, the subject of the present work. Yet another major development which derived from this momentous invention, affecting virtually all aspects of life, is globalization. The World Wide Web holds in its name its global dimension; it has built bridges that can connect people standing on opposite sides of the world, reducing distances and making anything accessible to anybody, anywhere. This has changed the perspective individuals have of themselves, others and the space around them, and human organizations have been affected just as much. Globalization is a huge opportunity for many businesses: as multinational corporations become more and more prominent – this, too, as a result of the globalizing trend –, expanding into new markets has become a requirement for many companies in order to stay competitive and prosper; as we have seen in the previous chapters, in a crowded space as today's web, there is little room for error.

International SEO is a branch of SEO that, building on on-page and off-page optimization, focuses the adaptation of a website to one or more foreign target audiences[44]. SEO is, in its essence, heavily dependent on the resource it aims to optimize; let us think of how different the process and outcome of keyword research might be for a small florist or a large company that sells cars, even if they both operate in Bruxelles. When dealing with entirely different regions, the differences in culture, customs, and consumption models can be tremendous, and must be carefully considered when creating and optimizing high quality content. International SEO encompasses various practices meant to manage the online presence of a business in different markets; in fact, its practices can vary greatly from case to case, as internationalization strategies change according to each individual business's goals. Conducting thorough

44 The very notion of "foreign" is, of course, relative to the object of reference; here, it is intended as foreign to the home market of the individual brand.

market research is essential before embarking on such a journey: based on the data collected, informed decisions should be made regarding which markets are worthy of investment. If significant traffic is coming from a different location to where the business is based, or from foreign-speaking users, then it might be wise to pursue it. At the same time, however, website owners must be aware that expanding one's business does not simply mean translating a few pages or adding new shipping options: internationalization implies creating content that is tailored onto the new audience, and being able to maintaining it over the course of time, but it also means providing the necessary support to customers, keeping in mind language, customs, time zones (Rudzki, 2019).

When talking about website localization, we can distinguish between two types of adaptation: if the target segment is defined by a certain language, then we talk about language targeting, the most obvious and common example being targeting all English-speaking countries; conversely, if the target segment coincides with a specific country, then we talk about country targeting, or geotargeting. The distinction between language targeting and geotargeting marks the difference between multilingual and multiregional websites: multilingual websites are those websites that are available in multiple languages, such as a Belgian website that is available both in Dutch and French. Multiregional websites are those that are accessible from various locations due to the fact that the local language is the same, such as a British website that features content in English for both the United Kingdom and Ireland[45]. Which strategy to implement depends on the goals and resources of each individual business: websites can be both multilingual and multiregional if they address different countries and different languages at the same time. A Canadian company, for instance, can target multilingual audiences if it offers an English version and a French version of it website to English-speaking Canadians and French-speaking Canadians respectively, and can target multilingual audiences if, additionally to the Canadian market, it targets customers in the United States with a dedicated website (Morris, 2016).

45 Source: Google Search Central.

The key factor that makes international SEO so powerful is that it can dramatically expand the area of influence of a brand. As it has already been profusely discussed, search engines work to provide the most relevant and suited results to each user; this means that offering tailored content is pivotal to try to rank competitively on the SERP. In an international context, this tailoring process includes not only needs and desires, but also linguistic, cultural, and geographical factors, all of which are taken into account to ensure the best user experience possible. When managing an international business, the main goal is to appeal to the right audience: whether one single site targets all Spanish-speaking users or Chilean and Mexican users separately, it is fundamental that the search engine understands which version ought to be shown to whom. To do this, Google relies on a number of signal: these can be summarized in technical signals – which encompass the formal structure of multilingual, multiregional websites –; content signals – which derive from content –; and off-page signals – that is, backlinks coming from other websites operating in the same market and/or in the same language. It is within these three pillars that international SEO operates (Rudzki, 2019).

1. Technical localization signals

The first set of implementations required to launch an international version of a website concerns the structure of said website. In addition to all other fundamental concepts in SEO, such as information architecture, internal linking, and page speed, multilingual and multiregional websites need to be designed so as to clearly indicate what audience they are serving. As we have seen in Chapter Three, URLs are a web page's ID, which states the website's domain and indicates all the other specific directions to locate the page on the web. When it comes to international businesses, Google uses URLs to decode for whom each version of the website is meant. To facilitate its job, webmasters can use a variety of domain structures, each with its advantages and disadvantages, and can fundamentally be divided into multiple-domain strategies and one-domain strategies – where a single website holds multiple versions of the content.

A multiple-domain strategy involves using multiple individual websites – one for each market; each version of the original website has its own domain which includes a two-letter code to indicate the specific country of language that is being targeted. Amazon.it and Amazon.es, for example, form two entirely different websites, which are marked by so-called country code top-level domains (ccTLDs) – .it and .es, respectively. These codes are used by Google to localize the websites, that is, to associate them with the countries those TLDs are associated with, and consequently show them in the SERP for users navigating from that country[46]. Despite the country code being a strong geotargeting signal for Google, ccTLDs show a number of negative sides. Compared to others, this approach tends to be more expensive, both in terms of budget – as it is necessary to acquire and manage multiple domains – as well as effort: indeed, each alternate URL needs to be established and marketed separately to rank well in the new market, because the new domain does not inherit equity from the original one, thus its authoritativeness has to be built from scratch. Additionally, ccTLDs can only be used to target countries but not languages and their purchase is subject to restriction[47]

A cheaper and more practical alternative are multiple subdomains under a single global domain, for example en.wikipedia.org and es.wikipedia.org. In this case, the country – or language – code precedes the seconds-level domain, and despite being technically registered under one single generic top-level domain, each alternate can be managed as separate websites. This implies higher ease of use, but on the other hand, it could also cause confusion on the user's part, who might not be sure about whether the site is intended for him; therefore he might not feel appropriately targeted and less trusting of the brand (Morris, 2016).

46 Some ccTLDs are treated by Google as gTLDs, which means that using one of these domains could negatively affect the website's ability to rank in the intended country. This reclassification is due to the fact that these domains are used for purposes beyond their original country-specific intentions – for example, ai (Anguilla) for AI-related websites, or .tv (Tuvalu) for television-related content. By treating these ccTLDs as gTLDs, Google can ensure that the search engine results are more relevant to users' queries, meeting their expectations and reflecting the actual use and target audience of the websites using these domains. At the same time, the consequence for SEO is that using one of these domains could negatively affect the website's ability to rank in the intended country (Southern, 2023).

47 There may be a range of policies and restrictions when registering specific ccTLDs. Some of these include residency requirements and particular renewal conditions.

One of the most common options for international websites is grouping all international versions under one global domain while using subdirectories to differentiate the various versions, such as in Netflix.com/de or Netflix.com/fr, for instance. In this case, the generic TLD is further specified so as to target different audiences based on country or language. This one-domain strategy is less expensive, more practical and manageable, although the targeting might not be immediately apparent to the user; in fact, gTLD with subdirectories is less strong of a signal to search engines than a ccTLD (*ibidem*).

Based upon the information contained in the URL, Google tries to predict the precise target of a website by matching the URL's directions to the IP address of the user server. As previously mentioned, domains are only one of the signals taken into account by the search engine, but they are indeed of significant relevance among technical signals. Another very important element is "hreflang", arguably the most powerful tool for language targeting. The hreflang attribute is a metatag that specifies which language – and, optionally, the region to specify a regional variety– is used on the website, so that search engines can display the result to users searching in that particular language. It can be included in the HTML code, the HTTP header, or on the sitemap of a website, together with all the different website alternates, so as to communicate to Google that those are localized versions of the site and therefore avoid duplication issues[48]. This is particularly relevant when the same content is addresses to different targets, as is the case for versions intended for different countries that speak the same language – for instance,

```
<link rel="alternate" href="https://www.adidas.be/fr"
hreflang="en-AU">
```

and

```
<link rel="alternate" href="https://www.adidas.be/fr"
hreflang="en-BE">.
```

48 The current alternate is signaled on the HTML code as "canonical", while all the other versions as "alternate".

As we can see, the tag generally includes a language code and a country code; while the former is always required – since it is necessary to, indeed, language target the website – the latter is optional and can be used to specify the language variant of a country – in the example, Australian English and British English, which, despite being fundamentally the same language, might differ for word use or certain meanings (Morris, 2016)[49].

```
<link rel="canonical" href="https://www.adidas.it">
<link rel="alternate" href="https://www.adidas.com.ar" hreflang="es-AR">
<link rel="alternate" href="https://www.adidas.at" hreflang="de-AT">
<link rel="alternate" href="https://www.adidas.com.au" hreflang="en-AU">
<link rel="alternate" href="https://www.adidas.be/en" hreflang="en-BE">
<link rel="alternate" href="https://www.adidas.be/fr" hreflang="fr-BE">
<link rel="alternate" href="https://www.adidas.be/nl" hreflang="nl-BE">
<link rel="alternate" href="https://www.adidas.com.br" hreflang="pt-BR">
<link rel="alternate" href="https://www.adidas.ca/en" hreflang="en-CA">
<link rel="alternate" href="https://www.adidas.ca/fr" hreflang="fr-CA">
<link rel="alternate" href="https://www.adidas.ch/de" hreflang="de-CH">
<link rel="alternate" href="https://www.adidas.ch/en" hreflang="en-CH">
<link rel="alternate" href="https://www.adidas.ch/fr" hreflang="fr-CH">
<link rel="alternate" href="https://www.adidas.ch/it" hreflang="it-CH">
<link rel="alternate" href="https://www.adidas.cl" hreflang="es-CL">
<link rel="alternate" href="https://www.adidas.co" hreflang="es-CO">
<link rel="alternate" href="https://www.adidas.cz" hreflang="cs-CZ">
<link rel="alternate" href="https://www.adidas.de" hreflang="de-DE">
<link rel="alternate" href="https://www.adidas.de/en" hreflang="en-DE">
<link rel="alternate" href="https://www.adidas.dk" hreflang="da-DK">
<link rel="alternate" href="https://www.adidas.es" hreflang="es-ES">
<link rel="alternate" href="https://www.adidas.fi" hreflang="en-FI">
<link rel="alternate" href="https://www.adidas.fr" hreflang="fr-FR">
<link rel="alternate" href="https://www.adidas.gr" hreflang="el-GR">
<link rel="alternate" href="https://www.adidas.ie" hreflang="en-IE">
<link rel="alternate" href="https://www.adidas.co.in" hreflang="en-IN">
<link rel="alternate" href="https://www.adidas.it" hreflang="it-IT">
<link rel="alternate" href="https://www.adidas.co.kr" hreflang="ko-KR">
<link rel="alternate" href="https://www.adidas.mx" hreflang="es-MX">
<link rel="alternate" href="https://www.adidas.com.my/en" hreflang="en-MY">
<link rel="alternate" href="https://www.adidas.nl" hreflang="nl-NL">
<link rel="alternate" href="https://www.adidas.no" hreflang="no-NO">
<link rel="alternate" href="https://www.adidas.co.nz" hreflang="en-NZ">
<link rel="alternate" href="https://www.adidas.pe" hreflang="es-PE">
```

Figure 43: Example of <link rel="alternate"> in an HTML code. Screenshot taken from: https://www.adidas.it/

Including hreflang tags for all URLs on each localized website creates a net of interconnected page variants and helps the search engine understand that they are related to each other but directed to different audiences, and it facilitates discoverability by web crawlers, thus indexation. This is extremely important, because if a page is missing return links, Google will ignore the one-sided tag and consider the website as unrelated. This requirement was introduced because having a set of mutually linking,

49 The language code must comply with the ISO 639-1 format, and the country code must comply with the 3166-1 Alpha 2 format (source: Google Search Central).

multidirectional URLs prevents unauthorized people from arbitrarily creating a website and presenting it as a localized version of somebody else's site by linking to it.

Additionally to country-specific metatags, a generic hreflang tag – for instance, `<link rel="alternate" hreflang="x-default" href="https://adidas.com/" />` – is usually included to provide a "default", generic alternate intended for international users, those who are not located in any of the other targeted countries. X-default tags are generally used on language selector pages, that is, pages where users can choose and select their language of preference before entering the website to be redirected to the relevant page. X-Default hreflang are not a requirement for multilingual websites, but that can be a strategic implementation to expand market reach by setting a website for a global audience and attract new customers who have not been specifically targeted yet (Wruck, 2018).

A third fundamental tool in the international SEO domain is geotargeting. This implementation is instrumental to address a specific website variant to a specific country. By explicitly selecting which regions a specific website version targets, this will be shown exclusively to users coming from that region. The reason behind this is relevance: the Spanish alternate of adidas.com – which shows prices in euros and lists stores located in Spain – will be relevant to users navigating from Spain only (Rudzki, 2019). The difference between language targeting and geotargeting not only lies in the targeting criteria; in fact, with geotargeting Google adjusts ranking according to the user's location; this means for which a website with a high ranking might be taken over by a competitor with lower ranking but more locally relevant than its competitor. In a way, Google promotes the geolocalized website. On the other hand, however, it is not possible to geotarget across countries or regions, as geolocalization only allows to target one country and primarily addresses visitors located elsewhere, which could potentially lead to an opportunity loss of traffic coming from other countries. For this reason, it is advisable not to configure it if one single website is meant to target multiple locations – Spain and Mexico, for example. Geo-targeting is, furthermore, often accompanied by highly localized content: with language targeting the content is merely translated, and

hreflang is used to signal the fact that it is, indeed, a translation or a slight adaptation, and avoid duplication issues.

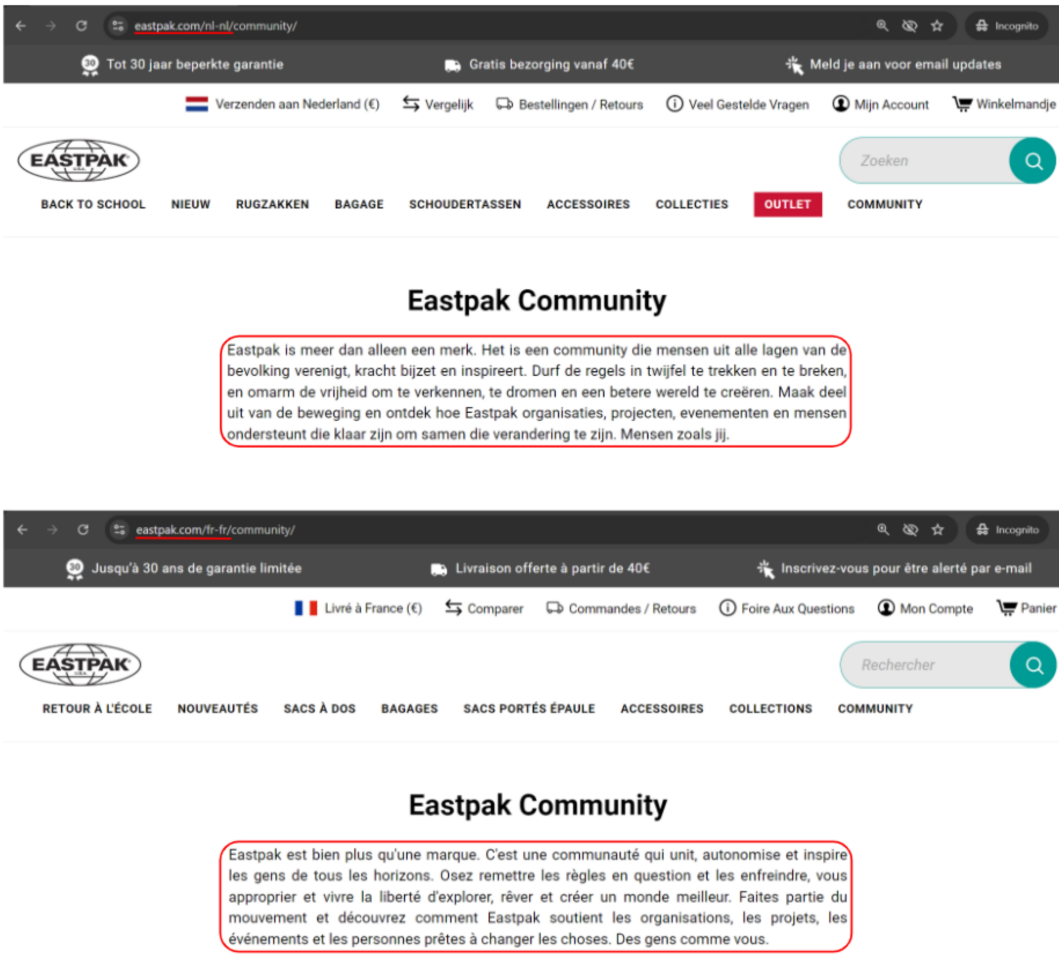


Figure 44: In the top picture, the community page on the Dutch (Netherlands) version of the website; in the bottom page, the French (France) version; as we can see from the two URLs, the two alternates are subdirectories of the main website, which has been tagged with hreflang, and the text has been accurately translated. Screenshots taken from: https://www.eastpak.com

Conversely, with geotargeting, the content is ideally created from scratch, to offer material that is country specific (Wruck, 2018). This marks the distinction between translation and localization, which will be discussed in Section 3.

Figure 45: In the top picture, the sustainability page on the Italian website; in the bottom page, the Spanish website; as we can see from the two URLs, each website is independent from the other, and the material has not been translated but replaced completely. Screenshots taken from: https://www.mcdonalds.it and https://www.mcdonalds.es

URLs, hreflang, and geotargeting are all part of the signals that search engines use to deliver the most relevant results to users based on their geographic and linguistic preferences; however, they are not directives: this means that each of them gives a cue to Google about the website, but not an explicit command. It is ultimately Google's decision whether to accept the suggestion; none of these elements by itself is sufficient to correctly match content to the intended audience, but rather they must be designed and integrated so as to maximize the chances of reaching the target (McMichael, 2016).

1.1 Geotargeting for local PPC

Geotargeting is an extremely powerful tool to restrict searchability on the part of non-targeted users and thus reach those who are more likely to become prospective customers. Given its potential, this opportunity for increased efficiency is used in online advertisement to target very specific customer segments, and concentrate effort and resources on more promising leads. Major brands, who dispose of larger budgets, might decide to run a marketing campaign worldwide, but much more frequently ads and promotions are limited to a certain area – a region, a country, or even a city. This practice/technique/strategy is often called local PPC, and consists in addressing highly personalized ads to local customers. As discussed in Chapter two, online advertisement – and more specifically PPC – tends to be very expensive; ad prices vary according to keyword difficulty and competition, but the actual cost depends, indeed, on how many clicks each banner gets. Consequently, manipulating ad placement so that it is shown exclusively to those users whose click might be fruitful and lead to a conversion is pivotal in the strategic management of SEA. On the contrary, if geotargeting is not configured on the advertising platform, any user anywhere might see an ad not intended for him, and if someone clicks on it without converting, this will cost the advertiser precious resources without offering any kind of return. A clothing brand shipping exclusively to Ireland does not have any interest – and, possibly, no budget to waste – in advertising its products and services to users navigating from South Africa, and would much better geotarget local prospects[50].

In Google, geotargeting for advertisement can be configured through Google Ads, Google's advertising platform, which allows to set one or multiple locations where to run the PPC campaign, or, on the contrary, which locations to exclude. These locations can encompass entire regions, countries, individual cities, or even areas within a city, and, depending on the IP address of their devices, users can or cannot qualify as potential customers and see the ad. This targeting should, of course, be paired with highly localized copies: high specificity of the ad means higher chances of conversions, higher return on the investment (ROI) and lower cost-per-click (CPC) (Voltina, 2018).

50 Source: WordStream.com/geo-targeting

Additionally, Google Ads enables to incorporate so-called location extensions, additional assets which provide additional information such as location (location extension) and phone number (call extension) and to ad-viewers for the highest quality experience possible (Irvine, 201). Local PCC is particularly relevant for local businesses, who serve customers in a specific area. Proximity and personalization tend to drive customers to interact more with geotargeted content, increasing engagement and brand loyalty. These signals are, in turn, valued positively by the search engine and can improve local rankings (Saxon, 2024).

2. Geo-redirection and UX aspects of localization

In the pursuit of enhancing user experience and meeting diverse needs, it might be tempting to automatically present individual users with the most suitable version of your website. Having an international website means having to deal with an international audience, made of users located in various countries and speaking different languages, and it may happen that Google presents the users with the wrong version. Geo-redirecting is a method used to automatically direct website visitors to specific versions of a site based on their geographic location. This is typically done by identifying the user's location through their IP address, GPS data, browser's language and region settings or other methods. Pages which implement this technique are called locale-adaptive pages, that is, pages that return different content depending on the perceived country or preferred language of the visitor. A Brazilian user navigating from India would, for example, be redirected to the Indian version, in spite of the fact that he clicked on the Brazilian site on the SERP.

Geo-redirecting international customers, while intended to provide a localized experience, can create significant issues for both users and businesses. On the users' part, automatic redirection can cause confusion and frustration due to the fact that the customer feels he is not in control of its own online journey. The language-country preference detected might not be accurate, and the redirected site might not offer the same products, services, or content as the site they originally intended to visit.

Additionally, user experience might suffer from longer load time, due to all these redirects. On the other hand, redirecting server requests according to the device's IP also means redirecting spiders trying to access and crawl the website. In Google, this might lead to issues with indexation: Googlebot crawls the Internet from location all over the world, but mainly and more traditionally it is based in the United States, which means that it is more likely that a site will be crawled as it is available to users in the US: Google will be redirected away from any non-US content[51]. For locale-adaptive pages, this might result in having only one version – en-us – properly indexed, which consequently affects ranking[52]. To prevent this from happening, it is advisable to let the user explicitly choose which website version to visit. When correctly implemented, hreflang annotation should eliminate the need for automatic redirection; generally speaking, a Moroccan user will navigate from Morocco and will click on the Moroccan version, which should rank higher on the SERP than any other alternate. Even in cases like this, however, where the user's preferences might be rather straight-forward, it is possible that he wishes to view content in another language, and it is therefore important to allow him to access other versions, either by accessing them directly from the SERP, or by selecting another language and country on the homepage from the navigation menu.

51 It is possible to enable geo-based auto-redirection and, at the same time, disable it for Googlebot using robot.txt files; however, Google strongly advises against it as, in order to properly serve their function, bots should be treated as regular, human users.

52 Source: Google Search Console.

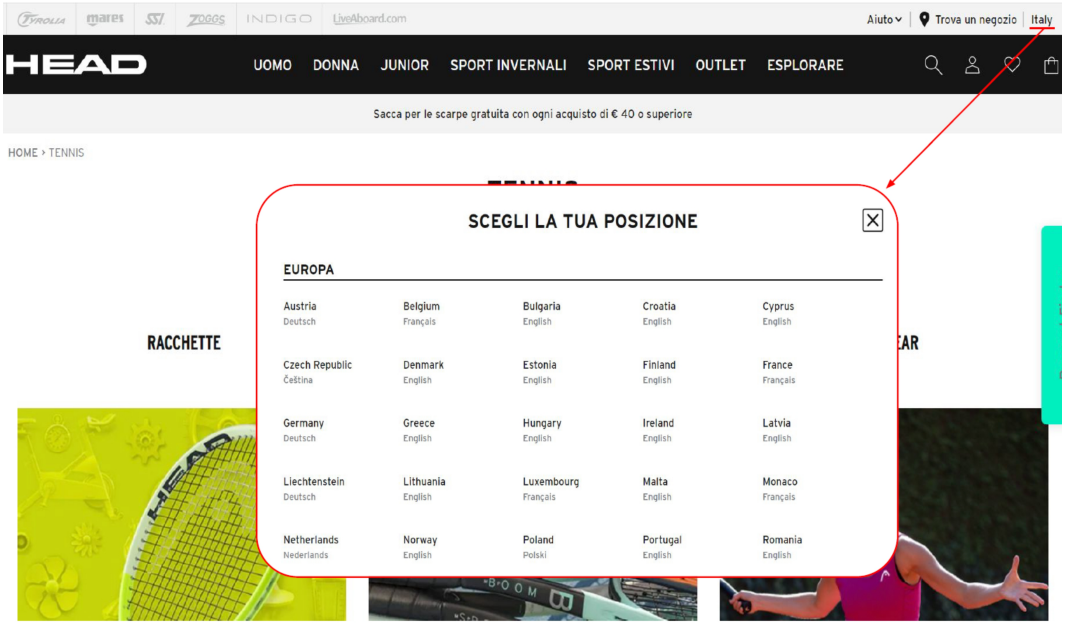


Figure 46: Example of language menu. Screenshot taken from: https://www.head.com/it_IT/tennis.html

When the site detects user may find another version more relevant, it is also possible to suggest the user to move to the correct one for his location using a non-intrusive pop-up message to prompt him, so that, if he inadvertently clicks on the wrong link, he will be able to seamlessly switch to the right page; therefore, he will be less likely to bounce back and abandon the site and will instead stay and continue navigating (Grand & Gavoyannis, 2023).

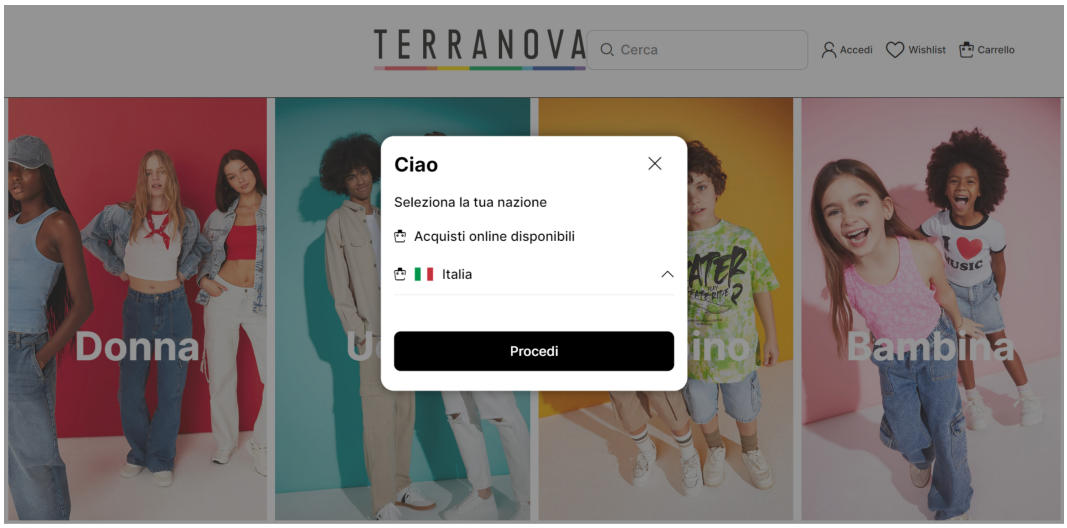


Figure 47: Example of pop-up language-country selector. Screenshot taken from: https://www.terranovastyle.com/it_it/

Both these techniques increase retention and enhance user experience, and can be stored in the cookies, text files where information about users' browsing session are recorded, such as login details, or language-country preferences (McMichael, 2016).

3. Content localization: adapting content to a foreign market

As it has been profusely discussed, user experience stands at the core of – more or less directly – all SEO practices, from understanding search intent to creating valuable content, to reducing loading speed. When taken to an international level, the layers of difficulty multiply: serving multiple audiences can bring great advantages in terms of market share, and is an essential step for global expansion. On the other hand, however, it increases management complexity and costs. One of the areas that is often – and can easily be – overlooked is content localization. Content localization for international SEO involves adapting your content to meet the linguistic, cultural, and regional preferences of your target audience in different markets. It is more than just translating text; it requires a deep understanding of local customs, idioms, and consumer behavior to ensure relevance and engagement. This process includes modifying keywords to align with local search terms, adjusting cultural references to resonate with local audiences, and ensuring that the content adheres to regional regulations and norms. By tailoring your content in this way, you enhance user experience and improve your chances of ranking higher in local search engine results, ultimately driving more targeted traffic and fostering stronger connections with international users.

The first, preparatory step towards localizing a website is selecting for which audience to localize. When embarking on an internationalization journey, businesses must evaluate and identify those markets which – from a business perspective – are worth expanding to. Expansion costs vary greatly from one business type to the other, with e-commerce websites having to face higher investment costs than purely editorial websites due to sales, legal and shipping costs, and targeting a market in which there is no demand for the offered service can result in a major failure. To this end, Google

provides a strategic market evaluation tool, Google Market Finder, which allows webmasters to check if and how a certain product is being searched in a specific country or language Alternatively, Google Analytics also provide useful insight on where visitors of a website are located: intuitively, a large stream of traffic coming from Saudi Arabia might signal a valuable opportunity to create a localized version in Arabic (Arosio, 2020). The insight gained from these analyses is key in deciding which new audiences to approach to ensure profitable returns: in fact, it is useless and wasteful to try and land on markets without any prospects of growth. It is worth noticing that not all markets are the same; in fact, demand can vary greatly depending on culture, values, and customs. Likewise, offer can vary accordingly: products, services, or content do not necessarily have to be the same across countries, as they are designed to respond to individual desires and needs (Rudzki, 2019). More often than not, however, modern businesses comply with the requirements set by globalization trends, and tend to standardize their offer; in this case, their products are differentiated in terms of marketing.

3.1 The concept of localization

According to LISA (Localization Industry Standards Association)[53], "localization involves taking a product and making it linguistically and culturally appropriate to the target locale [...] where it will be used and sold." (Jiménez-Crespo, 2013: 13). From this definition, we can derive the pillars of this process: its object, its linguistic and cultural components, and the locale. The objects concerned in localization not 'texts' – as it would be for translation –, 'products': the need of meeting the linguistic and cultural expectations related to introducing and selling this product in other markets is what fundamentally distinguishes localization from simple translation. These various markets are better defined as 'locale', meaning a space-time combination or language and region to which the adapted content will be distributed. We could, in

53 LISA was a Swiss-based trade organization specialized in the translation of computer software and related content into various natural languages. It was active from 1990 until its dissolution in February 2011.

other words, say that localization is a form of translation aimed at facilitating sales in a foreign market.

The localization industry was born in the 1970s, and primarily revolved around the adaptation of computer programs, for the most part products in the United Stated and then exported to other markets. This represented a novelty in the field of translation, as the very medium that hosted the material to be translated, which switched from paper to a digital format. For the following couple of decades, the term localization inherently involved exclusively technological products and the translation thereof, originated from the need of expanding the developing high-tech industry; other uses in other fields was, until that moment, not contemplated. It is after the widespread distribution and use of the personal computer that, in the 1990s, that the concept of 'localization' started assuming new meanings and multiple applications, eventually overcoming the purely textual element (Jiménez-Crespo, 2013: 7-8). It is then that the localisation industry firmly established itself as a significant segment of the translation industry. A further push derived from the invention of the World Wide Web, which introduced the need of localizing website and, with this, gave a rapid and considerable boost to the industry, a trend that is in constant growth due to the increasing number of web users: more than five billion[54]. A thriving and extremely important sector of localization is the one concerned with website localization, a field that today occupies a significant portion of the localization market.

Technically, web globalization typically involves two key integration processes functional to achieving an international web presence: web internationalization and web localization. Web internationalization is the process of preparing digital content for localization, cultural customization, and international deployment without needing further extensive redevelopment. This is done by creating or adapting a product from an internationalization perspective, that is, generalizing it so that it can handle multiple languages and cultural conventions. This process is integrated early in the design and development phases, preventing the need for costly and time-consuming re-designing that might arise if addressed later. More concretely, this can be exemplified by

54 Source: Statista, 2023.

elimination of all overly specific cultural references, in terms of text, but also colors and graphics, that could cause translation difficulties or misunderstandings in the target locale. The website should be neutral and functional, designed to make it easier for localizers to add or modify effects, colors, and references. This approach results in a 'generic' product that is more easily adaptable to various target markets. Web localization, on the other hand, is the adaptation of digital content for websites and application programs to fit the cultural, linguistic, legal, and other specific needs of a target market. While translation is an essential component of localization, it's just one part of a broader process (Benmamoun, 2021: 216).

In fact, localization itself can be conceptualized as one of a larger set of practices, summarized by the acronym GILT: Globalization, Internationalization, Localization and Translation. This framework describes a conceptual model where developers, search engineers, and translators work together to ensure the global localization process (Jiménez – Crespo, 2013: 9). In this sense, globalization indicates the strategy aimed at making a product, indeed, global: to achieve this, it is first of all necessary that the product undergo an internationalization process; this standardization ensures that web content remains uniform and consistent across all global markets, meaning users see the same web layout no matter where they are in the world. The product is then adapted to the specific target locale localized, and by this thus made global (Pym, 2004: 30).

3.2 Translation and localization

As mentioned above, translation and localization are two practices that, although very much similar and inherently intertwined, differ for a number or aspects. Translation concerns, first of all, a text, a written piece of content that is to be transferred from one linguistic code to another, so as to make it intelligible to a new audience. In this process, the content of the translated object remains fundamentally unvaried in its core; indeed, translation quality has traditionally been defined in terms of equivalence, and accompanied by the age-old controversies surrounding whether the

translator should strive for terminological equivalence – that is, produce a translated text that is closer to the original in terms of form – or textual equivalence – a translate text that is closer to the original in terms of message. A development of this theory – proposed to bridge the gap between theory and practice that existed in equivalence theory – was Hans Vermeer's skopos theory, or *skopostheorie* (1978). While both models represent a functional approach to translation, the skopostheorie underline the function of a translated text as the determining criteria for its evaluation, and thus as leading theoretical principle upon which to base one's work. As such, it fundamentally agrees with the concept of localization. Localization is usually viewed as distinct from translation on the basis of two major aspects: first, it is characterized by a commercial emphasis on products and related materials rather than only texts; second, localization often concerns the non-linguistic, cultural elements of a product. It is to be noted, however, that these are not absolute, necessary conditions, and that labels and boundaries between translation and localization remain somewhat blurred (Lommel, 2006: 223).

As illustrated by the definition provided by LISA, localizing a product means making a product suitable for the target local market by adapting all those extra-textual components that belong to a specific culture. These elements generally encompass, for instance, currency, number and date format, measurements, legal regulation that are specific to a particular system, but also symbols, traditions, colors, imagery, and generally items that might go against a specific culture's value system, customs, or consumption patterns, which – as opposed to systematic difference – are less easily identifiable. In this context it is highly influential the contribution of Hofstede, which has shed light on multiculturalism and differences across the world. In his work, Hofstede systematically classified cultures and group organization according to six principles, which he labeled as 'cultural dimensions':

- Power Distance, meaning 'the extent to which the less powerful members of organizations and institutions expect and accept that power is distributed unequally';
- Uncertainty Avoidance, meaning 'intolerance for uncertainty and ambiguity';

- Individualism vs. Collectivism, meaning ‘the extent to which individuals are integrated into groups’;
- Masculinity vs. Femininity, meaning ‘assertiveness and competitiveness versus modesty and caring’;
- Long Term vs. Short Term Orientation, meaning ‘the fostering of virtues oriented towards future rewards, in particular, perseverance and thrift [as opposed to] the fostering of virtues related to the past and present, in particular, respect for tradition, preservation of ‘face’ and fulfilling social obligations’;
- Indulgence vs. Restraint, meaning ‘a tendency to allow relatively free gratification of basic and natural human desires related to enjoying life and having fun [as opposed to] a conviction that such gratification needs to be curbed and regulated by strict social norms’.

(Hofstede, Hofstede, & Minkov, 2010)

This framework allows to acquire a global perspective on cross-cultural differences, and provide a powerful and practical tool to approach international audiences[55]. Taking into consideration these aspects is a fundamental part of any localization process; negligence in this respect could not only result in ineffective marketing actions, thus undermining any efforts, but it could also actively damage a business’ reputation and brand image if the lack of consideration is deemed as offensive.

55 Hofstede's cultural dimensions, while widely influential in cross-cultural research, have faced significant criticism. One key critique is that the model oversimplifies the complexities of culture by reducing it to a limited set of dimensions, potentially leading to stereotyping and an overgeneralization of national cultures. Critics argue that culture is dynamic and multifaceted, not easily captured by static dimensions. Additionally, Hofstede's research is often criticized for its methodological limitations, such as the reliance on data from a single company – IBM, where he was working at at the time – and during an extremely restricted time period – during the 1960s and 1970s –, which raises questions about the generalizability of the findings. The model has also been challenged for its ethnocentric bias, as it was developed based on Western perspectives, which may not accurately reflect the cultural realities of non-Western societies. Moreover, the assumption that national culture is homogeneous is problematic, ignoring the diversity within nations and the impact of regional, social, and organizational cultures. (McSweeney, 2002)

3.3 The role of the translator

Drawing from what has been said so far about the contrast between translation and localization, one might conclude that, in all major business contexts and essentially in any field except those purely literary and editorial, the former has fundamentally been relegated as an underdeveloped version of the latter. Localization indubitably often includes linguistic adjustments, perhaps because the target audience speaks a different language, but it is, nonetheless, interesting to note that the concept of translation is absent from this definition of localization, a detail that can be considered indicative of the role it plays within the theory of localization. In fact, if on one hand the theory relegates it to being merely a language problem – since in a perfectly internationalized product it becomes merely an automatic process – practice shows the opposite. Indeed, in reality, complete internationalization is not always possible, and therefore the translation process becomes more complex and an integral part of localization (Pym, 2004: 54). This is for two main reasons: firstly, cultures are extremely complex and multilayered systems, of which it is often hard to gain awareness for an untrained individual, who tends to unconsciously follow their language habits and thought patterns. Let us think of a collocation, an idiom, or a culturally specific reference that one might use without noticing, thus failing to recognize that such expression will most probably not produce the same rhetorical outcome in the target language, if not make sense at all, leading to the negative effects mentioned above. This instance shows the fundamental, critically important role that professional translators play as cultural mediators and experts, able to skillfully choose one word over the other according to the desired effect (Zhu, 2009: 57-58). Secondly, experienced communicators have the experience and expertise necessary to maximize the power of communication. As inherently intertwined with marketing, localization holds persuasion as one of its principles. People of different cultures hold different values, assumptions, beliefs, and attitudes towards the world; if language is used to meet their expectation in terms of worldview, the target audience is more likely to find content more engaging and be more easily persuaded by it. Returning to Hofstede's dimension, when localizing a

product for an audience that is characterized by a strong collectivistic trait – such as the Japanese culture, for instance – it may be strategic choice to implement references to group cohesion rather than to individual self-interest; for users coming from a highly uncertainty avoidant culture – for example, Germans – are best addresses with precise and detail information (*ibidem*: 62-63). A translator is not only a linguist, he is a connoisseur of what there is behind the languages he works with.

This trivial ability is also what differentiated a human translator and automated translation. Through innovation in the field of AI technology and machine learning, machine translation has become more and more advanced and sophisticated. However, despite its widespread use and its convenience in terms of both costs and timing, it presents significant limitations when compared to human translation, particularly in terms of linguistic nuance and cultural sensitivity. These systems, which rely heavily on algorithms and large datasets, often produce translations that are literal interpretations of the source text and fail to capture the subtleties of meaning, context, and tone, which are crucial for accurate communication. Studies have shown that machine translation softwares struggle with idiomatic expressions, complex sentence structures, and wordplay, leading to translations that may be awkward, inaccurate, or even culturally insensitive; while neural machine translation has improved fluency, it still falls short in handling ambiguities and context-specific interpretations (Gaspari et al., 2014). As such, while automated translation can greatly assist the translator in performing tasks – in the form of Computer-Assisted Translation (CAT) tools, it cannot yet replace him fully, lacking the expertise, the depth of understanding, and cultural awareness provided by humans.

3.4 Web localization for SEO

As introduced in Section 3.1, web localization involves customizing digital content on websites and mobile applications to meet the specific cultural, linguistic, legal, and other needs of a target market. Adaptation of formats or language or both are fundamental components of the localization process, which acquire additional layers of

difficulty when it comes to websites. To currency, date, number formats, SEO localization adds optimizing the target keywords, title tags, meta descriptions, URLs, imagery and alt text, external links and anchor tags, so as to optimize it for local search algorithms and audience preferences. To accommodate these needs, it is necessary to create a design that is responsive and flexible and can be easily adapted to serve other versions; this includes a page formatting that is spacious enough to host longer texts; practices to facilitate page speed to serve countries with slower internet connections; different payment methods to support audiences' preferences and encourage conversions (Jones, 2023). Holistically customizing the experience offered to the user can be critically important. Effective localization is a key factor in boosting your rankings in international SEO. In this sense, localization is the most visible expression of the practical application of the functionalist approach of Skopostheorie to professional reality: the site for the new target market and its function or *skopos* are the primary focus, and can very concretely affect performance and business results (Sandrini, 2005). It has been shown that localization can improve consumers' intentions toward websites and their willingness to pay, explained by the link between cultural congruence and reduced effort to process information and resulting in up to a 200% increase in a company's e-sales outside its language borders (Tixier, 2015: 16). In addition to making a site more relevant and attractive to the foreign audience, these procedures also increase the likelihood to be linked by reputable pages in those regions; as explained in Chapter three, such backlinks are vital for SEO, as they indicate to search engines that that content is valuable and trustworthy (Silva, 2023).

3.5 Types of web localization

In its essence, a website consists of a collection of pages, usually united by a common theme and design; when approaching website localization, it's essential to recognize that not all content serves the same purpose, and therefore, each type requires a unique strategy. The lowest on the localization spectrum is, generally, brand content such as homepages and "about us" pages; adjustment on these pages concerns mainly

simple translation of short texts and section titles, possibly reading direction – typically left-to-right in western cultures, as opposed to right-to-left in some eastern cultures –, although particular attention must be paid to the use of colors and imagery.

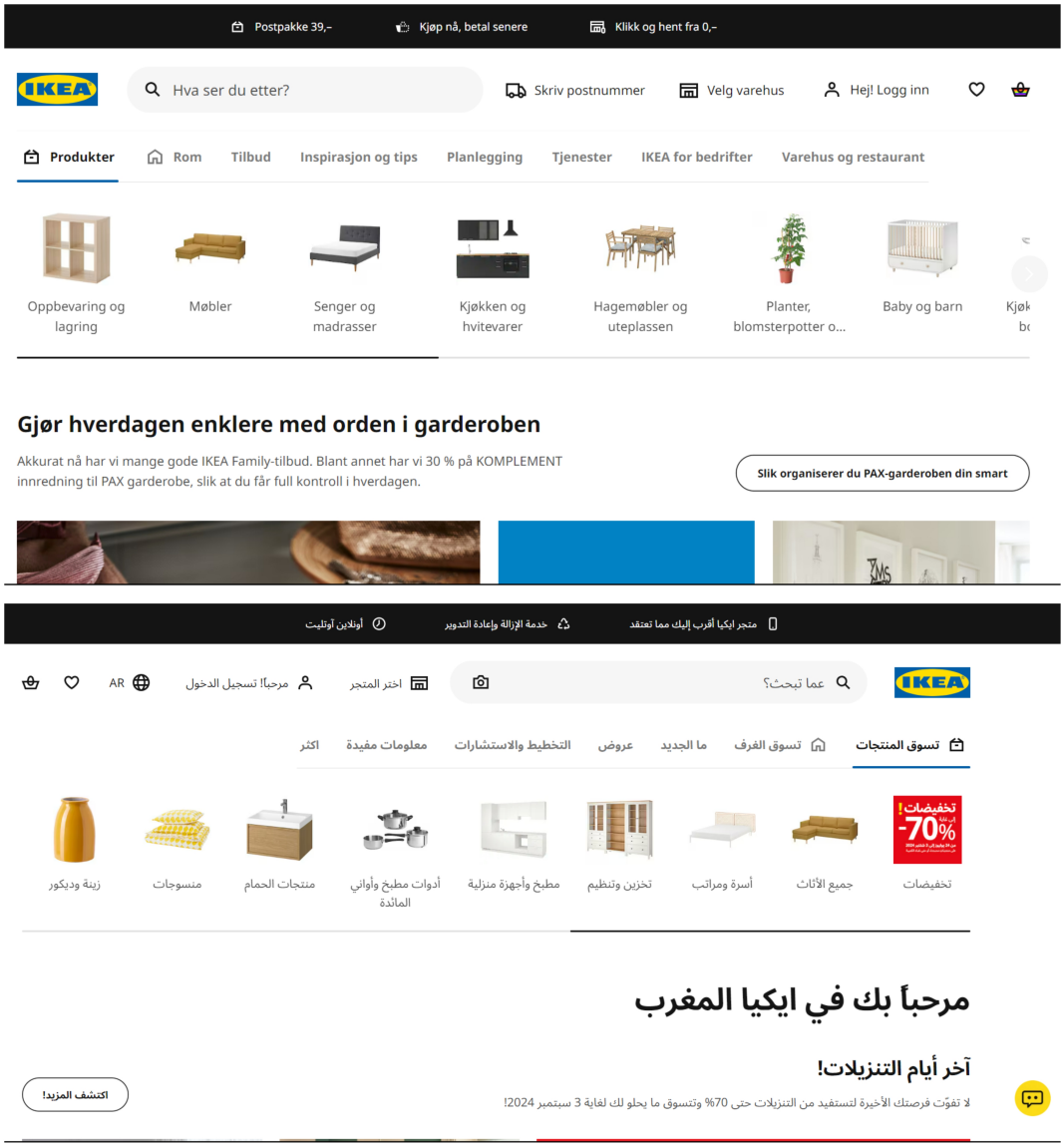
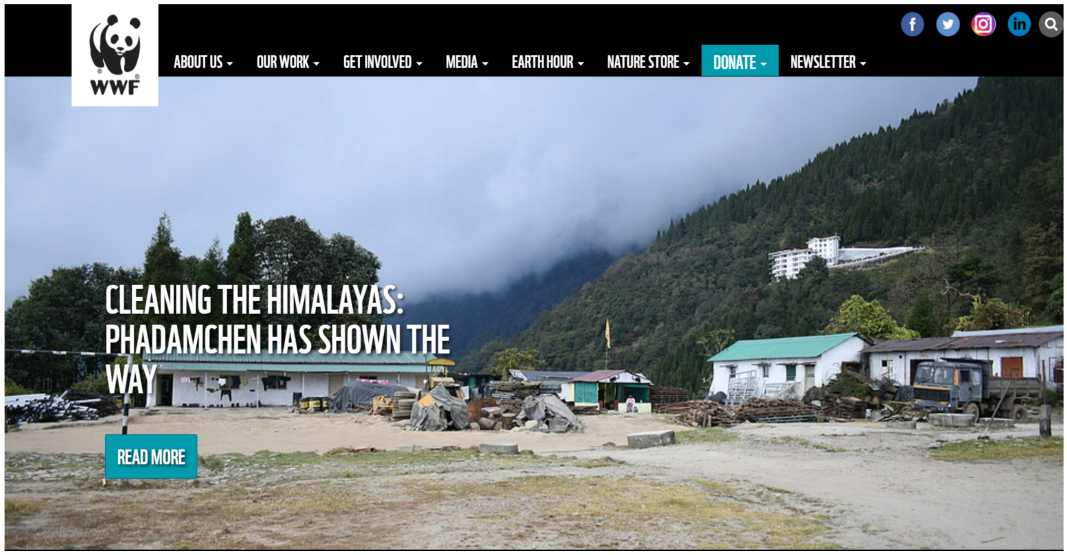


Figure 48: In the top picture, the homepage to the Norwegian version of the website, with products, descriptions, and the navigation menu written in Norwegian and following the right-to-left reading direction of the language; in the bottom page, the homepage to the Moroccan version of the website, with products, descriptions, and the navigation menu written in Arabic and following the left-to-right reading direction of the language. Screenshots taken from: https://www.ikea.com/no/no/ and https://www.ikea.com/ma/ar/

Since these pages do not fulfill any specific function if not that of welcoming the visitor and assisting him in navigating through the website, it is advisable to opt for a strong internationalization strategy, so as to reduce to the minimum the efforts required.

Highly localizable is, on the other hand, technical and legal content; pages such as contact, product information, guides, terms of use documents, data protection, refund policies is inherently specific to the locale, for which it is essential to translate it with high precision to avoid any misunderstandings and in compliance with local regulations. Lastly, marketing content is the most challenging to localize, and the one that requires the most mindful and sophisticated interventions. Regardless of what the website is about, the purpose of marketing is to persuade and engage, and in order to maintain this impact and effectiveness across different regions, it is essential to craft messages that reflect the local cultural norms, preferences, and buying behaviors, and that ultimately truly connect with the target (Jones, 2023). For this reason, marketing content is often entirely rewritten to ensure it effectively resonates with the new audience, and at the same time to maximize organic performance. Just like customs vary across countries, so do search habits and trends, which call for a thorough redesigning of keyword schemes that cannot be achieved with literal – and certainly not automated – translation. Search engines prioritize relevant and high-quality, original content; poor translation might be deemed as low-quality by the ranking algorithm, as well as risking duplicate content issues if the result is a product that too closely resembles the original[56].



56 In this respect, as we have seen in Chapter three, hreflang tags can be implemented to signal localized versions of a same piece of content. However, hreflang tags are a signal, not a directive, and as such concur with other signals to influence ranking; if the content is badly translated or duplicate, regardless of any hreflang attributes, it can still impact SEO negatively.

Figure 49: In the top picture, the homepage to the Indian version of the website written in English (one of the official languages of the country) and featuring an article on a topic relevant to the target, namely a community-driven initiative to address environmental issues by undertaking a massive cleanup of the Indian Himalayan Region; in the bottom picture the homepage to the Japanese version of the website written in Japanese and featuring an article on a topic relevant to the target, namely a community-driven campaign encouraging individuals, businesses, and local governments to get involved and support the transition towards carbon neutrality in the city of Hiroshima. Screenshots taken from: https://www.wwfindia.org/ and https://www.wwf.or.jp/

3.6 Global SEO strategy for international markets

Managing multiple websites to appeal to an international audience can be challenging. Coordinating these localized versions requires constant updates and consistency across all platforms, which can be complex and resource-intensive. Ideally, a website aimed at foreign audiences should be 100% localized to ensure it meets the specific needs and preferences of each target market and increase its relevance and effectiveness. However, in reality achieving such a high level of localization is frequently impeded by constraints in resources, including financial limitations, time constraints, and the availability of skilled personnel. As a result, many organizations are compelled to implement varying degrees of localization based on their resource availability and strategic priorities. This often leads to a spectrum of localization practices, ranging from minimal adaptation to more extensive modifications. In practice, some websites may adopt a less localized approach, involving only basic translation and

the inclusion of local contact information while retaining uniform design elements and a standardized user interface. Conversely, other websites might engage in a more nuanced level of localization, focusing on key markets by tailoring content, user interactions, and critical elements to address specific regional needs while maintaining a consistent overall structure across different locales. Conceptually, we can classify five different levels of localization, from which five corresponding types of websites emerge:

- Standardized Websites, that is, sites which present the same content to all users in a single language, generally – and understandably – English, given its given its widespread use; a monolingual and culturally independent site is considered the most practical and cost-effective solution;
- Semi-Localized Websites, that is, sites that provide contact information and company details specific to certain countries, but maintain a single language for all countries, similar to standardized websites, and lack cultural references or adaptations; due to the only partial localization action required, this solution is also relatively less costly than other, more radical options;
- Localized Websites, that is, sites that offer a translated version for each target country including only the pages that are relevant to the target audience, leaving untranslated pages not deemed of interest are not translated; less localized websites might feature local language and content, but they usually maintain consistent design elements across all versions;
- Highly Localized Websites, that is, sites that provide not only translations of content but also specific cultural adaptations for the target country, such as local currency, numbers, dates, and other regional specifics;
- Customized Websites, that is, sites that are fully localized, with the content and design extensively tailored to fit the target audience; this kind of websites are effectively rebuilt and rewritten to suit the cultural context of the target market, placing a strong emphasis on cultural adaptation, which is absent in all other levels of localization.

(Singh & Pereira, 2005: 10-15)

Companies must balance the ideal with the achievable, prioritizing key areas for localization that will have the most significant impact on user engagement and satisfaction, given their resource limitations. The disparity between ideal and practical localization underscores the need for organizations to strategically prioritize their localization efforts. Companies must balance the ideal with the achievable, prioritizing key areas for localization that will have the most significant impact on user engagement and satisfaction within the constraints of their operational capabilities. Localization implementation can also vary from country to country, defending on the attractiveness and the value of each specific market: a business marketing ski equipment will most likely prioritize the localization of a website for the Swiss, German, or Austrian market, and maybe tackle the Irish or UK market only at a later time.

An important consideration to make, and fundamental to avoid wasting resources and completely missing the target, is pending those efforts on the right platform. As anticipated in Chapter one, many are the search engines active and constantly developing around the world. In the present work the focus was set on Google, currently the most popular of all search engines[57] and the one that is the most relevant to what is typically called the Western world, where the intended public for this dissertation and I, the author, are currently based. Relatively speaking, if one was to approach the global market with limited resources, that it can be considered a generally wise choice to concentrate SEO efforts on Google as it would maximize the chances of appealing to a wider audience. However, when choosing to target foreign markets using a localizing approach, as part of the preliminary market research it is important to assess which search engine is the most popular in that particular context, the one which has the most probabilities to reach the most potential prospects, and consequently tailor a SEO strategy on that specific search engine (Arosio, 2020). As of 2024, in China, for instance, Baidu is the market leader in the online search industry, whereas Google, despite its unquestioned dominance worldwide, is relegated to a mere 1.83% of the market[58].

57 Source: Statista, 2023.
58 Source: Statista, 2024.

The process of managing and localizing multiple websites for an international audience is a multifaceted challenge that requires a careful balance between ideal aspirations and practical realities. While fully localized websites tailored to each specific market offer the greatest potential for user engagement and relevance, they also demand significant resources in terms of time, money, and skilled personnel. Consequently, organizations must strategically prioritize their localization efforts, focusing on key markets where the impact of these efforts will be most substantial. Moreover, it is essential for companies to recognize the importance of targeting the right platforms and search engines when approaching different international markets. By aligning localization efforts with the most effective platforms for each market, companies can maximize their reach and engagement, even within the constraints of limited resources.

CONCLUSION AND REFLECTION FOR FUTURE IMPLICATIONS

The aim of this thesis was to provide a thorough understanding of both the foundational elements of SEO and the advanced strategies required to succeed in the increasingly globalized digital environment. The exploration of SEO and International SEO has underscored the vital role these strategies play in the digital marketing landscape. As businesses continue to navigate the complexities of the online world, the ability to effectively leverage search engines to enhance visibility, drive traffic, and achieve business objectives has become indispensable.

The current work has first and foremost provided understanding of the mechanics of search engines, with a particular focus on Google, the dominant player in the search market. Understanding how search engines operate is the foundation of any SEO strategy. Google's complex algorithms have evolved significantly over the years, and dictate how web pages are ranked in search results. These updates have emphasized the importance of high-quality content, authoritative backlinks, and a positive user experience. The discussion in Chapter one provided insights into its historical development, the key components of its functionality—crawling, indexing, and ranking—and the critical importance of understanding user intent.

Chapter two of this thesis explored the critical elements of keyword analysis and content creation – the pillars that are essential for successful SEO. Content remains king in the world of SEO, as it is the primary means by which businesses communicate with both search engines and users. The discussion on content creation emphasized the need for originality, relevance, and comprehensiveness in order to meet the expectations of both users and search engine algorithms. Nonetheless, keyword analysis emerged as a crucial component of this process, enabling websites to target their content to specific audiences by understanding the terms and phrases users are searching for. However, it is not enough to simply insert keywords into content; effective SEO requires that these keywords be seamlessly integrated into well-crafted copy that engages and informs the reader. The art of SEO copywriting, as explored in this thesis, involves balancing the

needs of search engines with the expectations of users. As Google's algorithms increasingly prioritize content that reads naturally and provides real value, the role of natural language processing (NLP) in content creation has become more prominent.

Optimization, the process of refining a website to enhance its search engine rankings, was the focus of chapter three. This chapter broke down the three main areas of optimization: on-page, off-page, and technical SEO, each of which plays a distinct but interconnected role in a comprehensive SEO strategy. On-page optimization involves the direct management of elements within a web page, such as title tags, meta descriptions, headers, and internal links. The importance of mobile optimization and user experience (UX) was also discussed, highlighting the need for fast-loading pages and intuitive navigation. On the other hand, off-page optimization, which involves activities conducted outside of one's own website, primarily focuses on building high-quality backlinks. The process of link building, as outlined in this thesis, requires careful planning and execution to ensure that links are obtained from authoritative and relevant sources. Off-page optimization also plays a crucial role in establishing domain authority, which in turn influences a site's ability to rank well in search engine results. Technical SEO, the third component of optimization, ultimately supports any SEO practice, addressing the underlying infrastructure and functions of a website. This chapter explored the various technical elements that contribute to SEO success, including site speed, XML sitemaps, robots.txt files, and canonical tags, emphasizing the need for websites to help search engines better understand and present their content. Technical SEO ensures that a website is not only accessible to search engines but also performs efficiently for users.

In the final chapter of this thesis, we turned our attention to International SEO, an increasingly important area as businesses expand their reach into global markets. International SEO is more complex than domestic SEO, as it requires a deep understanding of the cultural, linguistic, and behavioral differences across various regions. This chapter provided a comprehensive guide to the strategies and techniques necessary for optimizing websites for a global audience. On the more technical side, one of the key aspects of International SEO is the correct implementation of hreflang tags,

which signal to search engines the intended language and geographical targeting of a web page. The discussion on the use of hreflang tags, along with the considerations surrounding country code top-level domains (ccTLDs), subdirectories, and subdomains, as well as the use of geotargeting strategies, provided a framework for structuring international websites in a way that avoids common pitfalls such as duplicate content and improper targeting. Localized keyword research emerged as another critical component of International SEO, which enables the creation of localized, high quality content. Content localization, a key topic in this chapter, goes beyond mere translation, but rather involves adapting content to fit the cultural context of the target audience, using native speakers for translation, and customizing images and messages to align with local customs and values. This thesis provided best practices for content localization, ensuring that businesses can create region-specific content that not only ranks well in search engines but also resonates with local users. The present work also emphasized the importance of understanding local search behaviors and conducting keyword research in the native language of the target market. This approach ensures that content resonates with local audiences and aligns with their search intent.

The comprehensive exploration of SEO and International SEO in this thesis has demonstrated that success in the digital age requires more than a basic understanding of how search engines work. It requires a strategic, multifaceted approach that encompasses content creation, keyword analysis, on-page and off-page optimization, and technical SEO, all of which have to be mindfully adapted to each targeted locale, facing the challenges of operating in a global market. As the digital landscape continues to evolve, SEO strategies must also adapt to keep pace with changes in search engine algorithms, user behavior, and technological advancements: while the insights provided in this thesis offer a solid foundation for understanding and implementing effective SEO strategies, the field of SEO is dynamic, and continuous learning and adaptation are essential for long-term success. Looking forward, the incremental integration of artificial intelligence (AI) and machine learning into search engine algorithms will likely become more and more pronounced, further emphasizing the importance of creating user-centric content and maintaining high technical standards.

For businesses seeking to expand their global presence, International SEO will remain a critical component of their digital strategy. As markets become more interconnected, the ability to effectively reach and engage audiences across different regions and languages will be a key determinant of success. This thesis has provided a comprehensive guide for navigating these complexities, but as the global digital landscape continues to shift, businesses must remain agile and responsive to the changing needs and expectations of their international audiences. SEO and International SEO are not static disciplines: they are ever-evolving fields that require ongoing attention and refinement. By understanding the principles outlined in this thesis and keeping track of industry developments, businesses can position themselves to thrive in the competitive and dynamic world of digital marketing.

Bibliography

Abouobaia, Jenny. 2022. “Digital PR: The Beginner's Guide to Making Your Brand Unmissable.” Ahrefs. Retrieved at https://ahrefs.com/blog/digital-pr/.

Alderson, Jono. 2020. “How to optimize browser caching.” Yoast. Retrieved at https://yoast.com/optimize-browser-caching/.

Alderson, Jono. 2021. “How to use meta robots tags: the ultimate guide.” Yoast. Retrieved at https://yoast.com/robots-meta-tags/.

Ali, Asif. 2023. “SEO vs. SEM: Differences in Techniques, Costs, & Results.” Semrush. Retrieved at https://www.semrush.com/blog/seo-vs-sem/.

Andrienko, Olga. 2020. “How To Optimize for People Also Ask.” Semrush. Retrieved at https://www.semrush.com/blog/how-to-maximize-people-also-ask-seo-opportunities-study/.

Arosio, Emmanuele. 2020. “Ecommerce multilingua: gli errori che affossano una strategia SEO.” Semrush. Retrieved at https://it.semrush.com/blog/ecommerce-multilingua-errori-strategia-seo/.

Benmamoun, Mamoun, Hadi Alhor, Christine Ascencio, and Woojong Sim. 2021. “Social enterprises in electronic markets: web localization or standardization.” *Electronic Markets* 31:215-231. https://doi.org/10.1007/s12525-020-00430-7.

Berysevich, Aleh. 2021. “How To Identify & Eliminate Keyword Cannibalization To Boost Your SEO.” search engine Journal. Retrieved at https://www.searchsearch enginejournal.com/on-page-seo/keyword-cannibalization/.

Besik, Heidi. 2017. “How Brands Use User-Generated Content.” search engine Journal. Retrieved at https://www.searchsearch enginejournal.com/user-generated-content-examples/214876/.

Bickart, Barbara, and Robert M. Schindler. 2001. “Internet forums as influential sources of consumer information.” *Journal of Interactive Marketing* 15 (3): 31-40. https://doi.org/10.1002/dir.1014.

Boroda, Nikolai. 2022. “What Are SERP Features? 18 Google Search Features to Know.” Semrush. Retrieved at https://www.semrush.com/blog/serp-features-guide/.

Burton, Winston. 2024. “Will AI Replace SEO Specialists?” search engine Journal. Retrieved at https://www.searchsearch enginejournal.com/will-ai-replace-seo-specialists/509934/.

Byers, Kyle. 2023. “What Is a SERP? search engine Results Pages, Explained.” Semrush. Retrieved at https://www.semrush.com/blog/serp/.

Camarena, Ana. 2023. “What Is Page Speed & How to Improve It.” Semrush. Retrieved at https://www.semrush.com/blog/page-speed.

Chetvertak, Daria. 2024. “How to Find Competitors Backlinks.” SE Ranking. Retrieved at https://seranking.com/blog/find-competitor-backlinks/.

Cialdini, Robert B. 1984. *Influence: The Psychology of Persuasion*. N.p.: Business Library.

Clark, Jon. 2023. “18 Online Review Statistics Every Marketer Should Know.” search engine Journal. Retrieved at https://www.searchsearch enginejournal.com/online-review-statistics/329701/.

Cojocariu, Adrian. 2023. “SEO Visibility: What It Is & How to Improve It.” Semrush. Retrieved at https://www.semrush.com/blog/seo-visibility/.

Collins, Sean. 2024. “What Is Keyword Difficulty? (& How to Measure It).” Semrush. Retrieved at https://www.semrush.com/blog/keyword-difficulty/.

Crowe, Anna. 2021. “What Exactly Is E-A-T & Why Does It Matter to Google?” search engine Journal. Retrieved at https://www.searchsearch enginejournal.com/google-eat/what-is-it/.

Cunningham, Camille. 2020. “Blog post structure: How to set up an easy-to-read text.” Yoast. Retrieved at https://yoast.com/writing-blog-creating-clear-blog-post-structure/.

Cunningham, Camille. 2022. “Homepage SEO: Does it exist?” Yoast. Retrieved at https://yoast.com/homepage-seo/.

Cunningham, Camille. 2023. “Make a great website menu for your users and SEO.” Yoast. Retrieved at https://yoast.com/seo-website-menu/.

Cunningham, Camille. 2024. “How does Google understand text?” Yoast. Retrieved at https://yoast.com/how-does-google-understand-text/.

Davydov, Dmytro. 2023. “Dofollow and nofollow links and their impact on SEO.” SE Ranking. Retrieved at https://seranking.com/blog/dofollow-nofollow-links/.

Dean, Brian. 2023(c). “Broken Link Building: The Complete Guide.” Backlinko. Retrieved at https://backlinko.com/hub/seo/broken-link-building.

Dean, Brian. 2023(b). “Off-Page SEO: The Definitive Guide (2024).” Backlinko. Retrieved at https://backlinko.com/off-page-seo-guide.

Dean, Brian. 2024(a). “Google RankBrain: The Definitive Guide.” Backlinko. Retrieved at https://backlinko.com/google-rankbrain-seo.

Dean, Brian. 2024(c). “What is Crawl Budget? And Why Does It Matter for SEO?” Backlinko. Retrieved at https://backlinko.com/hub/seo/crawl-budget.

Dean, Brian. 2024(b). “Duplicate Content and SEO: The Complete Guide.” Backlinko. Retrieved at https://backlinko.com/hub/seo/duplicate-content.

Dearie, Kari. 2023. “What Is Link Bait? 6 Real-World Examples + Link Bait Ideas.” Semrush. Retrieved at https://www.semrush.com/blog/link-bait/.

de Valk, Joost. 2023. “The ultimate guide to robots.txt.” Yoast. Retrieved at https://yoast.com/ultimate-guide-robots-txt/.

Einorytė, Aurelija. 2023. “Browser vs. search engine: What’s the difference?” NordVPN. Retrieved at https://nordvpn.com/it/blog/browser-vs-search-search engine/.

Eva, Viola, and Joshua Hardwick. 2021. “How to Create & Optimize Landing Pages for SEO (Step-by-Step Guide).” Ahrefs. Retrieved at https://ahrefs.com/blog/landing-page-seo/.

Fitzgerald, Anna. 2024. “Website Navigation: The Ultimate Guide [Types & Top Examples].” HubSpot Blog. Retrieved at https://blog.hubspot.com/website/main-website-navigation-ht.

Frankel, Alexis. 2023. “What Is Organic SEO and How to Get Started.” Semrush. Retrieved at https://www.semrush.com/blog/organic-seo/.

Gabe, Glenn. 2020. “Google's New Page Experience Signal: An Object-Oriented Approach to UX Ranking Factors, But Will It Have Teeth?” Semrush. Retrieved at https://www.semrush.com/blog/google-page-experience-signal/.

Gaspari, Federico, Antonio Toral, Sudip Kumar Naskar, Declan Groves, and Andy Way. 2014. “Perception vs. reality: measuring machine translation post-editing productivity.” *Proceedings of the 11th Conference of the Association for Machine Translation in the Americas*, 60-72. Retrieved at https://aclanthology.org/2014.amta-wptp.5.

Go, Sydney. 2024. “What Is a Dofollow Link? (& Why It Is Important in SEO).” Semrush. Retrieved at https://www.semrush.com/blog/dofollow-link/.

Grant, Jamie, and Despina Gavoyannis. 2023. “10 International SEO Best Practices + Checklist.” Ahrefs. Retrieved at https://ahrefs.com/blog/international-seo/.

Grubbs, Patrick. 2022. “How to Disavow Links: A Guide.” Semrush. Retrieved at https://www.semrush.com/blog/how-to-disavow/.

guide, step, and Brian Dean. 2023(a). “What Is Bounce Rate? And How to Quickly Improve It.” Backlinko. Retrieved at https://backlinko.com/hub/seo/bounce-rate.

Haines, Chris. 2023. “Internal Links for SEO: An Actionable Guide.” Ahrefs. Retrieved at https://ahrefs.com/blog/internal-links-for-seo/.

Hallebeek, Willemien. 2022. “Hosting guide: how to pick the right host for your site.” Yoast. Retrieved at https://yoast.com/hosting-guide-how-to-pick-the-right-host-for-your-site/.

Handley, Rachel. 2023(b). “The Marketing Funnel: What It Is & How It Works.” Semrush. Retrieved at https://www.semrush.com/blog/marketing-funnel/.

Handley, Rachel. 2023(c). “How to Do Keyword Clustering & Why It Helps SEO.” Semrush. Retrieved at https://www.semrush.com/blog/keyword-clustering/.

Handley, Rachel. 2023(d). “Alt Text: What Is It & Why It Matters for Accessibility & SEO.” Semrush. Retrieved at https://www.semrush.com/blog/alt-text/.

Handley, Rachel. 2023(e). “Breadcrumb Navigation for Websites: What It Is & How to Use It.” Semrush. Retrieved at https://www.semrush.com/blog/breadcrumbs-for-websites/.

Handley, Rachel. 2023(a). “Keyword Stuffing in SEO: What It Is & How to Avoid It.” Semrush. Retrieved at https://www.semrush.com/blog/keyword-stuffing/.

Handley, Rachel. 2024(l). “XML Sitemap: What It Is & How to Generate One.” Semrush. Retrieved at https://www.semrush.com/blog/xml-sitemap/.

Handley, Rachel. 2024(d). “What Keywords Are & How to Use Them.” Semrush. Retrieved at https://www.semrush.com/blog/what-are-keywords/.

Handley, Rachel. 2024(i). “What Keywords Are & How to Use Them.” Semrush. Retrieved at https://www.semrush.com/blog/what-are-keywords/.

Handley, Rachel. 2024(c). “Why Is SEO Important? 10 Reasons It Matters.” Semrush. Retrieved at https://www.semrush.com/blog/importance-of-seo/.

Handley, Rachel. 2024(e). “Keyword Research for SEO: What It Is & How to Do It.” Semrush. Retrieved at https://www.semrush.com/blog/keyword-research/.

Handley, Rachel. 2024(h). “Keyword Research for SEO: What It Is & How to Do It.” Semrush. Retrieved at https://www.semrush.com/blog/keyword-research/.

Handley, Rachel. 2024(b). “Google E-E-A-T: What It Is & How It Affects SEO.” Semrush. Retrieved at https://www.semrush.com/blog/eeat/.

Handley, Rachel. 2024(f). “Semantic Keywords: What They Are & How to Find Them.” Semrush. Retrieved at https://semrush.com/blog/semantic-keywords/.

Handley, Rachel. 2024(g). “Keyword Search Volume: What It Is & How to Find It.” Semrush. Retrieved at https://www.semrush.com/blog/keyword-search-volume/.

Handley, Rachel. 2024(a). “Google Algorithm Updates Timeline [2024-2009].” Semrush. Retrieved at https://www.semrush.com/blog/google-algorithm-update/.

Hardwick, Joshua. 2020. “What Is Anchor Text? Everything You Need to Know (No Jargon!).” Ahrefs. Retrieved at https://ahrefs.com/blog/anchor-text/.

Hardwick, Joshua. 2023. “Resource Page Link Building: The Only Guide You Need.” Ahrefs. Retrieved at https://ahrefs.com/blog/resource-page-link-building/.

Hardwick, Joshua. 2024. “301 Redirects Explained: How They Impact SEO.” Ahrefs. Retrieved at https://ahrefs.com/blog/301-redirects/.

Harnish, Brian. 2020. “When, Why & How to Use the Noarchive Tag.” search engine Journal. Retrieved at https://www.searchsearch enginejournal.com/noarchive-tag/358972/.

Harsel, Luke. 2023. “What Is Organic Traffic (And How to Increase It).” Semrush. Retrieved at https://www.semrush.com/blog/organic-traffic/.

Haynes, Marie. 2022. “How to improve E-A-T for YMYL pages.” search engine Land. Retrieved at https://searchsearch engineland.com/google-ymyl-eat-seo-380569.

Hines, Kristi. 2022. “6 Reasons Why Blogging Is Important For Marketing And SEO.” search engine Journal. Retrieved at https://www.searchsearch enginejournal.com/7-reasons-why-blogging-is-still-important-in-2012/39225/.

Hofstede, Geert, Gert J. Hofstede, and Michael Minkov. 2010. *Cultures and Organizations: Software of the Mind, Third Edition*. N.p.: McGraw-Hill Education.

Irvine, Mark. 2021. “NEW Display Location Extensions & Ads: Drive More Local Business.” WordStream. Retrieved at https://www.wordstream.com/blog/ws/2017/03/14/display-location-extensions.

Janssens, Maddy, and Chris Stayeart. 2014. “Re-considering language within a cosmopolitan understanding: Toward a multilingual franca approach in international business studies.” *Journal of International Business Studies* 45:623-639. https://doi.org/10.1057/jibs.2014.9.

Jiménez-Crespo, Miguel A. 2013. *Translation and Web Localization*. N.p.: Routledge.

Jones, Padrig. 2023. “Website Localization 101: How to Expand Your Reach Globally.” Semrush. Retrieved at https://www.semrush.com/blog/website-localization/.

Jovicic, Aleksandra. 2024. “Keyword Gap Analysis: What It Is & How to Do It.” Semrush. Retrieved at https://www.semrush.com/blog/keyword-gap-analysis/.

Knezevic, Aida. 2024. “Topic Clusters for SEO: What They Are & How to Create Them.” Semrush. Retrieved at https://www.semrush.com/blog/topic-clusters/.

Lahey, Connor. 2021. “What Are Stop Words in SEO? [+ Stop Words List].” Semrush. Retrieved at https://www.semrush.com/blog/seo-stop-words/.

Lahey, Connor. 2023. “Email Outreach: How to Write Outreach Emails that Convert.” Semrush. Retrieved at https://www.semrush.com/blog/email-outreach/.

Lahey, Connor. 2024. “Meta Tags: What They Are & How to Use Them for SEO.” Semrush. Retrieved at https://www.semrush.com/blog/meta-tag/.

Larkin, Kayle. 2022. “How to Find and Fix Orphan Pages (The Right Way).” Ahrefs. Retrieved at https://ahrefs.com/blog/orphan-pages/.

Lindgaard, Gitte, Gary Fernandes, Cathy Dudek, and J. Brown. 2006. “Attention web designers: You have 50 milliseconds to make a good first impression!” *Behaviour & Information Technology* 25 (2): 115-126. https://doi.org/10.1080/01449290500330448.

Loktionova, Margarita. 2024. “What Is a Content Management System (CMS)?” Semrush. Retrieved at https://www.semrush.com/goodcontent/content-marketing-blog/content-management-system/.

Lommel, Arle. n.d. “Localization standards, knowledge- and information-centric business models, and the commoditization of linguistic information.” *Perspectives on Localization* American Translators Association Scholarly Monograph Series (13): 223-239.

Lyons, Kelly. 2022. “What Is a Title Tag & How to Optimize Your Title Tags for SEO.” Semrush. Retrieved at https://www.semrush.com/blog/title-tag/.

Lyons, Kelly. 2023(b). “Long-Tail Keywords: What They Are & How to Use Them in 2024.” Semrush. Retrieved at https://www.semrush.com/blog/how-to-choose-long-tail-keywords/.

Lyons, Kelly. 2023(a). “What Is a Competitive Analysis & How to Do It (With Template).” Semrush. Retrieved at https://www.semrush.com/blog/competitive-analysis/.

Lyons, Kelly. 2024. “Cost Per Click: Definition, Why It Matters + How to Measure CPC.” Semrush. Retrieved at https://www.semrush.com/blog/cost-per-click/.

McMichael, Kaitlin. 2016. “The 15 Most Popular Myths About International SEO, Debunked.” Moz. Retrieved at https://moz.com/blog/top-15-myths-international-seo.

McSweeney, Brendan. 2002. “Hofstede’s Model of National Cultural Differences and their Consequences: A Triumph of Faith - a Failure of Analysis.” *Human Relations* 55 (1): 89-118. https://doi.org/10.1177/0018726702551004.

Morris, Kate. 2016. “The Guide to International Website Expansion: Hreflang, ccTLDs, & More!” Moz. Retrieved at https://moz.com/blog/guide-to-international-seo.

Mustapic, Boris. 2023. “Largest Contentful Paint (LCP): What It Is & How to Improve It.” Semrush. Retrieved at https://www.semrush.com/blog/lcp/.

Nicole, Dana. 2024(b). “Pagination and SEO: A Complete Guide to Best Practices.” Semrush. Retrieved at https://www.semrush.com/blog/pagination-seo/.

Nicole, Dana. 2024(a). “What Is Website Architecture? (& How to Improve it for SEO).” Semrush. Retrieved at https://www.semrush.com/blog/website-structure/.

Ong, Quan. 2023. “How to Execute the Skyscraper Technique (And Get Results).” Ahrefs. Retrieved at https://ahrefs.com/blog/skyscraper-technique/.

Ong, Quan. 2024. “Content Promotion: There are Only 3 Ways to Distribute Content.” Ahrefs. Retrieved at https://ahrefs.com/blog/content-promotion/.

Ong, Quan. 2024. “The Marketing Funnel: What It Is, How It Works, & How to Create One.” Ahrefs. Retrieved at https://ahrefs.com/blog/marketing-funnels/.

Ong, Quan, and Joshua Hardwick. 2023. “Guest Blogging for SEO: Everything You Need to Know.” Ahrefs. Retrieved at https://ahrefs.com/blog/guest-blogging/.

Page, Larry, and Sergey Brin. 2021. “20 Years of SEO: A Brief History of search engine Optimization.” search engine Journal. Retrieved at https://www.searchsearch enginejournal.com/seo/seo-history/.

Pandya, Ravi. 2024. “Google PageSpeed Insights: What It Is & How to Boost Your Score.” Semrush. Retrieved at https://www.semrush.com/blog/google-pagespeed-insights/.

Pavlik, Vlado. 2022(b). “Semantic HTML: What It Is and How to Use It Correctly.” Semrush. Retrieved at https://www.semrush.com/blog/semantic-html5-guide/.

Pavlik, Vlado. 2022(a). “What Is Search Intent? A Complete Guide.” Semrush. Retrieved at https://www.semrush.com/blog/search-intent/.

Pavlik, Vlado. 2024(b). “Nofollow Links vs. Follow Links: All You Need to Know.” Semrush. Retrieved at https://www.semrush.com/blog/nofollow-links/.

Pavlik, Vlado. 2024(a). “What Is a Slug? URL Slugs and Why They Matter for SEO.” Semrush. Retrieved at https://www.semrush.com/blog/what-is-a-url-slug/.

Pecánek, Michal. 2020. “What is HTTPS? Everything You Need to Know.” Ahrefs. Retrieved at https://ahrefs.com/blog/what-is-https/.

Perco, Ajdin. 2023. “Content Creation: What It Is & How to Do It in 2024 [Tips + Tools].” Semrush. Retrieved at https://www.semrush.com/blog/content-creation/.

Pol, Tushar. 2024. “What Is a URL? A Complete Guide to Website URLs.” Semrush. Retrieved at https://www.semrush.com/blog/what-is-a-url/.

Pym, Anthony. 2004. *The moving text : localization, translation, and distribution*. N.p.: John Benjamins Publishing Company.

Rudzki, Tomek. 2019. “Ultimate Guide to International SEO.” Onely. Retrieved at https://www.onely.com/blog/ultimate-guide-to-international-seo/.

Salsi, Hava. 2023. “Duplicate Content: What Is It + 4 Ways to Address It.” Semrush. Retrieved at https://www.semrush.com/blog/duplicate-content/.

Sarfraz, Hamid. 2020. “Is 100/100 PageSpeed score enough? The answer is No.” Semrush. Retrieved at https://www.semrush.com/blog/pagespeed-score-enough.

Saxon, Joshua. 2024. “Geo-targeting SEO: A game-changer for local businesses.” Semrush. Retrieved at https://www.semrush.com/local/blog/geo-targeting-seo/.

Sheridan, Jake, and Michal Pecánek. 2023. “What Is Topical Authority in SEO & How to Build It.” Ahrefs. Retrieved at https://ahrefs.com/blog/topical-authority/.

Silva, Carlos. 2023. “SEO Localization: Tactics for Targeted Global Online Visibility.” Semrush. Retrieved at https://www.semrush.com/blog/seo-localization/.

Silva, Carlos. 2024(a). “What Is Content Marketing? Definition & How to Use It in 2024.” Semrush. Retrieved at https://www.semrush.com/blog/what-is-content-marketing/.

Silva, Carlos. 2024(b). “What Is an H1 Tag? Why It Matters & Best Practices for SEO.” Semrush. Retrieved at https://www.semrush.com/blog/h1-tag/.

Silva, Carlos. 2024(c). “What Robots.Txt Is & Why It Matters for SEO.” Semrush. Retrieved at https://www.semrush.com/blog/beginners-guide-robots-txt/.

Singh, Nitish, and Arun Pereira. 2005. *The Culturally Customized Web Site: Customizing Web Sites for the Global Marketplace*. N.p.: Elsevier Butterworth-Heinemann.

Solntseva, Daryana. 2020. “SEO per il blog: suggerimenti per ottimizzare i tuoi articoli.” Semrush. Retrieved at https://it.semrush.com/blog/seo-per-il-blog/.

Soulo, Tim. 2023. “Link Building for SEO: The Beginner's Guide.” Ahrefs. Retrieved at https://ahrefs.com/blog/link-building/.

Southern, Matt. 2021. “Google: Embedded Videos Have Same SEO Value as Uploaded Content.” search engine Journal. Retrieved at https://www.searchsearch enginejournal.com/google-embedded-videos-have-same-seo-value-as-uploaded-content/394216/.

Southern, Matt. 2023. “Google's John Mueller On Domain Selection: gTLDs Vs. ccTLDs.” search engine Journal. Retrieved at https://www.searchsearch enginejournal.com/googles-john-mueller-on-domain-selection-gtlds-vs-cctlds/488501/.

Southern, Matt. 2024. “Google E-E-A-T: What Is It & How To Demonstrate It For SEO.” search engine Journal. Retrieved at https://www.searchsearch enginejournal.com/google-e-e-a-t-how-to-demonstrate-first-hand-experience/474446/.

Stox, Patrick. 2023(b). “What Are Core Web Vitals (CWVs) & How To Improve Them.” Ahrefs. Retrieved at https://ahrefs.com/blog/core-web-vitals/.

Stox, Patrick. 2023(c). “What Is Cumulative Layout Shift (CLS) & How To Improve It.” Ahrefs. Retrieved at https://ahrefs.com/blog/cumulative-layout-shift-cls/.

Stox, Patrick, Michal Pecánek, and John Mueller. 2023(a). “Google Uses 20 Canonicalization Signals.” Ahrefs. Retrieved at https://ahrefs.com/blog/canonicalization/.

Tagliaferro, Luca. 2018. “4 UX Analysis Principles To Get Your SEO Started.” Semrush. Retrieved at https://www.semrush.com/blog/4-ux-analysis-principles-for-seos/.

Tan, Priscilla, and Mateusz Makosiewicz. 2023. “Content Optimization: The Complete Guide.” Ahrefs. Retrieved at https://ahrefs.com/blog/content-optimization/.

Tan, Siew Ann. 2023. “What Is Branded Search? (And How to Optimize Your Site for It).” Semrush. Retrieved at https://www.semrush.com/blog/branded-search/.

Tan, Siew Ann. 2024. “Error 404: What It Is, What It Impacts, and How to Fix It.” Semrush. Retrieved at https://www.semrush.com/blog/what-does-error-404-not-found-mean/.

Tixier, Maud. 2005. “Globalization and localization of contents: Evolution of major Internet sites across sectors of industry.” *Thunderbird International Business Review* 47 (1): 15-48. https://doi.org/10.1002/tie.20039.

Tober, Marcus. 2022. “Zero-clicks Study.” Semrush. Retrieved at https://www.semrush.com/blog/zero-clicks-study/.

Toonen, Edwin. 2018. “What is a content delivery network (CDN)?” Yoast. Retrieved at https://yoast.com/what-is-a-cdn/.

Toonen, Edwin. 2018(a). “First things first: writing content with the inverted pyramid style.” Yoast. Retrieved at https://yoast.com/inverted-pyramid/.

Toonen, Edwin. 2022. “What is an XML sitemap and why should you have one?” Yoast. Retrieved at https://yoast.com/what-is-an-xml-sitemap-and-why-should-you-have-one/.

Toonen, Edwin. 2023(b). “How to create the right meta description.” Yoast. Retrieved at https://yoast.com/meta-descriptions/.

Toonen, Edwin. 2023(d). “What are breadcrumbs? Why are they important for SEO?” Yoast. Retrieved at https://yoast.com/breadcrumbs-seo/.

Toonen, Edwin. 2023(a). “Internal linking for SEO: Why and how?” Yoast. Retrieved at https://yoast.com/internal-linking-for-seo-why-and-how/.

Toonen, Edwin. 2023(c). “Image SEO: Optimizing images for search engines.” Yoast. Retrieved at https://yoast.com/image-seo.

van de Rakt, Marieke. 2016. “Use Yoast SEO to make your content findable.” Yoast. Retrieved at https://yoast.com/using-yoast-seo-make-content-findable/.

van de Rakt, Marieke. 2017. “What are user signals?.” Yoast. Retrieved at https://yoast.com/what-are-user-signals/.

van de Rakt, Marieke. 2018. “The context of internal links.” Yoast. Retrieved at https://yoast.com/the-context-of-internal-links/.

van de Rakt, Marieke. 2021(b). “How does video SEO work?” Yoast. Retrieved at https://yoast.com/how-does-video-seo-work/.

van de Rakt, Marieke. 2021. “Does readability rank? On ease of reading and SEO.” Yoast. Retrieved at https://yoast.com/does-readability-rank/.

Vermeer, Hans Josef. 1978. “Ein Rahmen für eine allgemeine Translationstheorie.” *Lebende Sprachen* 23 (3). 10.1515/les.1978.23.3.99.

Vora, Ankit. 2023. “Organic vs. Paid Marketing: What Are the Differences?” Semrush. Retrieved at https://www.semrush.com/blog/organic-vs-paid-marketing/.

Voronina, Daria. 2018. “Geo-targeting: come avere campagne PPC locali di successo.” Semrush. Retrieved at https://it.semrush.com/blog/geo-targeting-suggerimenti-pratici-per-strategie-ppc-locali/.

Walsh, Shelley. 2023. “What Are Rich Snippets & How to Get Them in 2024.” Semrush. Retrieved at https://www.semrush.com/blog/rich-snippets/.

Wilkinson, Marilyn. 2023. “What Is PPC? The Definitive Guide to Pay-Per-Click Marketing.” Semrush. Retrieved at https://www.semrush.com/blog/what-is-ppc/.

Wilkinson, Marilyn. 2024. “Keyword Strategy in SEO: What It Is & How to Create One.” Semrush. Retrieved at https://www.semrush.com/blog/keyword-strategy/.

Wruck, Raoul. 2018. “An In-Depth Look At International SEO (+Step-By-Step Instructions).” Semrush. Retrieved at https://www.semrush.com/blog/an-in-depth-look-at-international-seo/.

Zarudnyi, Andrew. 2023. “SEO HTML Tags You Can't Do Without in 2024.” SE Ranking. Retrieved at https://seranking.com/blog/html-tags-in-seo/.

Zerkalenkov, Zhenya. 2024. “Google SEO: Best Practices to Help with Search Ranking.” Semrush. Retrieved at https://www.semrush.com/blog/google-seo/.

Zhu, Pinfan. 2009. “Language Problems to Be Coped with in Web Localization.” *Journal of Technical Writing and Communication* 39 (1): 57-78. https://doi.org/10.2190/TW.39.1.e.

Web resources

“Algorithm Definition & Meaning.” n.d. Merriam-Webster. Accessed March 29, 2024. https://www.merriam-webster.com/dictionary/algorithm.

“Another step to reward high-quality sites.” 2012. Google Search Insight. Accessed May 7, 2024. https://search.googleblog.com/2012/04/another-step-to-reward-high-quality.html.

“Application software | Definition, Examples, & Facts.” 2024. Britannica. Accessed March 29, 2024. https://www.britannica.com/technology/application-software.

“Biggest social media platforms by users 2024.” 2024. Statista. Accessed May 4, 2024. https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.

"China search engines market share 2024." 2024. Statista. Accessed March 29, 2024. https://www.statista.com/statistics/253340/market-share-of-search-search engines-in-china-pageviews/.

"CSS Introduction." n.d. W3Schools. Accessed March 29, 2024. https://www.w3schools.com/css/css_intro.asp.

"Deep web | Definition, search engines, & Difference from Dark Web." 2024. Britannica. Accessed March 29, 2024. https://www.britannica.com/technology/deep-web.

"Geo-targeting: What Is Geo-targeting?" n.d. WordStream. Accessed July 24, 2024. https://www.wordstream.com/geo-targeting.

"Global search engine desktop market share 2024." 2024. Statista. Accessed July 24, 2024. https://www.statista.com/statistics/216573/worldwide-market-share-of-search-search engines/.

"HTML basics - Learn web development | MDN." 2024. MDN Web Docs. Accessed March 29, 2024. https://developer.mozilla.org/en-US/docs/Learn/Getting_started_with_the_web/HTML_basics.

"Installed base of smart speakers worldwide in 2020 and 2024." 2020. Statista. Accessed June 7, 2024. https://www.statista.com/statistics/878650/worldwide-smart-speaker-installed-base-by-country/.

"Internet and social media users in the world 2024." 2024. Statista. Accessed March 29, 2024. https://www.statista.com/statistics/617136/digital-population-worldwide/.

"Introduction to Web Accessibility | Web Accessibility Initiative (WAI)." n.d. W3C. Accessed June 7, 2024. https://www.w3.org/WAI/fundamentals/accessibility-intro/.

"Spam Policies for Google Web Search | Google Search Central | Documentation." n.d. Google for Developers. Accessed September 4, 2024. https://developers.google.com/search/docs/essentials/spam-policies.

"State of Content Marketing 2023 Global Report." 2023. Semrush. Accessed May 4, 2024. https://lp.semrush.com/rs/519-IIY-869/images/StateofContentMarketing2023.pdf?_ga=2.162228415.2087421412.1714481045-1270565458.1714481045.

"Understanding Google Page Experience | Google Search Central | Documentation." n.d. Google for Developers. Accessed May 4, 2024. https://developers.google.com/search/docs/appearance/page-experience.

"WCAG 101: Understanding the Web Content Accessibility Guidelines." 2021. WCAG. Accessed June 7, 2024. https://www.wcag.com/resource/what-is-wcag/.

"World Wide Web | History, Uses & Benefits." 2024. Britannica. Accessed March 29, 2024. https://www.britannica.com/topic/World-Wide-Web.

Riassunto

SEO Internazionale: Ottimizzazione per i motori di ricerca e la sua applicazione ai siti web multilingua e multi-mercato

Sin dalla sua invenzione negli anni Novanta del secolo scorso, il World Wide Web ha assunto un ruolo sempre più centrale nelle nostre vite. Accessibile attraverso una moltitudine di dispositivi tecnologici, nel corso degli ultimi decenni internet di è trasformato in una fonte da consultare ogni qual volta ci sorga un dubbio o emerga una necessità. La sua centralità nelle attività umane quotidiane è una tendenza che va di pari passo con l'aumento esponenziale di risorse web che mirano a rispondere ai bisogni degli utenti e a trasformarli in opportunità di business. Oggi, la sfida per i webmaster, ossia coloro che gestiscono e amministrano siti web, è fare sì che il proprio sito raggiunga e attragga le persone giuste, coloro che hanno bisogno dei loro contenuti e servizi e che possono essere trasformate in clienti. Un obiettivo, questo, che può essere raggiunto mettendo in atto strategie di perfezionamento del sito stesso in modo da promuovere la fruibilità. E' in questo contesto che si colloca l'ottimizzazione per i motori di ricerca, (SEO, search engine Optimization), una disciplina nata con l'avvento massivo di internet e che è oggi uno degli strumenti più tattici del marketing digitale. La caratteristica principale della SEO, che la distingue da altre forme di online marketing, è il fatto che essa si pone come obiettivo quello di incrementare il traffico organico, cioè aumentare il numero di visitatori di una pagina che arrivano su un sito senza che questo sia di fatto pubblicizzato, bensì selezionandolo dalla lista di pagine che gli vengono proposte "spontaneamente" dal web. A questi si vanno a contrapporre i risultati sponsorizzati, che rientrano nella categoria degli annunci pubblicitari per i motori di ricerca (SEA, search engine Advertising), strumento complementare nella costruzione di questo tipo di marketing (SEM, search engine Marketing). Nonostante si differenzino per meccanismi, costi e tempistiche, tanto la SEO quanto la SEA sono risorse fondamentali nella gestione strategica della presenza online di un'attività, che permetta di essere competitivi e guadagnare una posizione di vantaggio rispetto alla concorrenza.

Collocandosi nel grande mondo del marketing digitale, la SEO si basa sostanzialmente su internet e sui motori di ricerca, per cui capirne il funzionamento è essenziale per operare in questo ambito. Il web è concepibile come una enorme biblioteca contenente innumerevoli libri (i siti web) con altrettanto innumerevoli pagine (le pagine web). Questo archivio estremamente vasto è ricercabile dagli utenti interfacciandosi con un browser, ad esempio Google Chrome, che a fronte della richiesta effettuata, la inoltra al motore di ricerca, in questo caso Google, che infine recupera e restituisce la risorsa sollecitata. Per far sì che una pagina sia consultabile, è tuttavia prima necessaria che essa venga registrata nelle banche dati di Google; questo avviene in due fasi: l'analisi (o crawling) e l'indicizzazione (o indexing). Durante il crawling, degli appositi robot esplorano il web seguendo i collegamenti link da una pagina all'altra e ne analizzano il contenuto; successivamente, queste pagine vengono catalogate e registrate nell'indice di Google per poter poi essere visitate dagli utenti.

Generalmente, tuttavia, l'utente non richiede una pagina specifica, ma consulta il web inserendo nella barra di ricerca delle termini; queste parole vengono poi utilizzate da Google come filtri per selezionare le risorse, tra quelle presenti nella banca dati, che reputa più utili per la ricerca effettuata e mostrare nella SERP, la pagina dei risultati di ricerca. Una terza fase è quindi costituita dalla effettiva creazione di un elenco ordinato delle pagine da proporre per la consultazione. Questi risultati possono essere presentati in una moltitudine di formati diversi, che vanno dal più semplice e comune link blu alle cosiddette "SERP features", ossia risultati di varia misura e funzione messi in evidenza e integrati da informazioni aggiuntive. L'ordine di classificazione dei risultati sulla SERP (in gergo, "ranking") è di primaria importanza per la SEO: i risultati che appaiono per primi, infatti, sono quelli che godono di maggiore visibilità e quindi maggiori visite da parte degli utenti. Il ranking è determinato, fondamentalmente, dal principio cardine del funzionamento degli algoritmi dei motori di ricerca, che è la rilevanza: più il contenuto di una pagina è rilevante alle necessità dell'utente, più alta sarà la sua posizione nella SERP. Questa accezione di rilevanza, tuttavia, si regge su una serie di diversi fattori tanto interni quanto esterni al sito che concorrono a renderlo altamente performante e affidabile, e che rientrano nel raggio di azione della SEO.

Solitamente, viene fatta una distinzione di tre macroaree: ottimizzazione on-page, tecnica e off-page.

La on-page optimization, che consiste nell'ottimizzazione di elementi all'interno della pagina web. Il pilastro di questa branca della SEO sono le parole chiave, in gergo, "keywords"; esse sono termini che fungono sostanzialmente da punti di contatto tra il brand e il consumatore, e devono essere strategicamente inseriti nei contenuti pubblicati per fare sì che, quando digitati dall'utente, lo conducano alla pagina ottimizzata. Di fondamentale importanza è, quindi, saper individuare quali parole sono utilizzate dal proprio pubblico target per ricercare il servizio che si offre, in modo da poterle integrare in maniera chiara e descrittiva nella struttura e nel contenuto della pagina. Questo è necessario per consentire ai robot (spesso semplicemente "bot") di Google che analizzano la pagina durante il crawling di comprendere più facilmente di cosa si tratta e per dimostrare il valore del contenuto per l'utente. Conoscere a fondo il proprio target, ed è pertanto cruciale per soddisfare i suoi desideri e bisogni, nonché per apparire rilevanti agli occhi del motore di ricerca. Ciascuna ricerca è, infatti, mossa da un intento, che, a seconda di dove si collochi l'utente nel suo customer journey, può essere, ad esempio, informativo o commerciale e che determina il tipo di contenuto che l'utente sta cercando. E' il contenuto, in ultima analisi, a farla da padrone nel determinare il successo di un sito, contribuendo ad elevarne la reputazione, l'autorevolezza e la fiducia dei visitatori: gli algoritmi usati da Google sono sempre più sofisticati ed efficaci nel valutare la naturalezza e l'originalità dei testi, e sulla base di un modello chiamato E-E-A-T (esperienza competenza, autorevolezza e attendibilità) ne determinano la qualità, quindi il ranking. Da un punto di vista tecnico, è certamente essenziale progettare un'infrastruttura solida, che permetta una navigazione fluida. Partendo dai codici che compongono un sito web, quali HTML, CSS e JavaScript, che sono utilizzati rispettivamente per assemblarlo, curarne l'estetica e renderlo dinamico e interattivo, è fondamentale sviluppare ogni componente in modo che sia ben organizzata e priva di stringhe inutili o non ottimizzate, ad esempio contenenti spazio vuoti dove questi non servono oppure immagini o video di grandi dimensioni. Questo permette di rendere più efficiente il processo di reperimento dell'informazione, poiché, essendo file che devono

essere scaricati dal browser per poter essere visualizzate dall'utente, dei file cosiddetti pesanti richiedono generalmente più tempo, una risorsa preziosa sia per le persone che per i motori di ricerca. Molte delle pratiche di ottimizzazione tecnica vertono sulla riduzione della velocità di loading della pagina, un aspetto considerato di enorme importanza dagli algoritmi Google per assicurare una buona user experience: qualsiasi potenziale attrito potrebbe essere motivo di frustrazione per l'utente e spingerlo ad abbandonare la navigazione, di fatto vanificando ogni sforzo. Sebbene la velocità non sia di per sé una metrica, essa viene misurata attraverso altri indicatori, tra cui Core Web Vitals, che misurano i tempi di caricamento, la stabilità e la reattività del sito e che quindi ne riflettono la velocità di risposta. I CWV rappresentano alcuni dei fattori di ranking più influenti per Google, e che hanno quindi un peso significativo sul traffico organico.

L'ottimizzazione si concentra poi su design del sito e della pagina, fondamentale anch'essa per garantire un'esperienza immersiva e fluida ai visitatori. Si ha in primo luogo la creazione di una struttura organizzativa che raccolga le varie sezioni e contenuti secondo una logica intuitiva e ordinata, al fine di facilitare il percorso dell'utente e persino orientare verso/a compiere le azioni di conversione desiderate, ad esempio agevolando la finalizzazione di un acquisto o la compilazione di un form di iscrizione ad una newsletter. Parte di questo processo di semplificazione dello user journey implica anche ridurre al minimo il numero di passaggi che l'utente deve compiere per arrivare dove desidera; una buona website architecture si sviluppa orizzontalmente piuttosto che verticalmente, e si serve di vari elementi di supporto alla navigazione come la paginazione, filtri e i breadcrumb. La paginazione permette di suddividere in pagine numericamente ordinate contenuti come cataloghi di prodotto o archivi di documenti, e permette in questo modo di muoversi più rapidamente all'interno di un elenco di risorse. Elenco che può poi essere ridotto e adattato alle esigenze del singolo utilizzando dei filtri che riducono la selezione di elementi mostrati e raffinino così la ricerca. I breadcrumb, infine, tengono traccia del percorso svolto dall'utente e lo localizzano all'interno del sito, così che egli possa in ogni momento ritornare sui suoi passi o avere un punto di riferimento con cui orientarsi.

Un ruolo fondamentale per la navigazione è, infine, svolto dai link. Ogni risorsa su internet è dotata di un URL, Uniform Resource Locator, ossia una stringa di testo che funge da indirizzo per localizzarla nella rete e per poterla recuperare quando questa viene richiesta dall'utente. Quando un URL viene inserito all'interno del codice HTML di una pagina, viene creato un link, cioè un collegamento, da quella stessa pagina alla risorsa, che è spesso un'altra pagina. Nella website architecture, l'internal linking, cioè i collegamenti interni ad un medesimo sito, permette di interconnettere le varie parti dello stesso e consente al visitatore di muoversi più liberamente. L'internal linking, inoltre, previene il rischio di avere pagine cosiddette "orfane", pagine analizzate e indicizzate dai bot e pertanto presenti sul web, ma che sono sono collegate ad alcuna altra risorsa; questo significa che possono essere sì visitabili, ma solamente digitando il loro URL nella stringa di ricerca, e che restano così, di fatto, virtualmente irraggiungibili dall'utente.

Il valore dei link si estende poi anche al di fuori del sito, prendendo il nome di backlink. In questo caso, le risorse menzionate fanno riferimento a contenuti presenti su pagine di altri siti, che vengono linkate perché ritenute utili e di qualità. Il backlink building, infatti, cioè il guadagnarsi menzioni da altri creatori online è uno degli obiettivi principali della SEO e rappresenta l'attività cardine dell'ottimizzazione off-page. Questa branca della SEO mira ad accrescere la brand awareness e a promuovere l'immagine a nuovi segmenti di pubblico. Ottenere link da altri siti, infatti, significa raggiungere nuovi visitatori che non sono stati acquisiti attraverso la SERP, e che sono molto probabilmente visitatori qualificati, cioè in linea con il target del contenuto linkato. Questo è deducibile poiché, logicamente, una certa risorsa verrà menzionata da una pagina (e quindi un sito) che tratta di un argomento simile o complementare, e che quindi sarà di interesse per i lettori. La maggiore esposizione data dai backlink non si limita al traffico organico, bensì si traduce anche in un miglior ranking. Come menzionato prima, un contenuto viene linkato dai creatori di altri siti quando questi lo considerano di qualità e ne vedono il valore per i propri visitatori: si tratta di un segnale molto potente per Google che dimostra l'autorevolezza e l'attendibilità della risorsa. Ciascuna menzione funziona come un vero e proprio voto di fiducia, che prende il nome

di link equity. Il valore che viene trasferito varia in base a diversi fattori, tra cui la quantità e alla qualità dei backlink, intesa come autorità del sito che fornisce il link e la rilevanza del contenuto, che non solo deve essere pertinente ma deve anche inserire la menzione in maniera naturale e armoniosa, e non artificiosa e forzata, pratica penalizzata dagli algoritmi. In passato operazioni illecite di acquisizione di link come il loro acquisto in cambio di denaro o l'inserimento non autorizzato su siti altrui sfruttando delle vulnerabilità di sistema, erano pratiche diffuse, volte a manipolare il ranking e motivate dal loro valore strategico ma anche dalla loro difficoltà di ottenimento, dato che spesso richiede molto tempo e impegno creare contenuti meritevoli di menzione. A partire dalla seconda metà degli anni duemila, Google ha imposto regole sempre più rigide al link spamming e implementato controlli sempre più stringenti granzie allo sviluppo di algoritmi sempre più sofisticati che utilizzano tecnologie di elaborazione del linguaggio naturale e che sempre di più impiegano l'intelligenza artificiale ed il machine learning. I modi per ottenere backlink in maniera legittima sono molti, e si sovrappongono ampiamente con l'ambito delle relazioni pubbliche digitali, o digital PR. Così come nel mondo offline, una delle tecniche più efficaci di promozione e pubblicizzazione di un prodotto sono le relazioni con il pubblico e con altre aziende del settore. Questi incontri prendono spesso la forma di collaborazioni o sponsorizzazioni, che permettono di ricevere maggiore visibilità e di generare click e traffico, aumentando così la popolarità del brand. Analogamente alle PR tradizionali, le iniziative consistono nella creazione di eventi, nelle relazioni celebrità e i media, interventi editoriali, sebbene declinati in chiave "digital": eventi online, influencer, social media, blog. Talvolta sono attività di marketing a generare in maniera spontanea la promozione dei media, come ad esempio il lancio di un nuovo prodotto o la creazione di un "linkable asset", un contenuto altamente valido come ad esempio strumenti, guide specializzate, report che attirano l'attenzione dei giornalisti proprio per la loro estrema utilità; spesso, tuttavia, è necessario agire proattivamente nella promozione del proprio brand. Nel vasto mondo digitale, i brand possono servirsi di piattaforme dedicate che mettono in contatto media e creator per favorire lo sviluppo di rapporti lavorativi sulla base di interessi comuni. La più celebre tra queste, sebbene ormai sorpassata da altre alternative

più efficienti, è HARO, Help A Reporter Out, in cui i giornalisti o bloggers possono pubblicare un annuncio sul loro prossimo pezzo editoriale e accogliere proposte di collaborazione da esperti in cerca di pubblicità.

Nonostante facciano riferimento a concetti diversi e indipendenti, digital PR e backlink building sono termini spesso usati in maniera intercambiabile, poiché, come è logico che sia, le iniziative di relazioni pubbliche portano spesso all'ottenimento di menzioni. Altre volte, tuttavia, queste non vengono considerate ai fini del ranking, come è il caso delle campagne su social media. Le comunità online rappresentano ormai uno degli asset più preziosi per l'engagement di un brand, ma i link generati da queste piattaforme sono spesso nofollow, il che significa che non considerati come backlink da Google. Pur non contribuendo direttamente al miglioramento del posizionamento SEO, queste attività sono di sono essenziali per rafforzare la presenza online, migliorare la percezione del marchio e stimolare conversazioni tra gli utenti, e tendono comunque influenzare positivamente il traffico organico e le conversioni nel lungo termine. In questo modo, si evidenzia come le attività di digital PR possano avere un impatto più ampio rispetto alla semplice acquisizione di backlink, contribuendo alla strategia di marketing digitale in modo olistico.

Le pratiche di ottimizzazione SEO sono, per loro natura, estremamente dipendenti dal contesto in cui si colloca la risorsa da ottimizzare. Gli algoritmi dei motori di ricerca, infatti, sono fortemente basati su un principio di personalizzazione: essendo il loro obbiettivo quello di presentare all'utente i risultati più in linea con le sue esigenze, questi cambiano in maniera più o meno marcata a seconda di una serie di fattori individuali. Tra questi, quelli tra i più determinanti sono la lingua e la posizione geografica. Come qualsiasi altro strumento di marketing, i canali online si devono confrontare con mercati diversi, fatti di indici di con tradizioni, usi e costumi diversi. La SEO internazionale è oggigiorno una parte integrante della strategia di business di qualsiasi brand orientato verso un pubblico mondiale. La globalizzazione ha significato per i brand la possibilità di espandere la propria sfera di influenza e di raggiungere nuovi target, ma ha anche implicato la necessità di sbarcare su nuovi mercati per poter

rimanere competitivi. La SEO internazionale va ad aggiungere alle varie pratiche di ottimizzazione una serie di accortezze e integrazioni essenziali per riuscire a catturare il pubblico estero. Il primo passo nel processo di internazionalizzazione è condurre ricerche di mercato approfondite per individuare quali paesi rappresentino un'opportunità, e decidere in che modo approcciarsi; generalmente, infatti, le aziende dispongono di risorse limitate che impongono di scegliere in che modo come investire, sulla base dei costi e dei potenziali guadagni. Le strategie di espansione possono essere varie a seconda dell'impegno dedicato a ciascun paese; ai due estremi di questo spettro, si collocano i due tipi fondamentali di internazionalizzazione di un sito web: un sito multilingua o un sito multimercato. I siti multilingua, come suggerisce il nome, prevedono essenzialmente la traduzione delle pagine in un altra lingua per poter servire un nuovo pubblico; essi richiedono un minore investimenti sia di tempo che denaro, ma sono, d'altra parte, anche meno targhettizzati e redditizi. I siti multimercato, invece, sono progettati per adattarsi in maniera più mirata ad una certa nicchia, il che implica maggiore attrattività per il target ma anche più impegno. Per ciascun nuovo mercato è necessario infatti studiare come i nuovi utenti si comportano online, come navigano, quali parole chiave usano, come fruiscono dei contenuti. Questo permette di impostare l'ottimizzazione on-page, integrando le keywords alla struttura delle pagine e creando contenuti di valore per il proprio target. Questo processo, in particolare, risulta cruciale per la la SEO internazionale, poiché non si tratta di "mera" traduzione di testi bensì di localizzazione, cioè di adattamento dei contenuti alle caratteristiche culturali specifiche di ciascun mercato. Questo adattamento riguarda elementi più formali, come valuta, misure, recapiti, note legali, ma anche colori, simboli, raffigurazioni, fino ad arrivare al messaggio stesso. Come concettualizzato dal sociologo Geert Hofstede, ciascuna cultura può essere pensata come un insieme di tratti che contribuiscono a definire la struttura sociale e che riflettono i valori e le credenze alla base. Queste differenze, unitamente a differenze dovute al contesto economico e geografico, formano sensibilità diverse, abitudini di consumo diverse, interessi diversi, che implicano un lavoro di localizzazione delle risorse indirizzate al pubblico per poterlo raggiungere e comunicare il proprio brand in maniera efficace. Per farlo, non è sufficiente quindi tradurre un

medesimo sito in varie lingue, bensì è necessario sviluppare nuove versioni da gestire sostanzialmente come siti a sé, curando in maniera indipendente la creazione di contenuti ottimizzati ma anche costruendo relazioni di valore con media e celebrità online locali per ricevere preziosi backlink, e incrementare così il ranking e l'autorevolezza in quello specifico mercato. Dal punto di vista tecnico è infine indispensabile comunicare ai motori di ricerca quale variante sia destinata a quale pubblico, sulla base sia della posizione geografica che delle preferenze individuali. Questo avviene integrando degli indicatori di lingua e di paese nei codici HTML delle pagine e negli URL, che, confrontandosi con le informazioni provenienti dal dispositivo e dalle impostazioni browser dell'utente, permette di targhettizzare le pagine in maniera accurata, evitando disallineamenti e ottimizzando le risorse.

Applicare queste tecniche di ottimizzazione è oggigiorno un'esigenza di mercato per poter rimanere competitivi e operanti. Tuttavia, è importante considerare che la SEO non è una disciplina statica. Con l'evoluzione continua del panorama digitale, le strategie SEO devono adattarsi ai progressi tecnologici, ai cambiamenti degli algoritmi dei motori di ricerca, ai comportamenti degli utenti. In futuro, l'integrazione dell'intelligenza artificiale e del machine learning diventerà sempre più significativa, sottolineando ulteriormente l'importanza di creare contenuti orientati all'utente e di mantenere elevati standard tecnici. Per le aziende che desiderano espandere la loro presenza globale, la SEO internazionale continuerà ad essere un elemento cruciale della loro strategia digitale, e in un mondo sempre più interconnesso, la capacità di raggiungere e coinvolgere efficacemente il pubblico in diverse regioni e lingue sarà un fattore determinante per il successo.