CALIFORNIA STATE UNIVERSITY, NORTHRIDGE

AN ANALYSIS OF THE APPLICATION OF SELECTED SEARCH ENGINE OPTIMIZATION (SEO) TECHNIQUES AND THEIR EFFECTIVENESS ON GOOGLE'S SEARCH RANKING ALGORITHM

A thesis submitted in partial fulfillment of the requirements
For the degree of Master of Science
In Computer Science

By

Edgar Damian Ochoa

May 2012

The thesis of Edgar Damian Ochoa is approved:

Gloria Melara, Ph.D.	Date
John Noga, Ph. D.	Date
Rick Covington, Ph. D., Chair	Date

California State University, Northridge

Acknowledgements

To Professor Covington, who through his useful feedback and guidance, helped me take this thesis from a “rock” into a polished piece of art. Thank you.

Dedication

I dedicate this thesis to my mother and father. Although I can't bring back all those long hours that I spent away from you, I hope it'll be worth it. To my sisters. Thank you for all your support. To my brothers. I hope I can make you proud. And finally to Chiquiya. Thank you for being there through the stressful times. You're my rock.

Table of Contents

List of Figures

List of Tables

ABSTRACT

AN ANALYSIS OF THE APPLICATION OF SELECTED SEARCH ENGINE OPTIMIZATION (SEO) TECHNIQUES AND THEIR EFFECTIVENESS ON GOOGLE'S SEARCH RANKING ALGORITHM

By

Edgar Damian Ochoa

Master of Science in Computer Science

Due to the exponential growth of the Internet in recent years, search engines have the complex task of sorting through billions of pages and displaying only the most relevant pages for the submitted search query. Google has become an essential link between people and the information they seek online. For this reason, any webmaster or search engine optimization (SEO) engineer should be actively learning the techniques that drive visitors to their site.

This paper describes the application of selected Search Engine Optimization (SEO) techniques to a newly created website across the entire website lifecycle, from inception, through development, and finally to launch and optimization of the site. Several SEO experiments were defined and evaluated by collecting and analyzing real traffic and visitor data through the use of Google Analytics. This research further analyzes search engine ranking factors and their effectiveness on Google's search ranking algorithm by analyzing the number of users who visit the site and site rankings. The metrics for effectiveness of the SEO techniques were *Number of Visitors*, *Pageviews* and *Ranking*.

The results of the research confirmed that there was a noticeable increase in the number of users who visited the site and the search engine rankings also increased. The implementation of SEO showed a positive effect on the Google search rankings and the increase in traffic.

The results of this research confirm and extend results of earlier SEO research. This paper provides a thorough analysis and step-by-step implementation of selected search engine optimization techniques that are shown to increase visibility, get more visitors and achieve higher rankings in search results for a general class of website. For this reason, it is hoped that this paper can be used as a guidebook for new SEO engineers and as a basis for later continued SEO research.

Chapter 1 - Introduction

This chapter provides the introduction to my research paper and is structured as follows: Section 1.1 presents the topic of the paper. Section 1.2 covers the purpose and motivation. Section 1.3 describes the target readers. Section 1.4 presents key terms and their definitions. Finally, section 1.5 gives an overview of this research paper's structure and outline.

1.1 Topic

This paper describes the application of selected Search Engine Optimization (SEO) techniques for a website and analyzes its effectiveness in the context of the Google search engine. It covers the entire development lifecycle of a newly created website and the effect the techniques have on the number of users who visit the site. Search engine rankings are also analyzed

1.2 Purpose and Motivation

Searching online has become part of the everyday lives of most people. Whether to look for information about the latest gadget to getting directions to a popular restaurant, most people have made search engines part of their daily routine. Beyond trivial applications, search engines are increasingly becoming the sole or primary source directing people to essential information. For this reason, search engines occupy "a prominent position in the online world" [2]; they have made it easier for people to find information among the billions of web pages on the Internet.

Due to the large number of websites, search engines have the complex task of sorting through the billions of pages and displaying only the most relevant pages in the search engine results page (SERP) for the submitted search query.

With the continued growth of the Internet and the amount of websites available, it has become increasingly difficult for sites looking for an audience to achieve visibility. According to a recent study, there are about 3 million new websites appearing on the Internet every month [4]. As a result of this continued growth, it has made it increasingly difficult for websites to stay visible among all the other competing sites.

Another study found that "more than 80% of first visits to a website come from web search. Of those visits more than 76% use Google's search worldwide" [3]. Furthermore, it shows that "84% of Google searchers never go beyond the second page of the search results, and

65% hardly ever click on paid or sponsored results" [3]. These studies show how achieving top rankings in the search engine results is key to a site's continued visibility. Therefore, getting top positions in the search engine results is critical to the constant flow of users to the websites, and this is where the value of SEO comes in.

In order for search engines to determine the most relevant pages, the search engine algorithm has the daunting task of parsing and analyzing HTML pages in order to categorize them. These steps are needed so that when searchers type their keyword (or search query) into the search engine text box, the pages most relevant to that keyword are displayed.

To bring order to the Internet by helping to categorize web pages and increase their visibility, search engine optimization (SEO) has increased dramatically in popularity in recent years. SEO is the process of optimizing a website by "editing its content and HTML to boost its relevance with the specific keywords" [3] to gain high rankings in the major search engines such as Google, Yahoo and Bing.

Most people are not using traditional guides to "brick and mortar" businesses such as the Yellow Pages anymore. If businesses are not adapting to the rapid changes taking place in the Internet, they're destined to become non-existent. Appearing on the first page of Google can make or break a business. According to one source, "SEO generates around 78% of site traffic" [11] to websites, increasing traffic by a factor of 3.

As the Internet continues to grow, SEO techniques will become essential, as will increased research into these techniques. Google has transitioned from a novelty used by small group of technical insiders to an essential link in the process by which customers and business locate each other.

Figure 1.0 gives a general idea of the critical role search engines play between searchers and websites. In this example, Google is the critical link that connects searchers to different websites and helps them find information. Or in the case of businesses, Google connects potential clients to businesses.

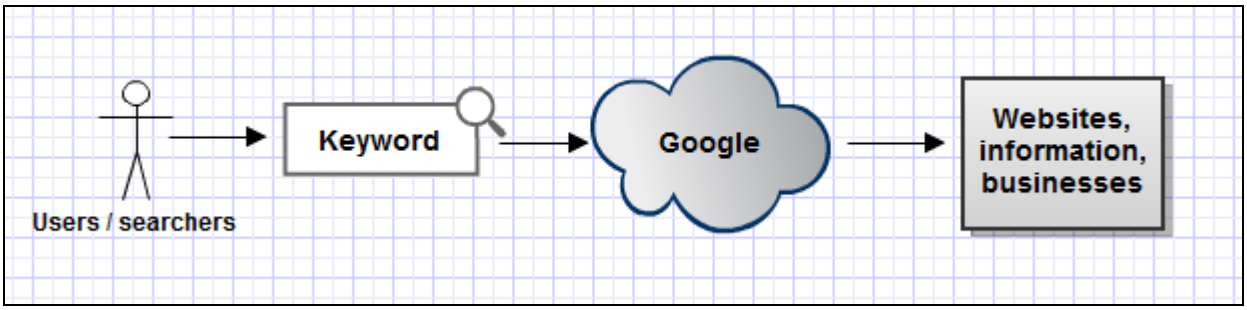


Figure 1.0: Google as the critical link that connects searchers to information.

Statistics show a high level of use of selected online services. A recent survey determined that "in January 2002, 52% of all Americans used search engines. In February 2012 that figure grew to 73% of all Americans. On any given day in early 2012, more than half of adults using the Internet use a search engine (59%). That is double the 30% of internet users who were using search engines on a typical day in 2004. And people's frequency of using search engines has jumped dramatically" [48]. This means that search engines will only become more prevalent in the lives of people looking for information online. Figure 1.1 below shows the online activity of users according to the survey. You can see that search engine use is high; 91% of the people said they use search engines every time they go online.

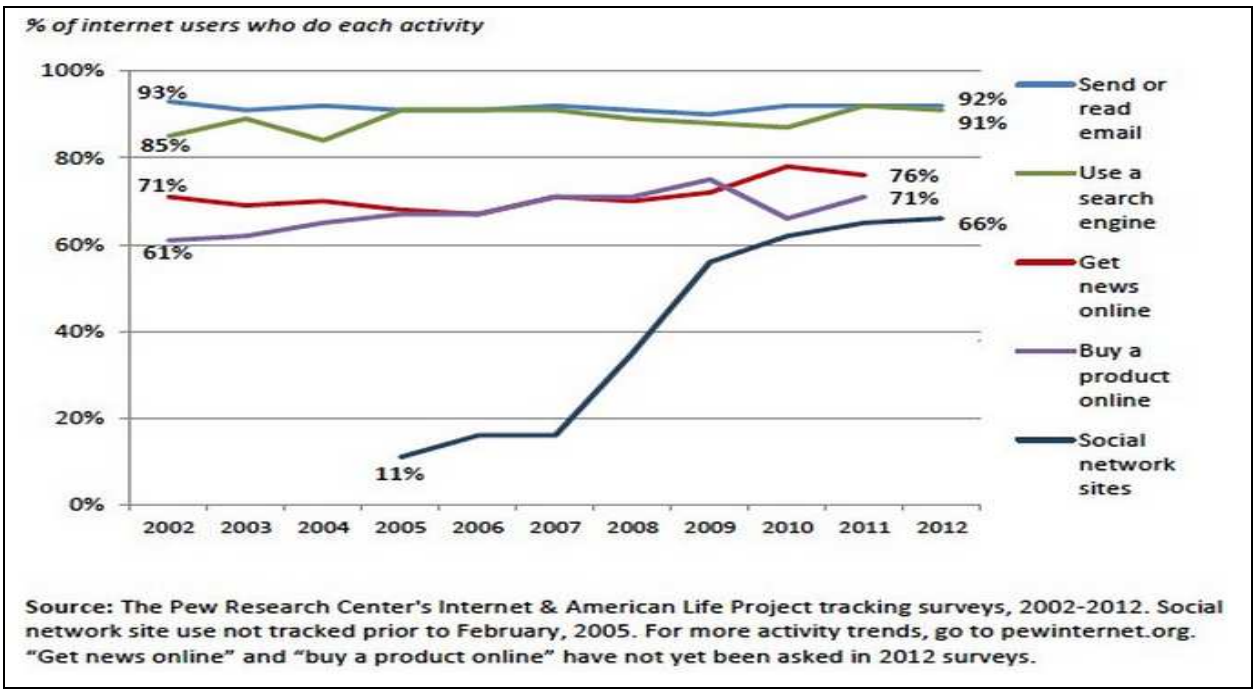


Figure 2.1: % of Internet users who do each activity.

Search engine use is an important subcategory of online services. “Asked how often they use a search engine to find information online, just over half of all search engine users (54%) say they do this at least once a day, a significant increase over 2004” [48]. Figure 1.2 shows additional results of the survey; as you can see, more than half of the people said they used search engines at least once a day, or more.

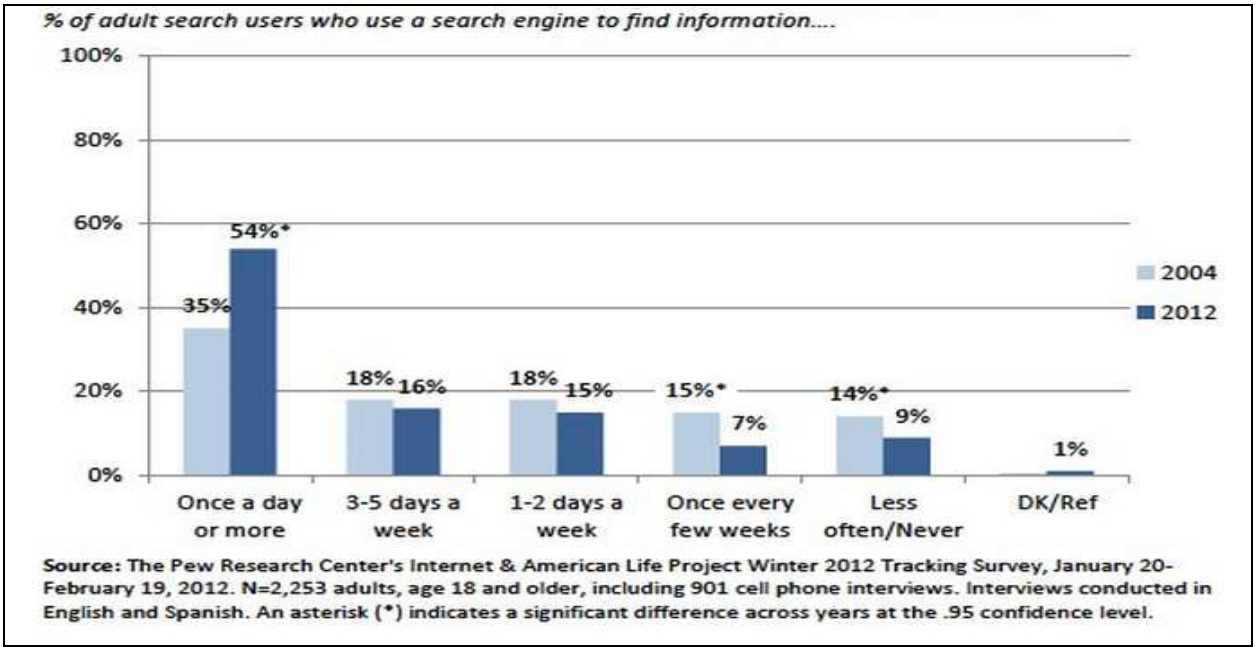


Figure 1.2: % of adult search users who use search engines to find information.

Given the high percentage of people who depend on search engine services, detailed accurate knowledge of SEO techniques will become essential for anyone who depends on the Internet.

In this paper, natural search engine ranking factors and their effectiveness on the Google search engine algorithm will be analyzed. I will use my experience with the creation and development of an actual website as a platform for the trial application of a set of selected SEO techniques. These SEO experiments will be evaluated by collection and analysis of real data from the website through the use of Google Analytics; furthermore, their effectiveness on the number of users who visit the site and search engine rankings will be analyzed.

Whereas other studies have focused on existing sites, examining the most important factors for SEO, and research on paid listings from major search engines, this paper extends that

earlier work by taking important factors (as determined from previous studies) and applies them to the development of a brand new site and analyzes the results.

To help the reader understand the overall process of SEO, this paper describes the entire website cycle, from initial inception, through development, and finally to the launch of the website. It goes through the entire SEO process focusing on the most important factors mentioned on previous research (for an overview, see [3,8,9]). It further analyzes and examines the effectiveness of SEO by taking important factors from previous research and applying them to the new site.

Search engines have become the primary vehicle people use to look for information online. As a result, search engines and their ranking algorithms have generated "great interest to the information science community" [1]. What role should SEO play? With the growth of the Internet and the availability of billions of web pages, competition for the top few positions of the search results is fierce; it may also be impossible for all users to find what they're looking for.

For this reason, search engines aim to display the most relevant pages for the user's search query. Therefore, the goal of SEO engineers is to make sure those key pages get indexed and if possible to get the pages onto the first page of the search results. Most of the popular search engines display 10 results on the first page. Fewer and fewer people go past the first page when looking for something, so if a website is not in the top 10 results, it is effectively invisible.

1.3 Target Readers

This paper focuses on the application of selected SEO techniques and their effectiveness on Google's *PageRank* search ranking algorithm. The results and analysis will be of interest to Computer Scientists because it provides an introduction to the science of information retrieval in Web search. With the increasing amount of information on the Web, this creates new challenges for people searching online and for web search engines striving to provide the most relevant search results.

Moreover, business owners, webmasters, or SEO engineers can further expand their knowledge on the ethical application of SEO methods to help improve their site's relevancy, which can lead higher rankings on the search engine results.

1.4 Key Definitions

Here are some key definitions and basic preliminaries that will be important to understand the ideas presented in this paper:

- **Search Engine Optimization (SEO)**: "SEO is the science of customizing elements of your website to achieve the best possible search engine ranking"[7] when a web user searches on a keyword.
- **Google Search**: Google is "a web search engine owned by Google Inc. Google Search is the most-used search engine on the World Wide Web" [49], which receives about 34,000 searches every second [50]. Its main purpose is to provide users with relevant web pages based on the search query used.
- **Search Engine Results Page (SERP)**: the page that displays a list web pages based on the user's search query. "The results normally include a list of web pages with titles, a link to the page, and a short description showing where the keywords have matched content within the page. A SERP may refer to a single page of links returned, or to the set of all links returned for a search query" [51].
- **Ranking**: the position of the webpage within the search engine results page (SERP).
- **Organic search results**: listings that are generated directly from the search engine's ranking algorithm based on search query relevancy, which the search engine policy promises were attained free from commercial payments.
- **Paid listings**: also known as sponsored listings, these are advertisements that appear adjacent or above the organic search results. These are paid advertisings that are displayed whenever a searcher's "keyword query matches an advertiser's keyword list" [52].
- **PageRank**: the proprietary search ranking algorithm used by Google Search "that assigns a numerical weighting to each element of a hyperlinked set of documents, such as the World Wide Web, with the purpose of 'measuring' its relative importance within the set" [53]; this numerical weight (its PageRank value) indicates the importance or authority of the web page, and it's also a determining factor of a page's ranking on the search results.

- **Alpiste**: the subject of the experimental website being optimized in this paper. *Alpiste*, known in English as canaryseed is primarily used as bird food, but in recent years people have begun to consume it for its potential health benefits.
- **Pay-per-click (PPC)**: it's an advertising model where search users are sent to the advertiser's page via paid listings. Each time a user clicks on any of the paid listings, the advertiser pays a certain amount for each click. An example of a PPC program is Google Adwords (explained next).
- **Google Adwords**: Google's main online advertising platform. It is Google's main source of revenue. Whoever wants to advertise on Google, can create their ads and choose the keywords that are related to their business. Whenever someone looking enters those keywords into the Google search engine, the ads may appear to the right side or the very top of the 'organic search results.'
- **Google Adsense**: A free service for publishers, or website owners, which allows them to earn money by displaying Google ads on their site. Website owners have to ability to choose which ads to display and in what format to display them. For example, the owner of a site about gardening called www.gardeningtips.com can signup with Google Adsense, create ads to be displayed on the site, install the Javascript code in the web pages and in a short time the site will have Google ads being displayed. Whenever a user arrives to the site and clicks on any of the ads, the owner will receive a certain percentage of the cost Google charges the advertiser.
- **Web crawler / bot**: a program that is "mainly used to create a copy of all the visited pages for later processing by a search engine that will index the downloaded pages to provide fast searches" [54] for users searching for information online.
- **Indexed pages**: search engine crawlers collect, parse and store web page data in the index database for use by the search engine to display on the search results. Once a web page data gets stored in the search engine index, the page has been indexed.

- **Keyphrase / Keyword / Search query**: these terms are used interchangeably; it is the word or set of words that a web user enters into the search engine text box for searching.
- **Inbound links / Backlinks / External links**: these terms are used interchangeably; these are links from other sites that point (or link) to your website.
- **Long-tail keywords**: these search queries that contain three or more words; a very specific search for which there is less competition. For example, search queries such as "*roses"* and "*red roses for mother's day"*; the latter would be considered a *long-tail keyword* because it's more specific. More on this subject will be covered in Section 4.3.1.
- **Web-based Content Management System (CMS)**: it's "a bundled or stand-alone application used to create, manage, store, and deploy content on Web pages" [14] such as video, text, images, etc… Examples of web-based CMS platforms are Drupal, Joomla and Wordpress.

1.5 Overview of the Rest of the Paper

This chapter has described the role played by search engines, the importance of SEO techniques, and has presented an overview of the research that was done. Here is an outline of the remaining chapters. Chapter 2 covers previous research done on the subject of SEO, a brief summary on the history of the Internet, the early beginning of Google and the importance of hyperlinks in its proprietary *PageRank* search ranking algorithm, and a brief summary on the early beginnings of SEO; its goals and key ideas such as white-hat SEO, black-hat SEO, on-page SEO and off-page SEO will be covered. Chapter 3 defines the questions and aims addressed by the research, research objectives, and data collection and analysis. Chapter 4 describes in detail the SEO approaches and methods implemented in this research: website topic research, website development, keyword research, data collection and implementation of SEO strategies. Chapter 5 presents the results of selected SEO experiments, and finally Chapter 6 provides some conclusions.

Chapter 2 - Background

This Chapter covers previous research done on the topic of SEO, search engine fundamentals, and a brief overview of the history of the Internet. This will help give a much better understanding of the importance of SEO in the current state of the Internet and in information search. Moreover, an analysis of the importance of becoming relevant and maintaining visibility among a large number of websites will also be discussed.

This Chapter is structured as follows: Section 2.1 will cover related research in the SEO field. Section 2.2 describes important events that took place for the development and history of the Internet, early browsers, the dramatic growth of the Internet and rise of search engines. Section 2.3 provides a brief summary of the beginnings of Google, how it ranks pages using its proprietary *PageRank* algorithm and the importance of hyperlinks in the scheme of the Internet. Section 2.4 discusses the history of SEO and its goals, the different types of SEO implementations (on-page SEO and off-page SEO), and the difference between white-hat SEO and black-hat SEO.

2.1 Related Work

This section analyzes the work of previous research done on the field of SEO. To get a better understanding of the critical role SEO plays in web search, a quick overview will be given on each of the studies, highlighting the importance of the research.

2.1.1 Study and Analysis of Key Influence Web Search Factors

The study in [3] used a reverse engineering approach to "study and analyze the key influence factors in the process of web search." Using this methodology, the researchers determined top five factors for SEO. The researchers developed a system that crawled all website factors (e.g. HTML structure, URL length, etc…) for 200,000 web pages using 10,000 search keywords as their sample set. Moreover, the keywords in the sample set were divided into the following three segments according to their Google search volume in the past three months:

1. ***Hot*** *– high search volume*
2. ***Middle*** *– medium search volume*
3. ***Cold*** *– low search volume*

The objective of this categorization was to discover the "different SEO factors on different segments" [3] of keywords. That is, the researchers were looking to uncover if

keywords with different search volumes (i.e. low search volume, medium search volume and high search volume) required different SEO strategies; it may be case that more competitive keywords (the *Hot segment*), the ones with the highest search volume, required a different SEO approach than the less competitive keywords (the *Cold segment*).

Using an empirical approach to develop a series of analysis, their study determined top five factors for SEO that have the greatest influence in the natural or organic search for increasing high search rankings. Based on their research, they obtained certain SEO rules and provided valuable guidelines for SEO engineers to help improve website rankings. Table 2.1 lists the top five factors from their research.

Rank	Hot	Middle	Cold
1	URL length	URL length	URL length
2	Keyword appear in URL domain	Keyword appear in URL domain	Keyword appear in URL domain
3	Keyword density in H1	Keyword appear in URL path	Keyword density in title
4	Keyword density in title	Keyword density in title	Keyword density in title
5	URL layers	URL layers	Keyword density in H2

Table 2.1: Top 5 SEO factors in study [3].

As you can see from Table 2.1, the study showed *URL length* as the most important factor. Another important factor was the importance of placing the keyword within important web page elements: within the *URL*, *heading tags*, *domain name* and *title tag*. This means that the content of the web page must be relevant to the keyword.

The difference between the SEO factors, among the three segments, may be due to the fact that high search volume keywords are more competitive; thus, they require a whole different SEO strategy. Surprisingly, one thing that was lacking from the paper was that the researchers didn't explain their definition of *URL layers*. Furthermore, *keyword density* refers to the number of times the keyword appears throughout the HTML page, HTML tag or SEO factor such as the URL.

In conclusion, this paper offers valuable insight and suggestions for SEO engineers to follow when optimizing websites; and if implemented correctly, high rankings for specific keywords can be achieved. Similar to my research and as will be seen in Section 4.5, the following three SEO techniques were also implemented: *keyword appearing in the site domain, keyword appearing in the H1 tag, and keyword in the HTML title tag*.

2.1.2 An Empirical Study on the SEO Technique and Its Outcomes

This study is based on use of the Chinese search engine Baidu; Google's *PageRank* algorithm is not considered. Instead, the authors defined a metric (*Page Interest*) and consider whether certain SEO methods have any influence on it. Even though this study doesn't apply to Google, I felt its relevancy was important to analyze similarities from selected SEO factors from different search engines.

Using data collected from 116 websites, the researchers sought to "analyze the impact of SEO techniques" and determine "which technique strategy was more effective" [2]. The following metrics were selected to measure the effectiveness of their SEO methods. The researchers believed these metrics were positive indicators of the SEO implementation:

- **Indexed pages**: the number pages crawled by the search engine bot.
- **Number of independent IP address (IP)**: the number of different IP addresses accessing the web site.
- **Pageview (PV)**: a user request to load a single HTML file.
- **Reach**: the percent of global Internet users who visit the site.
- **Page view per user (PV/U)**: the average number of pages viewed by the total number of visitors to the web site.

Furthermore, the researchers also tested any correlation of SEO techniques on *Page Interest*, an additional metric they derived, which will be explained next.

Page Interest "indicates the interest users show on Page" [2]; this means that the higher the *Page Interest* the higher the preference users will show to a website. According to the authors, *Page Interest* is related to *Pageview* and *Bounce Rate*, which is "the percentage of web surfers who visit websites and quickly leave" [2]. In other words, it's the percentage of page visitors who decide that the page is not relevant to their search query and quickly leave. Thus, a low *Bounce Rate* means users who visit a web page stay there longer because they found what

they were looking for; the keyword used for searching was found to be highly relevant to the content of the web page.

The authors defined a candidate metric, *Page Interest*, and then investigated the effect of several SEO techniques on this metric. Below is the formal definition of *Page Interest (I):*

- **I** = URL.pv x URL.time / URL.bounce
- **URL.pv** = the number of page views on average per day
- **URL.time** = the time users spend on the website (in minutes)
- **URL.bounce** = the bounce rate

From the above definition of *Page Interest*, one can see that if there is a high number of *Pageviews*, an increase of time spent on the website and a low *Bounce Rate*, there will be a high *Page Interest*. As stated previously, a high *Page Interest* can be seen as positive interest the users show on the website; in other words, the website in question is highly relevant to the user's search query.

Below were the six SEO techniques that the researchers implemented and analyzed in their research:

1. **Overall Links** – The total number of web pages linking to another website. This has been a huge determining factor on how search engines determine a site's position in the search results.
2. **Website Title Length** – Most search engines use the title tag on the search results page. Search engines also use the title tag to determine the theme or what the web page is about. Therefore, optimizing the title tag is important.
3. **Keyword Density** – This is defined as the percentage of times the keyword appears on the web page compared to the total number of words on the web page.
4. **Layer Number** – This is "related to the logical structure of a website designed according to the relationship between content relevance and link position" [2].
5. **Page Size** – The researchers defined it as "the sum of the file sizes for all the elements that make up a page" [2]. According to them, "most search engines will not fully index pages that are greater than a certain size" [2]; therefore, the smaller the web page size, the faster it will load.

6. **Customization of 404 Error Pages** – An Error 404 "Page not found" is displayed whenever a visitor requests a web page that no longer exists.

The final results of their study indicated that *Page Size*, *Customization of 404 Error Pages* and *Overall Links* are significant factors in the effectiveness of SEO. But as stated before, it's important to note that this study was focused on Baidu, China's most popular search engine. Although the research was performed on a different search engine, *Overall Links* seemed to coincide with Google's ranking algorithm as an important factor (more detail on this will be covered in Section 2.3.2). Furthermore, *Pageview* and *Bounce Rate* metrics were also used in my study to measure the effectiveness of SEO on my experimental website (more on this will be covered in Chapter 5).

2.1.3 SEO Research Based on Six Sigma Management

The authors of this paper conducted research and empirical analysis to determine what factors had the most positive effect on SEO and proposed a method for its implementation. The goal of the study was to help SEO engineers identify the most influential factors for SEO and how to manage the execution of these strategies by using Six Sigma Management model.

Originally developed by Motorola in the1980s as a business management strategy, Six Sigma "seeks to improve the quality of process outputs by identifying and removing the causes of defects (errors)" [56]; it was interesting to see that the researchers included this process model in their research as a way to execute their SEO methods more effectively. The top five SEO influence factors from their study can be seen in Table 2.2 below.

number	Influence factors
1	The application of keywords in website TITLE
2	The application of keywords in web page content
3	The number of external links
4	The number of backward links
5	The number of paper indexeded（take Baidu as an example）

Table 2.2: Top 5 SEO influence factors in study [55].

Based on their Six Sigma approach and their research, the researchers propose a "basic flow of the website search engine optimization (SEO)" [55] process using the following strategic steps:

1. **Keyword selection and application** – their tests suggests that keyword selection is the "most important factor which influences the search results ranking" [55]. This is similar to my SEO implementation in Section 4.3.
2. **Building external links** – According to their tests, they confirmed that "the number of external links, backward links and websites indexed have a positive correlation with the search results ranking" [55]. This idea is similar to Google's proprietary *PageRank* algorithm which will be covered in Section 2.3.2.
3. **Flow monitoring and search engine analysis** – The key to having a successful SEO implementation is to constantly monitor "the whole website and the flow of each page" [55]. Due to the frequent changes of search engine algorithms, it's important to constantly measure and analyze website data (Google Analytics) to verify any changes in the site rankings in order to "make instant adjustments to ensure rankings" [55].

The above suggestions for more effective SEO do correlate with my research. As stated in Section 2.1.2, although the research was done on Baidu and not on Google, it's important to know the similarities between different search engines and the critical page elements that have the greatest influence on search engine rankings. This knowledge will help SEO engineers understand the different algorithms and compare the most factors used in their ranking algorithm.

2.1.4 How to Improve Your Google Ranking: Myths and Reality

This study focused on the Google's ranking algorithm; the researchers sought to systematically "validate assumptions others have made about this popular ranking algorithm" [8] and identify what page factors or other criteria, had the most influence in its ranking algorithm. They designed and developed a ranking system to determine the most important factors Google uses to rank pages.

Using a reverse engineering approach, the paper showed how their ranking system "can be used to reveal the relative importance of ranking factors in Google's ranking function" [8]. Thus, the paper provides guidelines for SEO engineers on what factors are the most critical for optimizing web pages in order to achieve higher rankings.

Although it has been known that Google uses more than "200 factors" [57] in their search engine ranking algorithm, this study determined a subset of those factors. The researchers' top 5 SEO factors are listed below:

1. **PageRank** – how authoritive is the site (as determined by Google's algorithm)
2. **Domain** – the keyword appearing in the domain name
3. **Title Tag** – the keyword appearing in the title tag
4. **Description Tag** – the keyword appearing in the description tag
5. **URL** – the keyword appearing in the URL

PageRank was the dominant factor of what determines high page rankings (*PageRank* is discussed in more detail in Section 2.3.2). Most of the results of research paper [8] correlate with my findings and implementation of the SEO process. With the exception of *PageRank*, the SEO methods I used were focused on having the keyword within the *domain name*, *title tag*, *description tag* and *URL* of the page being optimized; the details of this will be covered in Section 4.5.

2.1.5 The Application of SEO for Internet Marketing: An Example of the Motel Websites

This study conducted an experiment similar to mine in that they used selected SEO techniques and applied them on a website over the course of a year and then analyzed the effects of the SEO. What was different from my study was that the authors implemented their techniques on an existing website; in my research, I developed a website (from the ground-up), applied selected SEO tactics and then measured the effectiveness of the SEO.

The authors used an existing website *mymotel.com.tw* as a case study to apply the SEO techniques and selected *Janfusum* as their target keyword to optimize the site. The motel is located in southern Taiwan and it's in close proximity to the Jansufum World, a famous amusement park in Taiwan [9].

Table 2.3 below shows the research variables and their definitions that were used as metrics to measure and analyze the effectiveness of the SEO. In my research, *Number of Visits* and *Ranking* variables were also tracked and analyzed; *Pageviews* is another metric that was measured in my study (Section 3.1 goes into more detail about the metrics used).

Research Variables	Operational Definitions
Number of visits	Total number of visits. Single visit to a website and then exit within a period of time. For example, it is counted as one time visit for reading all the information within an hour.
Pages	After deleting the unnecessary files, like pictures, the total number of files provided by the server.
Hits	The total number of files provided by the server.
Bandwidth	Total volumes of transmission
Ranking	After the users put keywords into the search engine, the website ranking order in the search engines shown on the result pages.

Table 2.3: Research variables and operational definitions of study [9].

Below is a list of SEO strategies that were implemented on the existing website (*mymotel.com.tw)* for their target keyword (*Janfusum)*:

1. Keyword was put in the HTML Title tag
2. Added the keyword to the ALT property of the image tags
3. Added keyword to the H1 header tags
4. Registered the website to the DMOZ open website catalog (DMOZ.org)
5. Directly submitted the website to the main search engines: Yahoo, Google and Bing
6. Executed WEB PING to the main search engines. Pinging notifies the search engines that the website has been updated; this increases the chance that the search engines will find and index the pages much faster.
7. Created profiles in popular discussion boards (forums) and put keywords into the signature lines (e.g. *Janfusum*)
8. Created a sub-domain with the keyword in it, http://janfusum.mymotel.com.tw
9. Created an XML sitemap for the search engines. Sitemaps lists all the pages in the website. Sitemaps tell search engines information about your site, how it's structured and how often to index (or crawl) certain pages.

According to their research results, the experimental website (*mymotel.com.tw*) moved higher in Google's search rankings results for the *jansufum* keyword; the bandwidth also increased after applying SEO. The ranking for *mymotel.com.tw* went from the No. 14 position to No. 2 position for their target keyword, and the bandwidth increased as a result of an increase of users who visited to the site.

This study in effect shows the importance of SEO as a way to increase website rankings and traffic. The only back-linking strategy discussed in the study was the creation of profiles in forums and back-linking from the signature. While this may help in the rankings, there are more effective back-linking strategies that can be used and they will be illustrated in Chapter 4.

2.1.6 Summary

This Section briefly summarized previous research done on the field of SEO. As seen from their studies, effective SEO implementation can help websites attain visibility by achieving higher rankings in the search results. Although Google has stated that it uses more than 200 factors in their ranking algorithm [57], there really is no consensus on what the most important

factors are that determine website rankings. As a result, SEO engineers have relied on their experience, or any research performed in this field, to determine the factors for a more effective SEO implementation.

Two of the five papers ([2,55]) discussed the Baidu search engine and sought to identify what factors were important in its ranking algorithm. Although the results were not collected from Google, it's important to analyze any similarities from certain SEO factors between different search engines. In this case, hyperlinks were seen as a critical factor in both search engines for determining high website rankings.

More on the importance of links will be discussed in Section 2.3, but in order to get a better understanding of hyperlinks and the importance search engines play in the current state of the World Wide Web, we need to discuss the history of the Internet and its beginnings.

2.2 History of the Internet

In this Section, a brief summary of history of the Internet and important events that laid the foundation to the development of the World Wide Web will be covered. The discussion is arranged as follows: a brief summary to the beginnings of the Internet, a timeline of key events, creation of the first top-level domains, early browsers, and explosive growth of the Internet and the rise of the search engines.

2.2.1 The Beginning

The history of the Internet begins during the 1950s and 1960s with the development of computers. It came about as a result of early visionaries who saw a great value sharing scientific and military research information via computers.

With the launch of the Sputnik by the USSR in 1957, the United States established the Advanced Research Projects Agency (ARPA) with the goal of becoming a leader in science and developing new technologies. In 1962, Dr. J.C.R. Licklider was chosen to lead ARPA's research efforts and was a key figure in laying the foundation for ARPANET, which would eventually become the Internet.

It wasn't until December of 1969 that it was brought online, and at the time, there were only four computers connected at the following universities: UCLA, Stanford, UCSB and the University of Utah; you can see the original four-node network in Figure 2.1below.

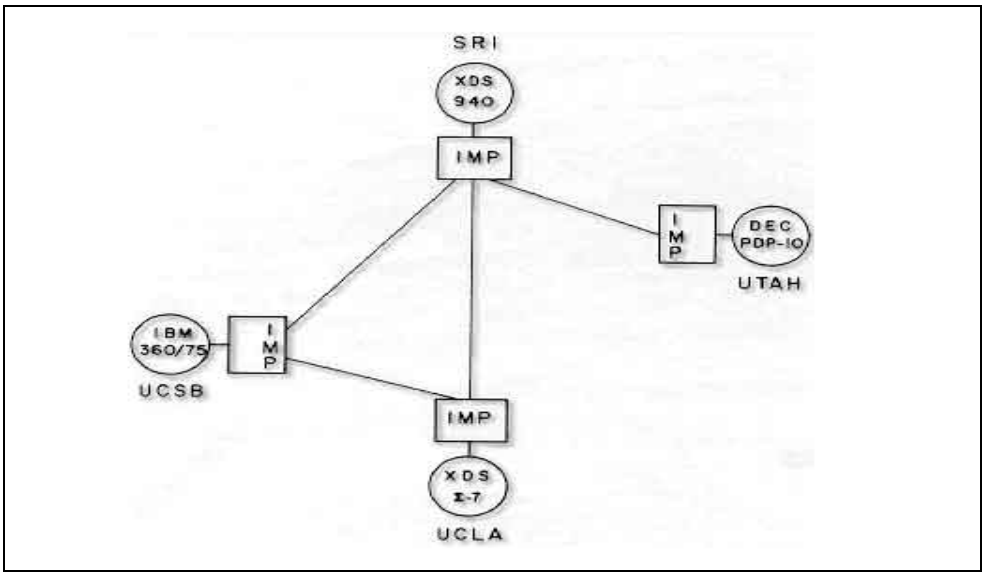


Figure 2.1: ARPANET: Four-node network in 1969 [30].

The Internet was designed to be a communications network among a network of computers and to be fault-tolerant against a possible nuclear attack. The idea was that it would continue sharing information even if some nodes were destroyed; routers would be able to transmit data through the network via different routes.

2.2.1 Internet History Timeline

Below is a list of key events in the history of the Internet taken from [29]:

- 1957 – After the successful launch of the Sputnik by the USSR, the USA saw the need to create Advanced Research Projects Agency (ARPA) "with the mission of becoming the leading force in science and new technologies."
- 1962 – J.C.R. Licklider of MIT proposes the concept of a "Galactic Network," then later elected to head ARPA's research efforts.
- 1962 - Paul Baran, a member of the RAND Corporation, determines a way for the Air Force to control bombers and missiles in case of a nuclear event. His results call for a decentralized network comprised of packet switches.
- 1968 - ARPA contracts out work to BBN. BBN is called upon to build the first switch.
- 1969 – ARPANET created - BBN creates the first switched network by linking four different nodes in California and Utah; one at the University of Utah, one at

the University of California at Santa Barbara, one at Stanford and one at the University of California at Los Angeles.

- 1972 - Ray Tomlinson working for BBN creates the first program devoted to email.
- 1972 - ARPA officially changes its name to DARPA Defense Advanced Research Projects Agency.
- 1972 - Network Control Protocol is introduced to allow computers running on the same network to communicate with each other.
- 1973 - Vinton Cerf working from Stanford and Bob Kahn from DARPA begin work developing TCP/IP to allow computers on different networks to communicate with each other.
- 1974 - Kahn and Cerf refer to the system as the Internet for the first time.
- 1976 - Ethernet is developed by Dr. Robert M. Metcalfe.
- 1976 – SATNET, a satellite program is developed to link the United States and Europe. Satellites are owned by a consortium of nations, therby expanding the reach of the Internet beyond the USA.
- 1979 - USENET, the first news group network is developed by Tom Truscott, Jim Ellis and Steve Bellovin.
- 1981 - The National Science Foundation releases CSNET 56 to allow computers to network without being connected to the government networks.
- 1983 - Internet Activities Board released.
- 1983 - TCP/IP becomes the standard for internet protocol.
- 1983 - Domain Name System introduced to allow domain names to automatically be assigned an IP number.
- 1984 - MCI creates T1 lines to allow for faster transportation of information over the internet.
- 1984- The number of Hosts breaks 1,000
- 1987- The number of hosts breaks 10,000
- 1988 - Traffic rises and plans are to find a new replacement for the T1 lines.
- 1989- The Number of hosts breaks 100 000

- 1989- Arpanet ceases to exist
- 1990 - Advanced Network & Services (ANS) forms to research new ways to make internet speeds even faster. The group develops the T3 line and installs in on a number of networks.
- 1990 - A hypertext system is created and implemented by Tim Berners-Lee while working for CERN.
- 1990- The first search engine is created by Mcgill Univeristy, called the Archie Search Engine
- 1991- U.S greenlight for commerical enterprise to take place on the Internet
- 1991 - The National Science Foundation (NSF) creates the National Research and Education Network (NREN).
- 1991 - CERN releases the World Wide Web publicly on August 6th, 1991
- 1992- Number of hosts breaks 1,000,000
- 1993 - InterNIC released to provide general services, a database and internet directory.
- 1993- The first web browser, Mosaic (created by NCSA), is released. Mosaic later becomes the Netscape browser which was the most popular browser in the mid 1990's.
- 1994 - First internet ordering system created by Pizza Hut.
- 1994 - Comet Shoemaker Levy photos distributed by NASA on the Internet
- 1994 - First internet bank opened: First Virtual.
- 1995 - NSF contracts out their access to four internet providers.
- 1995 - NSF sells domains for a $50 annual fee.
- 1995- Registration of domains is no longer free.
- 1996 - Internet Service Providers begin appearing such as Sprint and MCI.
- 1996 - Nokia releases first cell phone with internet access.
- 1998- Netscape releases source code for Navigator.
- 1998-Internet Corporation for Assigned Names and Numbers (ICANN) created to be able to oversee a number of Internet-related tasks

- 1999 - A wireless technology called 802.11b, more commonly referred to as Wi-Fi, is standardized.

2.2.2 The First Top-level Domains

In 1985, the first top-level domains introduced were `gov`, `mil`, `edu`, `org`, `net`, and `com`; this introduction occurred when the Domain Name System (DNS) was first implemented. Before the DNS was established, the way to access documents in the Internet was by typing the hostname or the IP address; this was somewhat difficult to do and for people to remember. When the DNS system was introduced, it made it easier for people to remember domain names and easier to access documents by translating "human-friendly" names into IP addresses. For example, the domain name *example.com* translates to the IP address 192.0.43.10 [58].

Although the `com` domain was first established in 1985, it was not until 1995 that it became available to the public for commercial use. "With commercialization and popularization of the Internet, the com domain was opened to the public and quickly became the most common top-level domain for websites, email, and networking." As seen from Figure 2.2, the Internet begun to experience its tremendous growth in 1995. This was due to its popularity and commercialization.

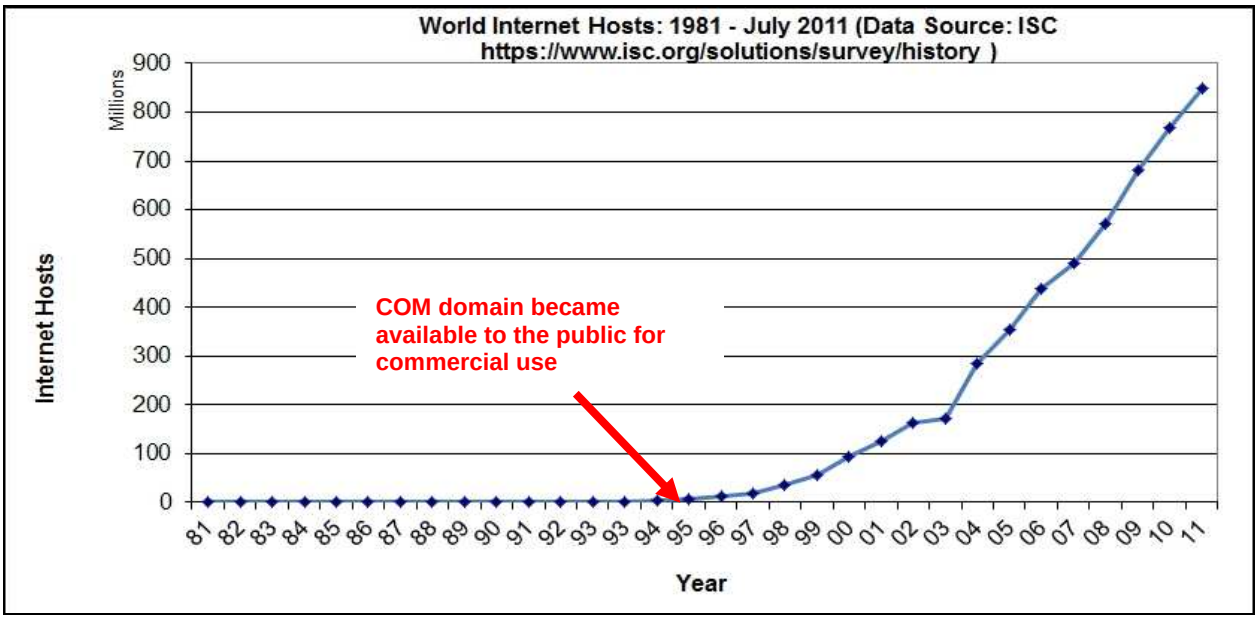


Figure 2.2: Growth in the number of Internet hosts worldwide.

2.2.3 Early Browsers

In 1990, Tim Berners-Lee created the first Web browser (and Web editor) originally called the WorldWideWeb and later renamed to Nexus in order to avoid “confusion between the program and the abstract information space (which is now spelled World Wide Web with spaces)” [31]; it was written in Objective-C using the NeXT computer. And at the time, this was the only way to browse the web. You can see a screenshot of the first browser in Figure 2.3 below.

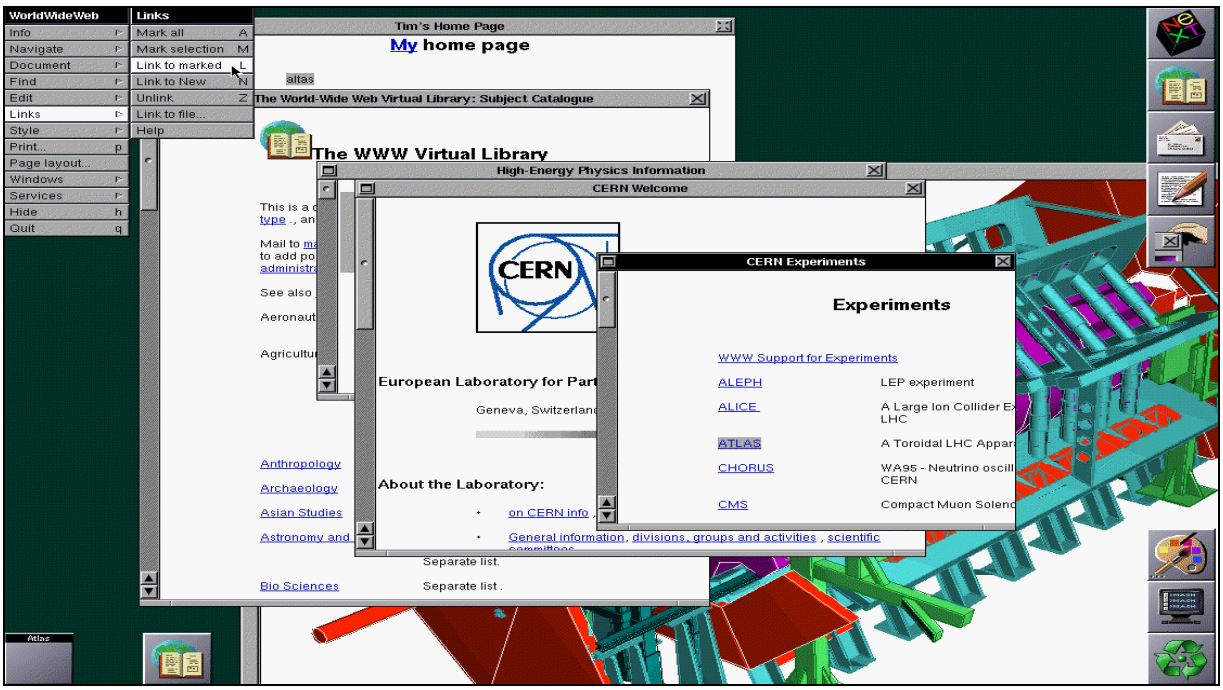


Figure 2.3: Screenshot of the first web browser called WorldWideWeb launched in 1990.

1993 marked an important turning point for the World Wide Web. The National Center for Supercomputing Applications (NCSA) at the University of Illinois, led by Marc Andreessen, introduced the Mosaic browser. It quickly became popular due to its graphical support and its ability to “display images inline with text instead of displaying images in a separate window” [31]. Mosaic made it much easier for people to navigate hyperlinked pages and it made the Web “easy to use and more accessible to the average person. Andreesen's browser sparked the internet boom of the 1990s” [33], as it was previously seen in the Figure 2.2.

A year later, Andreesen “started his own company, named Netscape, and released the Mosaic-influenced Netscape Navigator in 1994, which quickly became the world's most popular browser, accounting for 90% of all web use at its peak” [34]. Then in 1995, Microsoft got involved in the web browser business and released Internet Explorer which was “heavily influenced by Mosaic, initiating the industry's first browser war. Bundled with Windows, Internet Explorer gained dominance in the web browser market” [34].

2.2.4 Explosive Growth of the Internet

As explained before, during the mid-nineties the Web started experiencing a tremendous growth in both, number of users and number of websites. You can see in Figure 2.4 the growth of the Internet in its early years; by 1997 there were about 70 million users on the Internet and by 1998 this number had doubled.

DATE	NUMBER OF USERS	% WORLD POPULATION	INFORMATION SOURCE
December, 1995	16 millions	0.4 %	IDC
December, 1996	36 millions	0.9 %	IDC
December, 1997	70 millions	1.7 %	IDC
December, 1998	147 millions	3.6 %	C.I. Almanac
December, 1999	248 millions	4.1 %	Nua Ltd.
March, 2000	304 millions	5.0 %	Nua Ltd.
July, 2000	359 millions	5.9 %	Nua Ltd.
December, 2000	361 millions	5.8 %	Internet World Stats
March, 2001	458 millions	7.6 %	Nua Ltd.
June, 2001	479 millions	7.9 %	Nua Ltd.
August, 2001	513 millions	8.6 %	Nua Ltd.
April, 2002	558 millions	8.6 %	Internet World Stats
July, 2002	569 millions	9.1 %	Internet World Stats

Figure 2.4: Growth of Internet users worldwide.

In Figure 2.5 below, you can see the growth trend in number of Internet users (United States versus the World) as taken from [36]. Even though it’s nearly impossible to gather exact figures, you can see that its dramatic growth starts in the mid-nineties. This growth is seen in the number of users and number of websites.

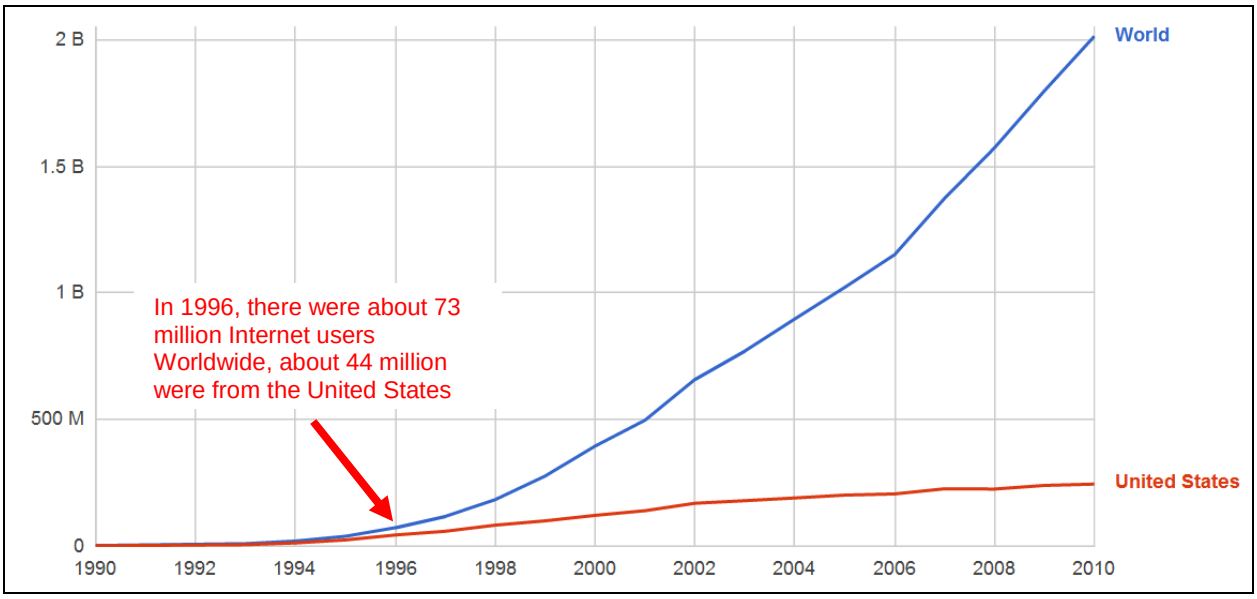


Figure 2.4: Number of Internet users the United States vs the World [36].

2.2.5 Early Rise of Search Engines

With the continued exponential growth of the Internet, it became apparent the need for classification of the content of the Internet. As result, search engines and Web directories started to appear in the early 1990s to organize pages and to make it easy for people to find information online.

In 1994, WebCrawler became the first widely popular "full text crawler based search engine" [37] which allowed users to search for words within the HTML page. This led to what "has become the standard for all major search engines" [37], to let users find information via search queries. Prior to WebCrawler, earlier search engines relied on the page titles and page headings.

Lycos, which was created in 1994 by Michael Loren from Carnegie Mellon University, began as a research project and then in 1995 became commercial. It was a search engine and a web portal that provided email, news, entertainment in addition to web search [38].

Soon after, other search engines started to appear and gain popularity among web surfers to find information scattered all over the web, some of these earlier search engines "included Magellan, Excite, Infoseek, Inktomi, Northern Light, and AltaVista. Yahoo! was among the most popular ways for people to find web pages of interest" [37].

It was during this time that Web directories and Web portals were becoming very popular. For example, Yahoo!'s "search function operated on its web directory, rather than full-

text copies of web pages" [37]. Users had the ability to browse the directory in the Web portal instead of doing a keyword-based search. This became widely popular and web portals became the starting point of the user's web browser experience. Since many web portals provided additional services such as email, news and entertainment in addition to search, people found themselves spending a lot of time in there. Figure 2.6 shows a screenshot of Yahoo! directory circa 1997.

Figure 2.6: Yahoo! directory circa 1997.

As the Internet continued to grow at a fast pace, the popularity of these early web portals and search engines started to decrease; people were looking for other pages of interest that existed away from these portals. And new approaches to search and find information began to develop. It was just not feasible to review full lists of results anymore, and it was the arrival of Google and its proprietary *PageRank* algorithm that would have the greatest impact on web search. Google would dramatically change the concept of search and this will be described next in Section 2.3.

2.2.6 Summary

As you have seen, the early development of computers during the 1950s and 1960s laid the foundation for the US-Government backed ARPANET agency, which would become the Internet. Key events such as the invention of the TCP/IP in 1983 made it easy for hosts to connect and communicate, making it the standard Internet protocol for communication. In the same year, the Domain Name System (DNS) was introduced which automatically assigned IP numbers to “human-friendly” domain names.

Then in 1990, with the introduction of the hypertext system by Tim Berners-Lee, interlinked documents could be easily accessible via the Internet. But it was the introduction of Mosaic web browser three years later that helped fuel tremendous World Wide Web usage; it was one of the first browsers to provide graphical support and made it easy for regular people to navigate the interlinked hypertext documents on the Internet.

During the 1990s, the Internet experienced a huge increase in the number of users and number of websites; ISPs brought the Internet to every home. As a result of this, early search engines and web directory portals started to appear; search engines became popular because they made it easier for people to find information and web portals (e.g. early Yahoo!) became the starting point into the web browsing experience for most users.

By the late 1998, the number of Internet users had reached about 147 million users and the number of websites had increased dramatically. Furthermore, new approaches to web search became available; this paved the way to relevancy ranking of web pages. But it was the launch of Google in 1998 (and the introduction of its proprietary *PageRank* algorithm) that would have the greatest impact on web search.

2.3 Birth of Google

This section will briefly discuss the beginnings of Google, will explain the *PageRank* algorithm and the importance of hyperlinks within the Internet world. Moreover, a recent case study will be provided to discuss the importance of hyperlinks in search rankings and how corporations are taking advantage of SEO to send more users to their sites.

2.3.1 How It All Started

Google was started by Larry Page and Sergey Brin while they were both students at Stanford University. The idea for Google came about as part of a doctoral research project that they began in 1996. Then in 1998 Google was incorporated.

The ranking algorithm behind Google was what differentiate it from other search engines and it was the key to its early success. Based on their experience while conducting their "academic research," they hypothesized.

The following statement, taken from [59] explains it best:

> *"Based on Larry and Sergey's experience with the process of academic research, they believed that Web page authority and relevance could be derived algorithmically by indexing the entire Web, and then analyzing who links to whom. This idea came from the fact that in academia authority is derived when researchers advance their own research by citing one another as part of the research process. Indeed, each piece of published scholarly work (including Larry and Sergey's dissertation) has a works-cited page at the end of each finished piece of written research, which includes a list of resources that were cited as relevant to the work being advanced."*
>
> *"Larry and Sergey took the process of citing in academic research, and hypothesized that Web pages with the most links to them from other highly relevant Web pages must be the most relevant pages associated with a particular search."*
>
> *"To further bolster the concept, Larry and Sergey created PageRank (named after Larry Page), which not only counts how many links point to any given page, but also determines the quality of those links."*
>
> *"Although the Google algorithm is more complex than just analyzing who links to whom, the process of algorithmically analyzing links was a great idea that has separated Google from its competition."*

As a result of this, presently Google is the leading search engine. It controls more than 60% market share here in the United States and is the preferred search engine in most parts of the world.

2.3.2 How Google Ranks Pages

As discussed earlier, search engines use their own proprietary algorithm to rank web pages to be displayed in the search results page. Search engine companies will never reveal the exact mathematical formula or algorithm that powers their search engine because it is one of their most guarded secrets. This is what Google has said about their ranking algorithm:

> "*Traditional search engines rely heavily on how often a word appears on a Web page. Google uses PageRank to examine the entire link structure of the Web and determine which pages are most important. It then conducts hypertext-matching analysis to determine which pages are relevant to the specific search being conducted. By combining overall importance and query-specific relevance, Google is able to put the most relevant and reliable results first.*"

As you can see from the previous quote, Google does not place great emphasis on how many times the keyword being searched appears on the web page, just like other traditional search engines, instead it uses its proprietary ranking algorithm *PageRank* to examine the link structure of the page to determine the relevancy and importance of web pages. The next section will cover *PageRank* in more detail.

2.3.3 Google PageRank Explained

In a research paper published in 1998, Larry Page and Sergey Brin write about how their ranking algorithm takes "advantage of the link structure of the Web to produce a global 'importance' ranking of every web page. This ranking, called *PageRank*, helps search engines and users quickly make sense of the vast heterogeneity of the World Wide Web" [23]. PageRank defines the weight, or level of importance, of a given web page or set of pages. It's similar to a voting system, but within Google search, it's based on a large scale voting system where web sites vote for one another.

PageRank resulted out of the idea of how academic papers were cited. If an academic paper was cited often, it can be concluded that the paper must be important. In the same way, the more links your website has from relevant and authority pages, the higher the *PageRank* it will have, and the higher the probability that it will rank higher as well. You "can simply think of every link as being like an academic citation" [23]; for example, a popular site such as

www.msn.com will contain thousands of links (or citations) pointing to it from other sites. From this, one can conclude that www.msn.com is important because it contains thousands of backlinks, or votes, from other sites.

Here's a brief description of *PageRank*: "a page has high rank if the sum of the ranks of its backlinks is high. This covers both the case when a page has many backlinks and when a page has a few highly ranked backlinks" [23]. Trying to explain the exact definition of *PageRank* and how it works is beyond the scope of this paper, but a simplified *PageRank* explanation of how it's calculated and it's transferred can be seen in Figure 2.7 as taken from [23]. Keep in mind that *PageRank* assigns a weight value between 0 and 10; the large numbers in Figure 2.7 are used for demonstration only.

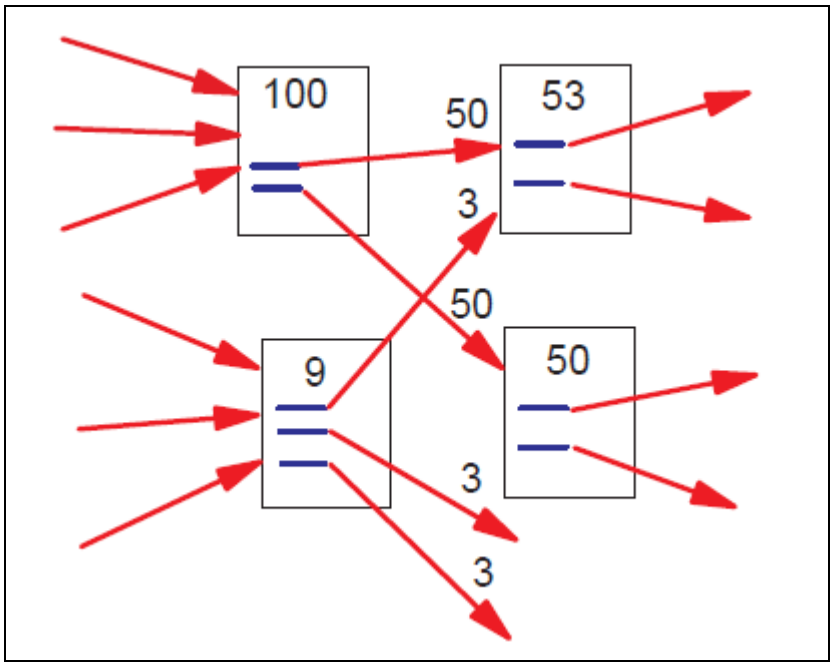


Figure 2.7: How *PageRank* gets transferred.

In this simplified version, note how the *PageRank* of a page is divided among its outgoing links evenly "to contribute to the ranks of the pages they point to." Furthermore, notice the transfer of rank from one page to another, which is determined by the total number of outgoing links from the page. For example, the page with PageRank 9 has three outgoing links, each one of them transferring a *PageRank* of 3.

We can see the importance of links and how a critical role they play in Google's ranking algorithm. Links are important factors in how Google's search engine algorithm determines relevancy and importance from pages that are displayed in the first page. Therefore, links are still one of the most important factors used in its ranking algorithm. As the web evolves, new ranking factors will continue to be introduced, but links and how they are configured will also continue to play a critical role in the Google search engine.

2.3.4 Links: The Currency of the Internet

As discussed, links pointing from one page to another can be considered as a "vote", so a page receiving the most links will always rank higher. In theory, this is the case, but it's not as simple as this because not all links have the same weight or value; some links have more weight (more PageRank) than others. For example, if a website with *PageRank* 6 places a link to your site, this single link has more value than if your page were to receive hundreds of links from sites with *PageRank* 1. In other words, what you're looking for is not just thousands of votes, but votes from high authority websites with a high *PageRank* because getting links from high PageRank sites can be the determining factor of whether your site achieves top rankings in the search results or not.

Getting thousands of links pointing to your web site doesn't automatically guarantee that your site will rank higher and achieve top placements in the search results. As you will see later in Chapter 4, in order to influence search rankings into your favor by getting higher positions in the search results, links that point to your site need to be from relevant and authority web sites; this means that they must be related to what your site is about and should have a high *PageRank*.

Figure 2.8 shows a visual presentation of backlinks; websites A and B are backlinks of C. In the same way, sites A and B have forward links to site C.

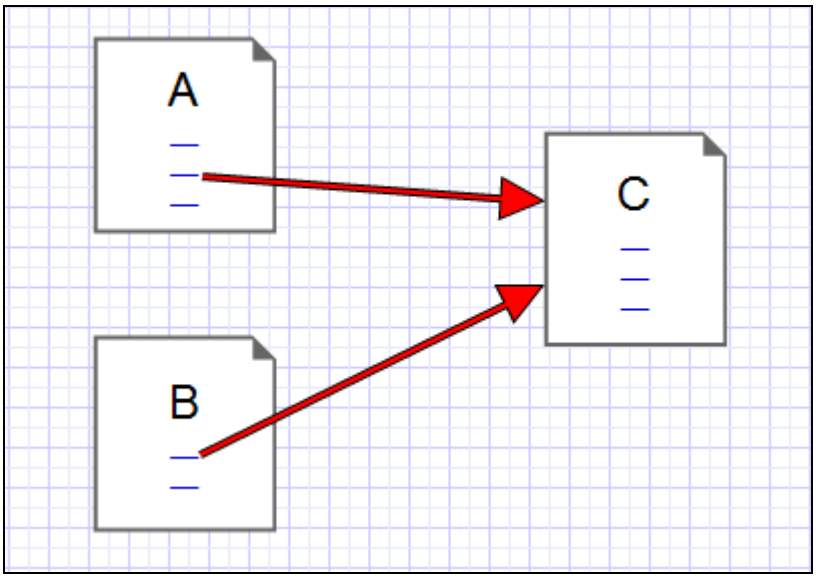


Figure 2.8: Visual representation of backlinks.

As Google performs its hypertext-matching analysis, it also analyzes the content of the pages within the website to determine their relevancy and make sure that all pages are all related. This analysis is performed to detect fraudulent links or any "black-hat" methods, optimization techniques that attempt game the search engine in order to gain top rankings. As will be discussed in the next section, these techniques are not approved and are condoned by all major search engines.

2.3.5 Link Importance: A Case Study

In early 2011, The New York Times published an article involving major retailer J.C. Penny (JCP) which was accused of using black-hat SEO methods in order to maintain top rankings for competitive keywords. The article talks about how JCP outranked millions of other sites for popular searches such as "*dresses, bedding* and *area rugs*. For months, it was consistently at or near the top in searches for '*skinny jeans*,' '*home decor*,' '*comforter sets*,' '*furniture*' and dozens of other words and phrases" [39], it even says that JCP's website appeared on top of manufactures' sites "in searches for the products of those manufacturers" [39]. For example, whenever users typed *samsonite carry on luggage* on Google, JCP would appear at the top of the search results, ahead of samsonite.com.

So how was JCP able to achieve top or near the top listings for competitive keywords? It was because JCP had acquired thousands of backlinks to its site. It was interesting to read that

most of the 2,015 links pointing to the JCP site were from totally unrelated pages. The article states that "the phrase '*black dresses*' and a JCP link were tacked to the bottom of a site called nuclear.engineeringaddict.com. '*Evening dresses*' appeared on a site called casino-focus.com. '*Cocktail dresses*' showed up on bulgariapropertyportal.com. '*Casual dresses*' was on a site called elistofbanks.com. '*Semi-formal dresses*' was pasted, rather incongruously, on usclettermen.org" [39].

Figure 2.9 shows the HTML source code of how these links might have been implemented along with the keywords. The target web site is the destination site (or page) that's receiving the backlink, and the keyword is the anchor text that's visible to the users. Having thousands of backlinks in this format (with a keyword as the anchor text), JCP was able to position its website in the top search listings, besting millions of other sites for popular and competitive keywords.

Figure 2.9: Backlink with keyword as anchor text.

The article further states that even though links to your site may come from unrelated pages, these backlinks "can bolster your profile if your site is barnacled with enough of them. And here's where the strategy that aided JCP comes in. Someone paid to have thousands of links placed on hundreds of sites scattered around the Web, all of which lead directly to JCPenney.com" [39]. JCP denied any involvement in this, saying that it was against their search policies and would work on taking all those links down.

When The NY Times sent Google the evidence it had collected about the JCP "link scheme," Matt Cutts (the head of the Search Quality Group in Google), stated that the links pointing to JCP did violate their search guidelines and that they would "take strong corrective action" [39]. And they did, because according to the article, a few days later JCP went from being at No. 1 for "*samsonite carry on luggage*" to No. 71 in the search results. Also, from being at No. 1 for "*living room furniture*" it had dropped to No. 68.

2.3.6 Summary

What started as a doctoral research project for Larry Page and Sergey Brin while they were both students at Stanford University in 1996, paved the way to the creation of the Google search engine which was founded in 1998. Its introduction and the implementation of the *PageRank* algorithm for displaying the most relevant pages in the search engine results page (SERP) would have the greatest impact on web search.

As was discussed, Google does not place great emphasis on how many times the keyword being searched appears on the web page to determine its ranking on the search results. Instead, it uses its *PageRank* algorithm to examine the link structure of the page to determine its relevancy and importance, which is the determinant factor of where it ranks within the search results.

In addition, the important takeaway from the JCP "link scheme" incident is that links are critical and will continue to play an important factor in how Google determines relevancy and importance of web sites. Search engines will continue to update and refine their algorithms in order to provide the most relevant results to its users, free from any deliberate manipulations to game the search engine and influence top rankings

Furthermore, as the Internet keeps expanding, search engines will continue to grow in importance because they are the main starting point of most users looking for information online. And in order to continue meeting user's search needs, search engines will continue updating their search algorithm for displaying the most relevant and unbiased results. So much that a whole new industry has emerged: search engine optimization (SEO) with the goal of influencing the ranking algorithm to help websites improve their rankings for selected keywords being searched for.

2.4 Search Engine Optimization (SEO)

SEO methods are not meant to deceive or manipulate the search engines in an unethical way; they are implemented to help improve the visibility and relevancy of a website in the organic search results by helping them achieve high rankings. SEO can be thought of as a collection of techniques for the strategic editing of the webpage; this process exposes the most relevant page factors to search engines and helps increase its importance in the search engine results page.

SEO is not a simple process to implement because it requires a lot of experience, background knowledge and patience. Search engines can be very unpredictable with their

ranking algorithms constantly being update and enhanced; so it's the job of the SEO engineer to keep up-to-date and stay current.

As discussed in section 2.1.4, Google has stated that its ranking algorithm takes into account more than 200 factors when determining website rankings. It is therefore important for SEO engineers to know what the most important factors in order to undertake a successful SEO implementation.

Although Google will not fully disclose all the factors that are taken into consideration, it does however provide guidelines for SEO engineers or webmasters to follow for improving the overall rankings of websites. The SEO strategies used in my research and their implementation are covered in Chapter 4.

This section is structured as follows: Section 2.4.1 covers the early beginnings of SEO. Section 2.4.2 discusses the importance of SEO and its goals on website rankings. Section 2.4.3 explains on-page SEO. Section 2.4.4 explains off-page SEO. Section 2.4.5 discusses white-hat SEO. Finally, Section 2.4.6 explains black-hat SEO.

2.4.1 SEO Early Beginnings

As was discussed in Section 2.2.5, the early 1990s marked the debut of the earliest search engines; some of the popular ones were Infoseek, Alta Vista and Yahoo!, which was more of a directory than what most people think of a search engine. "Like Yellow Page influencers, early SEOs took advantage of alphabetical order to get to the top of rankings. This included listed pages with names like 'AAA,' '1ForU,' and similar titles. In addition to this rudimentary tactic, early SEOs took advantage of chronological order by submitting websites at certain times (midnight), thus attaining the first result for the given query" [40]. Early SEOs devised and implemented tricky tactics in order to gain more visibility and appear on top of competing pages.

With the continued growth of the Internet, new search engines appeared that used more complex algorithms for ranking pages. "These algorithms used the metrics of keyword density (the number of times a specific word or phrase is used on a given page divided by the total number of words on the page) and meta tags like 'key-words' to supplement their understanding of the content of websites. SEOs followed pace and started the process of keyword stuffing (artificially adding given keywords to a page) in order to be seen more relevant" [40]. Again, you can see that as search engines evolved in the way they ranked pages and determine relevancy,

SEO engineers evolved as well by finding creative ways to influence search rankings, either by using unethical (black-hat SEO) or ethical (white-hat SEO) tactics.

In an article published by The New York Times that dates back to November 1996, it talks about how web developers went “to great lengths to try to get their Web site to appear at the top of the list that is displayed after a user submits a search query” [41] using a black-hat SEO technique called keyword stuffing. The article describes how web developers simply loaded “a site with certain keywords often hidden behind graphics or in black type on a black background,” so that "a search engine that simply counts the number of times a certain keyword appears in a single site will display such sites higher in the relevancy ranking” [41]. It’s interesting to see how these tactics were taking place in 1996, just as the Internet was experiencing its tremendous growth (as seen in Section 2.2.4), and two years before Google launched.

As you see, gaining high rankings for certain keywords was a much simpler task in the early days than it is today. Implementing tactics such as keyword stuffing on the web page would almost certain guaranteed top page placements of the page. Nowadays the algorithms of the search engines have become more complex and thus making it more difficult for SEO engineers to manipulate the search algorithms as before, but the “cat and mouse game between SEOs and search engines continues” [41] to this day, and it will continue to be the case.

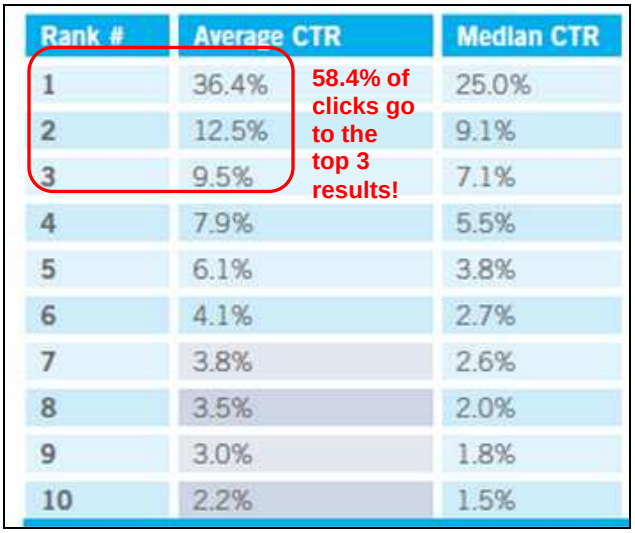
2.4.2 SEO Goals

The goal of SEO is to help websites or web pages achieve top placement in the organic search results by increasing the relevancy of a website or web page to the search query that users type on the search engine. Displaying the most relevant pages for the search query has many benefits to both the user and the search engine providing the results: the user finds the most relevant results for the keyword used, and the search engine is perceived as reliable and trustworthy because its algorithm displays the most relevant pages.

Since SEO is concerned with improving a site’s rankings on the organic or natural search results, the process requires time and knowledge of tactics to implement. Time is a huge initial investment to be made to the website or page being optimized, and an ongoing maintenance to maintain the site’s top rankings.

Depending on how competitive the target keyword is (or group of keywords) that will be used for optimization, SEO may take weeks or even months before seeing any results. Although

the results of the SEO may not be seen immediate, the long-term benefits of SEO can mean top rankings and a high volume of user visiting the site. According to [43], "it's critical for websites to appear on Page 1 of Google, especially in one of the top three organic positions, as these spots receive 58.4 percent of all clicks from users." In Figure 2.10, you can see the results of the recent study; it's interesting to note how being in the No. 1 position in Google is the equivalent to receiving all the traffic that goes from the second all the way through the fifth position.



Rank #	Average CTR	Median CTR
1	36.4%	25.0%
2	12.5%	9.1%
3	9.5%	7.1%
4	7.9%	5.5%
5	6.1%	3.8%
6	4.1%	2.7%
7	3.8%	2.6%
8	3.5%	2.0%
9	3.0%	1.8%
10	2.2%	1.5%

Figure 2.10: % of clicks received by the top 10 search results.

It is no wonder that websites with products or services to sell are all vying to reach to the coveted No. 1 spot. As an example, when a highly competitive keyword such as "*auto insurance*" that gets 1.5 million Google searches every month, means that the top three positions receive 876,000 (or 58.4%) of all the visitors, with 546,000 (or 36.4%) of the visitors going to the No. 1 position according to [43]. This can be extremely lucrative for sites appearing at the top of the results, especially when they have this constant flow of visitors going to their site every month, many of which become clients.

Another study determined that searchers are most likely to click on organic links as opposed to paid listings; it determined that "72.3% of Google users clicked on links generated through searching compared to 27.3%, who clicked on paid listings" [8]. This is another reason why companies that know the effectiveness of SEO spend time and money implementing

strategies so that their site achieves top rankings. Figure 2.11 shows the different areas between organic results and paid listings.

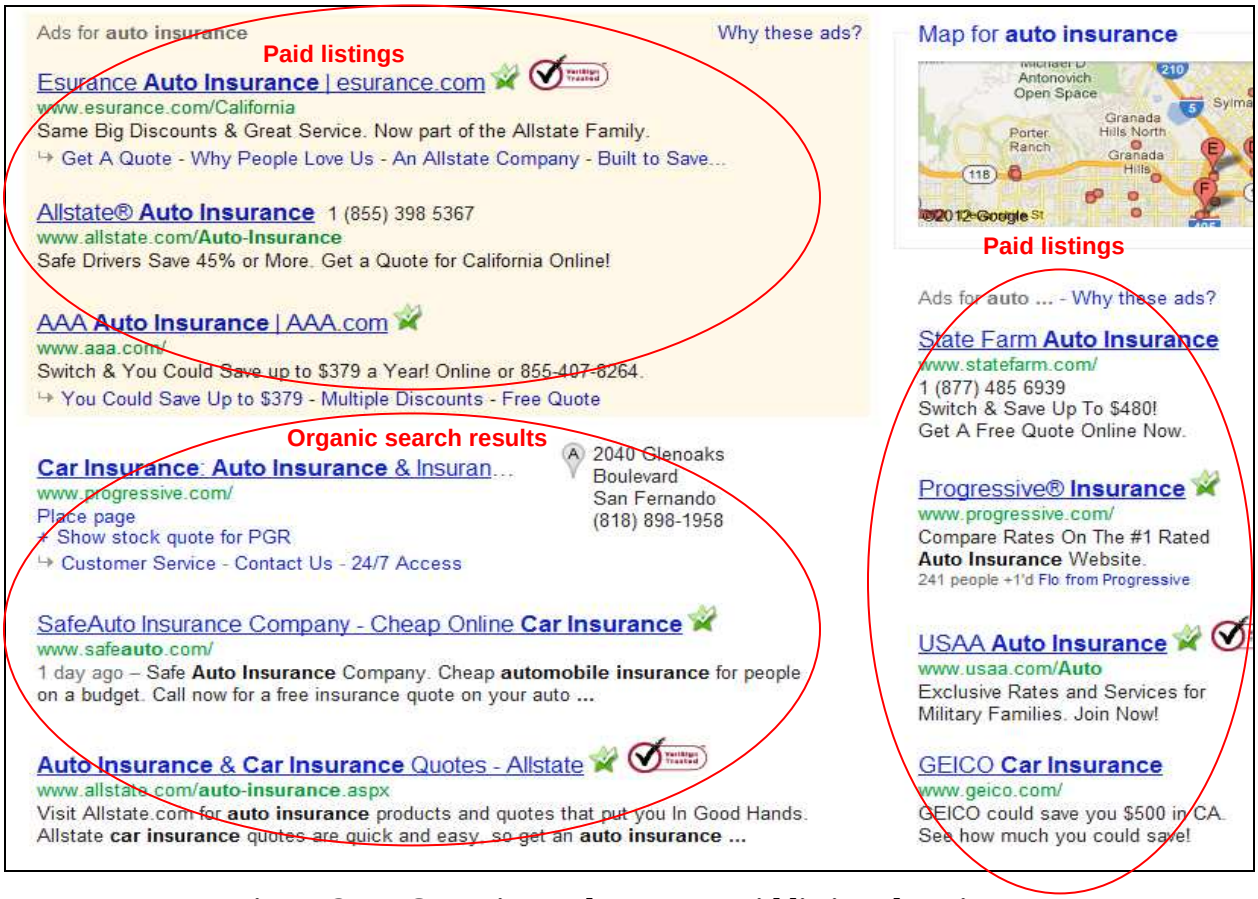


Figure 2.11: Organic results versus paid listings locations.

The previous studies show that if websites want to maintain visibility and a continuous flow of visitors, they need to start paying attention on ways to achieve top rankings through effective SEO execution. But before undertaking any SEO process and start implementing specific search engine optimization methods, it's important to distinguish between on-page SEO and off-page SEO methods. The next sections explain both implementations, along with their main differences and a brief summary of the techniques within each method. Then on Section 4.5 and Section 4.6, I introduce in detail the steps required for effective SEO implementation that were performed on the experimental site.

2.4.3 On-page SEO

On-page SEO deals with anything you have direct control of, in the code or content of your web site (e.g. text, headings, images, links, etc…); basically anything that you implement or upload to your site is considered on-page SEO. On-page SEO lays the foundation of all your SEO efforts because this is where you have most of the control; and as you will see in Chapter 4, any updates implemented on your site can either work for you or against you on the search results. Therefore it's important to get on-page correct before launching into off-page SEO.

As discussed in Section 2.4, Google uses over 200 factors to determine page relevancy and importance for deciding what pages will be displayed in the top search results; you will see that many of these factors are in direct control of the SEO engineer. The following list shows important factors that will be the focus of the on-page SEO strategies implemented in Chapter 4 for this study:

- Keyword research
- Title tag
- Description meta tag
- Robots.txt
- Optimized URLs
- Content
- HTML headings
- Images
- Correct use of the "*rel=nofollow*" attribute
- Keyword placement
- Sitemap

In Section 4.5 I will go into more detail into the importance of the previous list of factors and their on-page SEO implementation. In the next section, off-page SEO will be explained.

2.4.4 Off-page SEO

As opposed to on-page SEO which gets implemented on the website with a high direct control, off-page SEO consists of all promotion that takes place outside of your website; it can be considered as the combination of all the things that can help your site get higher rankings. It's primarily focused on acquiring backlinks (links pointing to your site from other websites) from authority sites in a gradual manner, and any technique that accomplishes this goal is by definition a candidate off-page SEO technique.

And as you will see in Section 4.6, off-page SEO is by definition an open-ended process and is more driven by the SEO engineer's individual creativity on how to achieve backlinks to the site. In contrast, on-page SEO is a more standardized set of technical procedures that can be studied and implemented on the site.

2.4.5 White-hat SEO

In SEO terminology, white-hat SEO refers to the set of techniques, methods or strategies that follows search engine rules and guidelines. These are not mean to deceive or manipulate the search engine or users in any way. It has also been called ethical SEO because it is implemented with a long-term in mind, by following correct guidelines and policies, your site will continue to appear in the search engine results and possibly top rankings.

2.4.6 Black-hat SEO

On the other hand, the term black-hat SEO refers to those set of techniques which are meant to deceive, game or manipulate the search engine's algorithm with the intention to gain high rankings. It's important to understand that these unethical methods are not approved by the search engines, and implementing any of them runs the risk of your site getting removed from the search engine's database index altogether.

Performing black-hat SEO tactics can result in sudden rise in the organic listings, but it may be a matter of time before the search engines determine the existence of shady tactics, which can result in the site can getting penalized and be removed entirely from the search database index. Following are some of the most popular and more common black-hat SEO techniques:

- **Hidden text or links**: text or links that are the same color as the background.
- **Keyword stuffing**: One of the earliest forms of search engine spam, this is when the spammer uses the target keyword in large instances all over the page with the hopes that the search engine finds it relevant.
- **Doorway pages**: These are low quality pages that offer no value to the users. They are created with the sole purpose of ranking for specified keywords; once the visitor arrives at the doorway page, he or she is taken to the homepage or other pages where products or services can be promoted.

- **Cloaking**: This refers to the practice of presenting a page to the search engines and a different page to the users. The purpose of cloaking is to deceive the search engines into displaying a page that it would not otherwise be displayed.
- **Link farms**: As seen in Section 2.4.4, the J.C. Penny incident being charged in running a link farm scheme. The intention of link farms is to create as many sites as possible, all of them linking to your site. The idea is to inflate the number of backlinks going to the spammy site in order to deceive the search engine by into thinking it's an authority site.

Google has stated that its "aim is to give users the most valuable and relevant search results. Therefore, we frown on practices that are designed to manipulate search engines and deceive users" [10]; therefore, it is recommended to stay away from any of the previously mentioned black-hat SEO tactics, or risk the removal from Google's search results.

2.4.7 Summary

You have seen saw how early SEOs implemented rudimentary search engine optimization methods, using naming conventions such as "AAA" to take advantage of alphabetical ordering to appear at the top of the results in the early web directory portals; keyword stuffing was another early tactic used, with the hope of making the page more relevant to the search query.

Implementing such tactics would almost certainly guarantee top page placements in the early Internet days, but search ranking algorithms have become more complex, which has made it more difficult for SEO engineers to influence search engines as before. Although this is the case, there are certain approved methods and guidelines that can be followed which are called white-hat SEO techniques. On the other hand, black-hat SEO methods refer to the set of techniques that are meant to deceive, game or manipulate the search engine's algorithm with the intention to gain high rankings.

Furthermore, SEO falls into two categories: on-page SEO or off-page SEO. As explained earlier, on-page SEO deals with any updates performed on the code or content of the web site (e.g. text, headings, images, links, etc…), and this is where SEO engineers have the most control. On the other hand, off-page SEO consists of all promotion that takes place outside of the website

and is primarily focused on getting backlinks (i.e. links pointing to your site from other websites).

In conclusion, ethical and approved SEO methods are not meant to deceive or manipulate the search engines in an unethical way; they are implemented with the goal of improving the visibility and relevancy of websites in the organic search results. SEO can be considered as a collection of methods for the strategic editing of web pages in order to increase their importance and their relevancy in the search engine results.

As a SEO engineer, it's important to stay current with search trends and changes in search technology in order to launch more effective SEO strategies. And as stated previously, the "cat and mouse game between SEOs and search engines continues" [41] to this day and it will continue in the coming years. This is the motivation for my current research as will be explained in Chapter 3.

Chapter 3 - Problem Statement

This chapter discusses the motivation behind my research and its importance to the current body of knowledge. As presented in Chapter 1, search engines have become part of the daily lives of people; it is through them that most people begin their online activities. People are increasingly dependent upon the Internet as their primary source of Information in many important domains, not just entertainment.

With billions of web pages available, search engines have made it easier for people to find information online. Sorting through the billions of pages and displaying only the most relevant results to the user is a complex task. This has made it increasingly difficult for websites to stay visible. A recent study says that there are about 3 million new websites appearing in the internet every month [4]. As a result, there has been an increase of competition between websites, all of them trying to achieve top rankings in the search engine results page (SERP).

In addition, fewer and fewer people are using traditional media such as newspapers, magazines, radio and TV to buy products, or find services. In the past, the previous media was the “link” that connected people looking to buy products or services to businesses.

Nowadays, most people turn to search engines whenever they’re looking for something. As was discussed in chapter 1, ranking in the first page of Google can make or break a business since more than 80% of users who first visit a website come from search. Of these visits, more than 76% use Google.

So the question becomes, what determines the ranking of a website? Or more importantly, how is it possible to stay visible and be found in the millions of web pages available? How can webmasters setup their sites so that searchers are able to find them? This is where proper and effective implementation of SEO comes into play, having a site optimized so that Google searchers are able to find it.

SEO techniques will only grow in importance, and research in this area is destined to grow dramatically. As online search continues to become more important and the traditional sources of advertising become less relevant, Google will continue to be the essential link in the process by which customers and business locate each other. Businesses that don’t stay relevant and don’t have a serious and up-to-date strategy for using Google in their marketing will not survive. For all others that embrace this constant change, they will be in good position for their continued growth.

3.1 Research Questions and Aim

In this paper, search engine ranking factors and their effectiveness on the Google search engine will be analyzed. I include my experience with the creation and development of an actual website as a platform for the trial application of a set of selected SEO techniques. Furthermore, these SEO experiments will be evaluated by collection and analysis of real data from the experimental website through the use of Google analytics. By following a sound keyword strategy (selecting the right keywords) and methodology, and applying selected SEO techniques, the effectiveness of SEO on Google search engine rankings will be analyzed.

3.2 Objectives

This section will discuss the objectives of my research and how each of them will in turn lead to achieve the aim of this paper.

- **Website topic research** – A step-by-step process of the selection of the website topic will be covered.
- **Keyword research** – How to identify and select possible keywords for conducting an effective SEO on the website will be explained.
- **Website creation and SEO implementation** - This research covers these phases: inception phase of website, development, launch, and SEO implementation. The application of SEO techniques will be outlined and explained in Section 4.5 and Section 4.6.
- **Questions analyzed** – These are the questions this paper attempts to answer:
    1. How effective is the implementation of SEO on a newly created website on Google's search ranking algorithm for achieving top placement in the search results page?
    2. What are the most important factors that have the greatest influence on Google's search ranking results?
    3. How quickly do SEO techniques take effect for newly created websites? How much time is required before the application of SEO techniques start producing measurable results?
    - **SEO Techniques** – In order to answer the previous questions, selected SEO methods will be implemented on a page of *www.casadelalpiste.com* (built

from the ground up) to test their effectiveness. In Section 4.5 and Section 4.6, on-page SEO and off-page SEO will be explained in more detail and the tactics used.

- **Metrics analyzed** – This paper will focus on the following research variables in order to validate and test the effectiveness of the selected SEO techniques; the metrics' definition is also given:
    - **Number of Visits**: The number of times your visitors has been to your site (unique sessions initiated by all your visitors). If a user is inactive on your site for 30 minutes or more, any future activity will be attributed to a new session. Users that leave your site and return within 30 minutes will be counted as part of the original session [7].
    - **Pageviews**: This field indicates the total number of pageviews for your site when applied over the selected time frame. For example, if you select this metric together with Request URI, it will return the number of page views over the time interval for the Request URI for your report [7].
    - **Ranking**: This refers to the position on the search engine ranking results. The goal will be to reach the No. 1 position (on the first page) in these search engine results page (SERP).

3.3 Data Collection and Analysis

The data collected for this research took place from March 2011 until May 2012. The data-gathering tool used for collecting, tracking and analyzing is *Google Analytics*, which is offered by Google. It produces in-depth statistics from website visitors; it's also the most widely used statistics tool, and it's currently in use on around 60% of the 10,000 most popular websites [6].

Chapter 4 - Description of Approach and Methods

In this chapter, the approach and SEO methods will be presented in detail. Section 4.1 covers the selection of the website topic; Section 4.2 covers domain name selection and site setup; Section 4.3 goes into the process of keyword research; Section 4.4 covers the configuration for the data collection; Section 4.5 covers the implementation of the on-page SEO strategies and the chapter ends with Section 4.6 which goes over the off-page SEO methods.

4.1 Topic Website Research

Early March of 2011, before the initial website development started for the experimental website where the SEO techniques would be implemented, I first needed to decide on a theme or topic for the site. After doing some brainstorming, I decided to choose *alpiste* as the main theme for the site. It is a Spanish word that means canary seed in English, it comes from the canary grass plant, "belonging to the family Poaceae. Originally a native of the Mediterranean region, it is now grown commercially in several parts of the world for birdseed, hence the name" [44].

Before continuing with the research, I needed to do market research to validate that there was some interest online, meaning that people were actually searching for *alpiste* in the Internet. The last thing I wanted to do was develop a new site about *alpiste* and not have people searching for it; in other words, the idea of this experiment was to create a website around a topic people were interested in.

I used *Google Insights* to begin the market research; this is a tool provided by Google to research and analyze worldwide (or local) trends of keywords that people are searching. For *alpiste*, I decided to perform a worldwide trend analysis for the date range: 2005 thru 2011. You can see this in Figure 4.1.

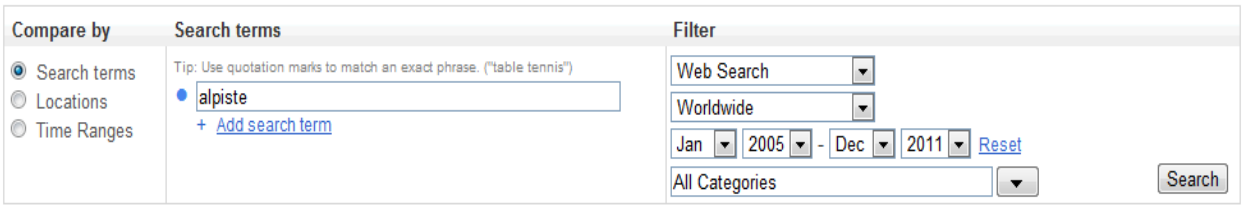


Figure 4.1: Trend analysis search for *alpiste* keyword.

The returned trend analysis graph can be seen in Figure 4.2. One can see from the graph that the search volume for *alpiste* was low from 2005 thru the beginning of 2009, then at the

beginning of the second quarter of 2009 that's when it experienced its largest increase in volume; that's when the interest for *alpiste* started to increase.

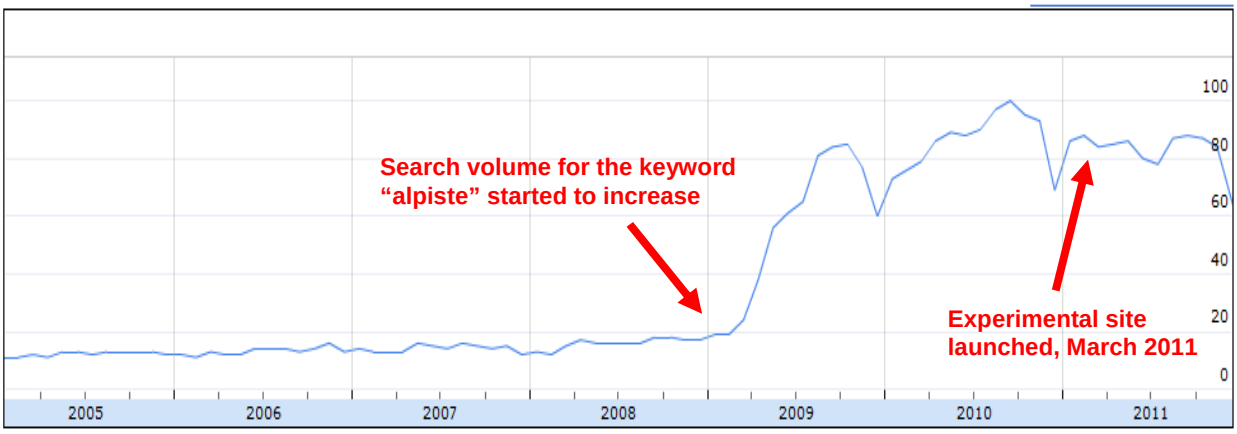


Figure 4.2: Search volume trend for the keyword: *alpiste*

Also, on the right-hand side of the trend graph you can see numbers in a scale from 0-100. According to Google, "these numbers reflect how many searches have been done for a particular term, relative to the total number of searches done on Google over time (over the selected date range). These numbers don't represent the absolute search volume numbers; instead the data is normalized and presented on a scale from 0-100. Each point on the graph is divided by the highest point, or 100" [10]. When not enough search data is available, the number 0 is shown.

As stated, it's important to note that the data is normalized, meaning that Google divides "sets of data by a common variable to cancel out the variable's effect on the data. Doing so allows the underlying characteristics of the data sets to be compared" [11]. Normalizing the data prevents regions with the highest search volumes from always ranking on top.

For example, let's say that Japan and Spain both show 90 for the keyword *vacation*. Even though both countries have the same normalized value for the same keyword, this doesn't mean the "absolute search volume" is the same. Both regions can have significantly different search volume for the same keyword, say 1000 searches per month for Japan versus 20,000 searches per month for Spain, but the search data "can be compared equally because the data has been normalized by the total traffic from each respective region" [11]. Since both countries show a

high normalized value of 90, one can infer that searchers in both Japan and Spain show a high interest and are likely to search for the keyword *vacation*.

A key feature of *Google Insights* is that it provides you with the option of researching the "Regional interest" of the keyword you're researching. In other words, you can see a list of countries (or you can drill down to the city-level) for which your keyword is most popular. This is great to know because now you can begin planning the content of your site to match the interest of the large majority of potential visitors to the site. As you can see from Figure 4.3, the top 10 countries that show the most interest in *alpiste* are located in Latin America; this is not surprising due to the keyword being written in Spanish and its popularity.

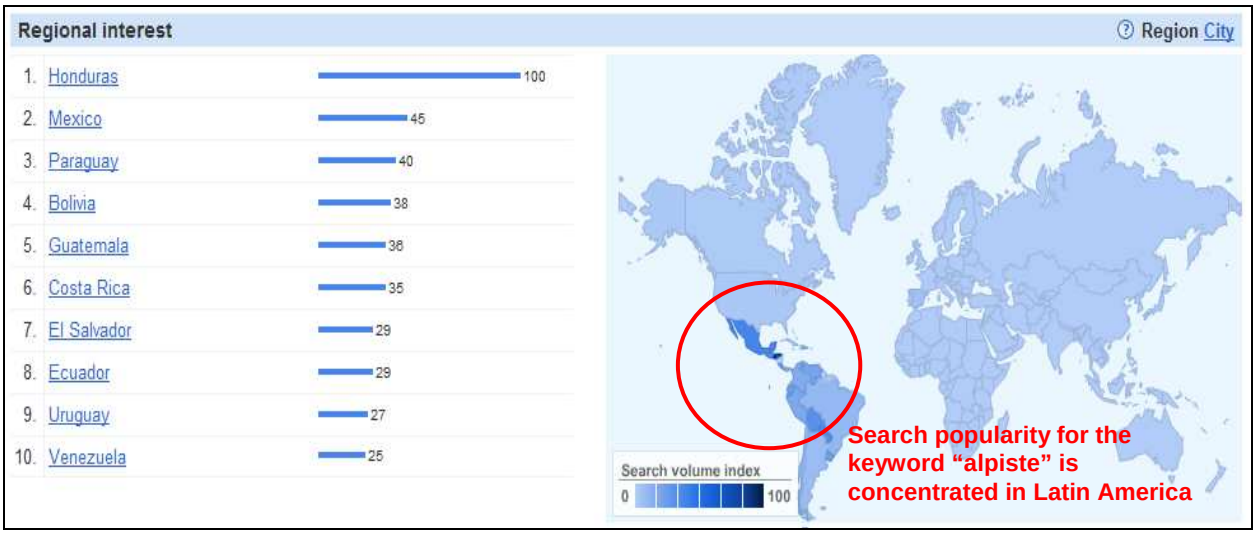


Figure 4.3: Regional interest for the keyword: *alpiste*

Lastly, *Google Insights* also provides you with a table of related search terms to the one you're researching; you can see this in Figure 4.4 below. It provides you with **Top searches,** which are popular keywords that are related to your search query, and to the right you can see the normalized search volume index. **Rising searches** are search queries "that have experienced significant growth in a given time period, with respect to the preceding time period" [12]. Since I selected the date range 2005 – 2011, the research tool is using 2004 as the benchmark for comparison. Also, any keyword that has a **Breakout** label instead of a percentage means that this particular search query has "experienced a change in growth greater than 5000%" [12]; you can also see this in Figure 4.4 below.

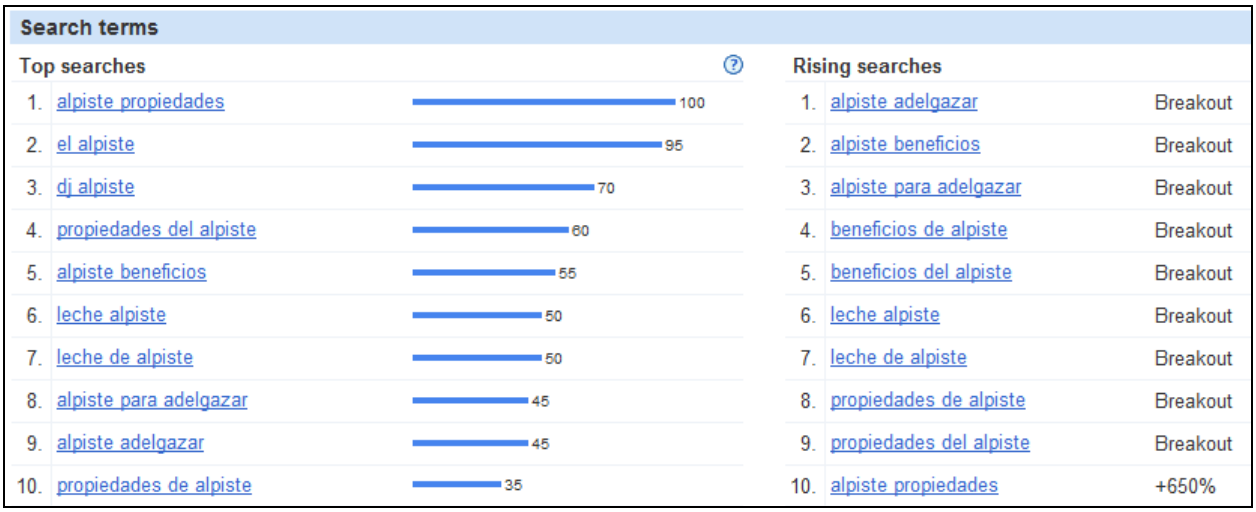


Figure 4.4: Top searches and rising searches related to the keyword: *alpiste*

The steps above culminated the topic research and validation process for the experimental site. There seemed to be good potential for developing the website based on *alpiste* for two important reasons; according to my initial keyword analysis, (1) people were searching for *alpiste* online (enough interest seemed present), (2) there was worldwide interest for the *alpiste*. Even though Latin America seemed to be the region most interested for this particular keyword, I was curious to see what additional visitors I would also generate from other countries through the SEO implementation. This was the start of the keyword research process, and what needed to be completed next was to obtain a domain name and start the website development.

4.2 Domain Name and Website Setup

This section describes selection of the domain name, and the technical details of setting up the website and establishing a web hosting service. The consequences of the choices made at this stage may influence the effectiveness of the SEO techniques being studied and those effects are also described.

4.2.1 Domain name selection

From the start, I was looking to purchase a .COM domain that had the keyword *alpiste* in it. I tried to purchase alpiste.com, but it was already taken. I continued my search for the right domain name and after brainstorming, I decided on *casadelalpiste.com* which translates to “house of alpiste” in English. I liked the name, it was available for purchase and I also felt it

would resonate well with the users searching for *alpiste*. Once I had secured the domain name for the experimental website, I needed to look for a web hosting provider to configure the domain and begin the site development.

4.2.2 Web hosting service

After looking at my options, I decided to host the experimental site using a shared web hosting service. A shared hosting "refers to a web hosting service where many websites reside on one web server connected to the Internet. Each site 'sits' on its own partition, or section/place on the server, to keep it separate from other sites" [45]. This seemed to be the best option for me due to its low cost and there was no need for me to do any major server configurations.

Shared hosting may be implemented in two ways, named-based or IP-based. The differences are explained below:

- In a *IP-based virtual hosting* (also known as dedicated IP hosting) setup, each site on the web server uses a different IP address
- In a *name-based virtual hosting* (also known as shared IP hosting) setup, the web server uses a single IP address to serve multiple domain names, or hostnames.

Some advantages of shared hosting is that it's relatively low-cost and the web hosting service provider is responsible for most of the system administrator tasks, such as managing users, technical support, installing server software and security updates. A disadvantage is that if file permissions are not configured correctly, other users (sharing the same server) may be able to access the system-wide files from different users, compromising the files.

One important thing to keep in mind (when choosing the share web hosting option), is making sure that the shared IP address and the server where the site will reside has not been blacklisted or banned from Google. When developing a new site, think of it like moving to a new neighborhood; you would want to live in a clean and crime-free area. This is the same with shared hosting; you want to avoid having blacklisted sites (sites that have been identified as spammy) or banned sites on the same shared IP address.

Even though Google bans sites, not IP addresses, you want to verify that your IP address is in good standing by not being blacklisted. Google put together a short video explaining

whether a spammy website might affect the rankings of other sites hosted on the same server (on the same shared IP address). It suggested that this is one of the issues that's not on their top list of priorities, but they "understand that shared hosting happens; you can control or help who else is on that (same) IP address or class-C subnet" [17].

However, there have been special cases when there were about 26,000 spammy and porn websites all under one IP address, and there was one normal website. In this special case, this can really affect the SEO of the one normal site because being surrounded by thousands of spammy sites invites scrunity. It exerts more effort to conclude such a site is legitimate. So in order to avoid future problems, it's always recommended to verify that the web server and IP address are not blacklisted.

Once I had purchased the shared web hosting service, I decided to analyze the type of sites that were hosted on the same IP address as the experimental site. To check on the sites, I used a tool provided by Majestic SEO that will display "the most backlinked domains hosted on any given IP or subnet" [18]; in other words, this tool will return the most popular sites (as determined by the total number of links pointing to them) that are hosted in the same IP address. By looking at the types of sites from the returned results, you may be able to identify which sites are spammy.

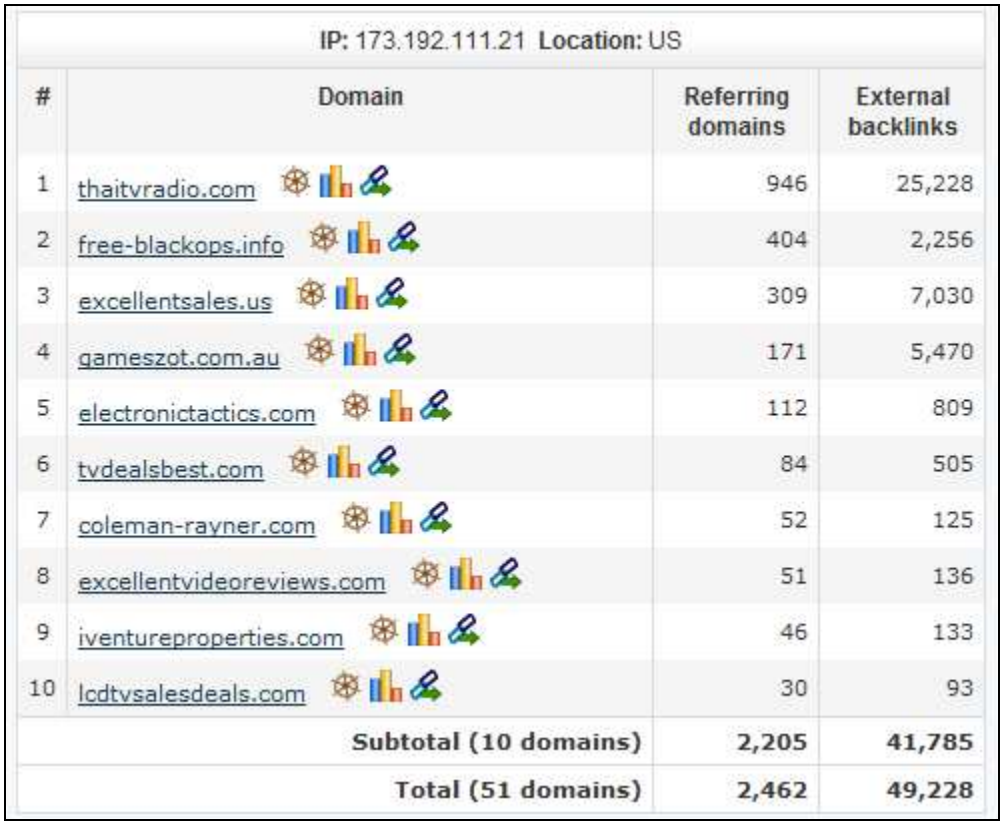
To start the analysis on the returned list of sites, I entered the domain name of the experimental site as seen in Figure 4.5 You also have two options for your analysis:

1. **Fresh Index**: is updated daily and it contains data from the most recent 30 days
2. **Historic Index**: contains all data available in their database

As you can see, I entered www.casadelalpiste.com and selected the *Historic Index* option because I wanted to get the historical data.

Figure 4.5: Searching for sites hosted on same IP as *casadelalpiste.com*

You can view the results of the query in Figure 4.6; it shows the top 10 most popular sites on the shared IP address (173.192.111.21), the same IP where the experimental site is hosted. I then visited each of the returned sites to determine whether they seemed spammy or not. It was interesting to see that out of the 10 sites, only 4 were live sites, the other 6 had been expired and were no longer available.



IP: 173.192.111.21 Location: US

#	Domain	Referring domains	External backlinks
1	thaitvradio.com	946	25,228
2	free-blackops.info	404	2,256
3	excellentsales.us	309	7,030
4	gameszot.com.au	171	5,470
5	electronictactics.com	112	809
6	tvdealsbest.com	84	505
7	coleman-rayner.com	52	125
8	excellentvideoreviews.com	51	136
9	iventureproperties.com	46	133
10	lcdtvsalesdeals.com	30	93
	Subtotal (10 domains)	**2,205**	**41,785**
	Total (51 domains)	**2,462**	**49,228**

Figure 4.6: Top 10 domains (by # of backlinks) hosted on same IP address as *casadelalpiste.com*

The important take away from this section is that choosing a web hosting provider is a crucial step in the development of a website, especially if you're about to implement a series of SEO strategies to improve the search rankings. Moreover, the ideal setup would be to host the site on a dedicated IP address in order to have total control of the site; this way you know exactly what domains are hosted on the server. The only drawback with this option is that it's more expensive; that's the reason I chose the more cost-effective option (shared web hosting), without

compromising the research goals by making sure not a lot of spammy sites shared the same IP address as the experimental site.

4.2.3 Website setup

Once the domain and web hosting were setup, the next step was to start the website development. At this point of the research, I had two choices: (1) create the site from the ground up using traditional HTML coding standards, or (2) use a web-based Content Management System (CMS). In order to speed up the implementation process and make site-wide updates easier, I decided on the second option and chose Wordpress (a web-based CMS) as the back-end for the experimental site.

Wordpress is an open-source and widely used web-based Content Management System (CMS) based on PHP and MySQL. "It has many features including a plug-in architecture and a template system" [46]; that is, it makes it simple to extend its functionality by installing plug-ins without having to code any HTML. Also, by utilizing a template system, it separates the design from the code, which makes it easier to make changes to the front-end (GUI) of the website.

As of this writing, the web server needs to meet the following requirements in order to run Wordpress:

- PHP version 5.2.4 or greater
- MySQL version 5.0 or greater

Once I made sure my host met the minimum requirements, I installed and setup Wordpress on the shared web server; the installation took no more than fifteen minutes to complete.

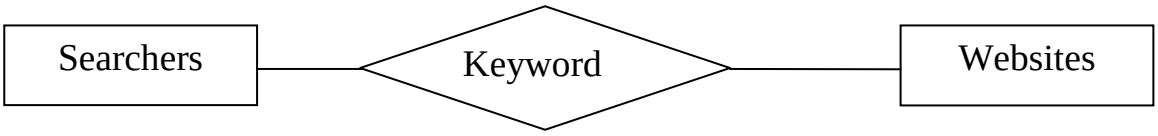
The next step was to start developing content for the site by adding web pages, but before starting this, I needed to complete the most important step in the SEO process: *keyword research*. The keyword research process would help me identify which keyword (or set of keywords) to focus on for this study, and that's the topic of the next section.

4.3 Keyword Research

Effective keyword research is at the heart of SEO. At the start of website development, the developer must carefully choose those keywords that will be critical to the success of the website. In other words, the SEO engineer needs to determine (through the keyword research process) which keywords are the most relevant to the website being developed. The goal here is to find relevant keywords that people are searching for. By relevant, this means choosing

keywords that are closely related to the theme of the website (i.e. to the services or products offered).

The selected keywords play a critical role because the developer assumes that most users will find the website by entering the critical keywords into a search engine, and the search engine will thus refer them to the site based on the strength of the keyword's affinity to the site. Keywords can be considered to be the link that will get the searchers to websites; they are the link that connects users looking for information online to specific sites. The diagram below shows a graphical view of this idea and the importance of keywords.



Not surprisingly, competition among similar websites to be associated and be found with popular keywords is fierce. The developer must now begin to make hard economic choices: how to use the available budget to target specific keywords for driving visitors to the site. Google will sell the developer access to its search results page by allowing them to bid on keywords for a price through Adwords (its Pay-Per-Click (PPC) advertising system), but PPC for highly popular keywords is expensive. The developer can also begin to promote his or her website through SEO and hope to rise in the organic search results, but this process takes time.

Through keywords, search engines drive visitors to websites in two different ways: SEO or PPC. As explained previously, SEO is more concerned with ranking high on the organic search engine results page (SERP) for selected keywords, but the process to achieve top rankings takes time and effort. With PPC, developers can start getting visitors to their sites immediately, but going this route you have to bid a determined amount on the selected keywords and as a result spend money each time a user clicks on one of your paid listings.

Implementing SEO effectively on a constant basis can guarantee you top search engine rankings in the organic search results for your selected keywords, without having to pay Google. But as explained before, SEO will take time and lots of effort to reach those goals. With PPC,

you can start getting visitors to your site right away but there is a high monetary cost involved, and that's where the developer needs to decide what strategy is best to implement.

A good strategy to implement can be seen in Figure 4.7. Along the x-axis we can see the *Time* variable, and along the y-axis we can see the number of *Visitors* variable. Both, PPC and SEO, may be implemented simultaneously (in parallel) so that you can start getting visitors immediately through PPC and at the same time be implementing a SEO techniques for the long-term.

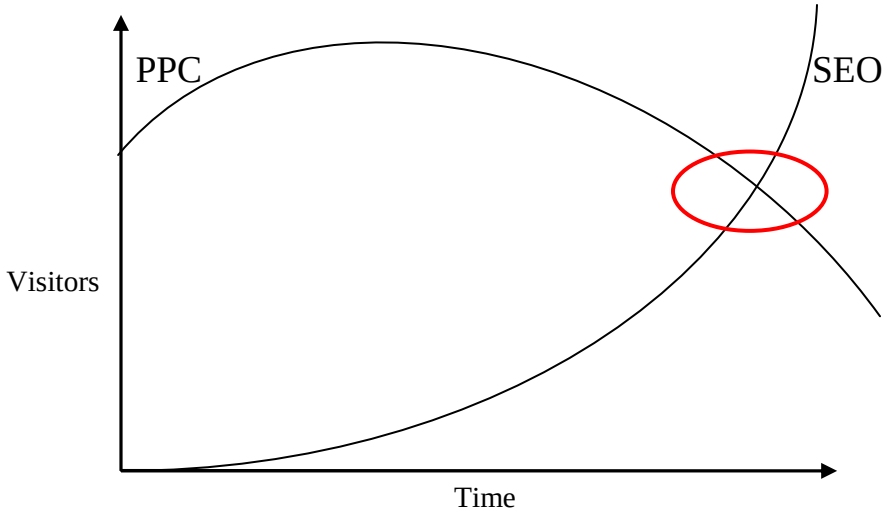


Figure 4.7: PPC traffic vs SEO traffic over time.

In the beginning, the site will not be ranking high on the search engine results so the number of visitors will be low, but with the right SEO implementation and with time, you can expect to see a constant increase of visitors to the point where the total number of visits to the site is the same or greater as PPC. Once this point has been reached, the site owner can then scale back on PPC and focus more time and energy on the SEO, further guaranteeing free traffic from high rankings on the search results. This idea can be seen in Figure 4.7 above.

Hence, effective keyword generation can be considered the most critical component when implementing SEO. Also, all keywords are not created equal; depending on which ones you have identified as possible candidates, some will be much easier to rank for and other keywords will be nearly impossible to achieve top positions on the organic search results.

As a general rule, when one is just starting on SEO, it's best to stay away from broad or general keywords because these are generally more competitive. Since general keywords have a larger search volume, lots of SEO engineers are trying to optimize for the same keyword.

As you will see next in Section 4.3.1, it's a more effective strategy to focus on more specific and less competitive keywords which are called *long-tail keywords*; these will be explained in more detail in the following Section.

4.3.1 Long-tail Keywords

Long-tail keywords are those search queries that contain three or more words; these are more descriptive and are generally much easier to rank for. Therefore, long-tail keywords should represent the bulk of any keyword generation and SEO efforts. Figure 4.8 shows an example of how *long-tail keywords* relate to more popular, high volume keywords. As you can see, the more popular keyword *vacation* has a USA monthly search volume of 13,600,000 compared to 1,300 for the long-tail keyword *best all inclusive vacation packages*.

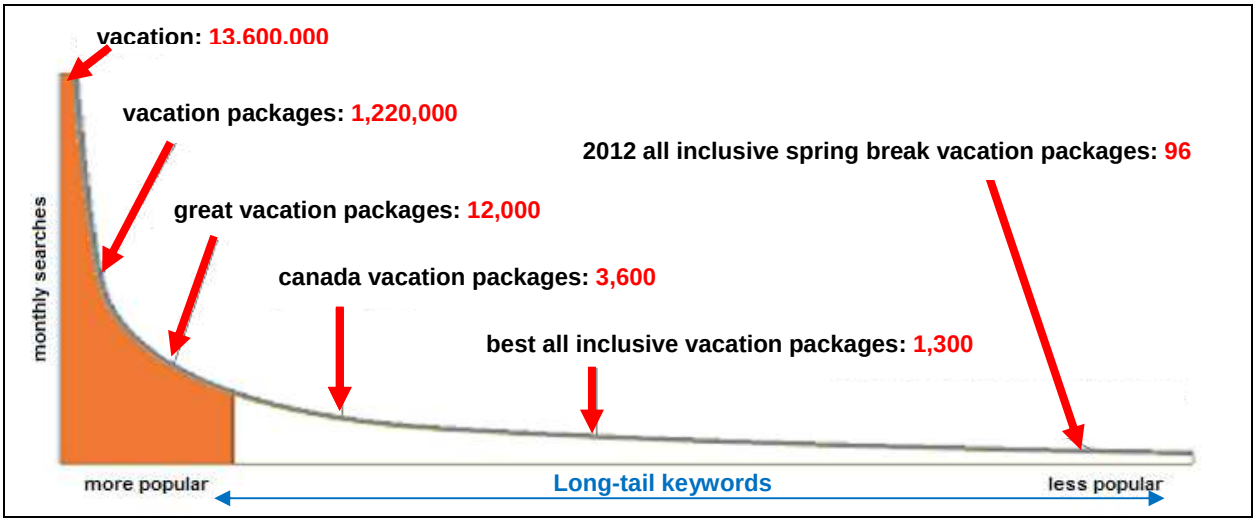


Figure 4.8: An example of *long-tail keywords.*

Even though general keywords tend to have larger search volumes, more specific keywords (long-tail keywords) as a group, can generate a significant search volume, sometimes more than general search queries. And when applying SEO on your site, it's easier to rank for these more specific keywords than it is for the more popular ones. Again, unless you have the

resources to do so (time and money), it's better to focus on more specific keywords to optimize your site.

Another key aspect of these long-tail keywords is that they also tend to *convert* better than broad keywords. The term *convert* is a marketing phrase that has several meanings (in different settings) depending what has been decided by the website owner. For example, a conversion can be described when users visit an e-commerce website and make a purchase; another example is when users visit a site and signup to receive a newsletter, this can also be considered a conversion.

In a search example, a user searching for the general keyword *vacation* may or may not be looking to purchase a vacation package or vacation deals. But a user that types "*2012 all inclusive spring break vacation packages*" into the search engine, we can infer from the search query that there's a higher probability that the user knows what he or she is looking for and is ready to buy, provided that the search engine displays relevant page results that match the query.

Thus, knowing this about keywords and how important it is before embarking on any SEO strategy, the next step is to start the keyword research process so we can identify the keywords that will be used for optimization of the experimental site.

4.3.2 Keyword Research Process

This section explains the keyword research process; as stated, it can be considered the most important part of the SEO process because it is here that keywords are discovered and identified to be used in your search engine optimization efforts. Selecting the wrong keywords, those that are not relevant to the theme of your site or worse, those that people are not searching for, can negatively affect your SEO strategy. You may implement SEO techniques perfectly on your website, but if you have selected the wrong keywords you will be sending the wrong visitors to your site, or none at all. So it's crucial to execute an effective keyword research from the start.

Keyword generation tools such as Google's Keyword Suggestion Tool (KST), Keyword Discovery, and Wordtracker can help you discover relevant keywords for your website; these tools will also help you research, analyze and select potential keyphrases. And even though these keyword generation tools can provide you with thousands of keywords, each page of the site should be optimized for only one or two closely related keywords. The goal is not to generate thousands of keywords, but rather a shorter and more focused list of keywords that are relevant

to the theme of the site; keeping these points in mind will help you maintain a your focus and will give you a better chance at ranking at the top of the search results for the selected keywords.

As discussed, even though there many keyword generation tools available, I used Google's *Keyword Suggestion Tool* for my keyword research process. It is the de facto tool for keyword research and analysis; it is also the most widely used.

Once I'm logged in to KST, I'm presented with three options to begin the keyword research. I decided to do a word or phrase search and type *alpiste* into the specified box; you can see this from Figure 4.9.

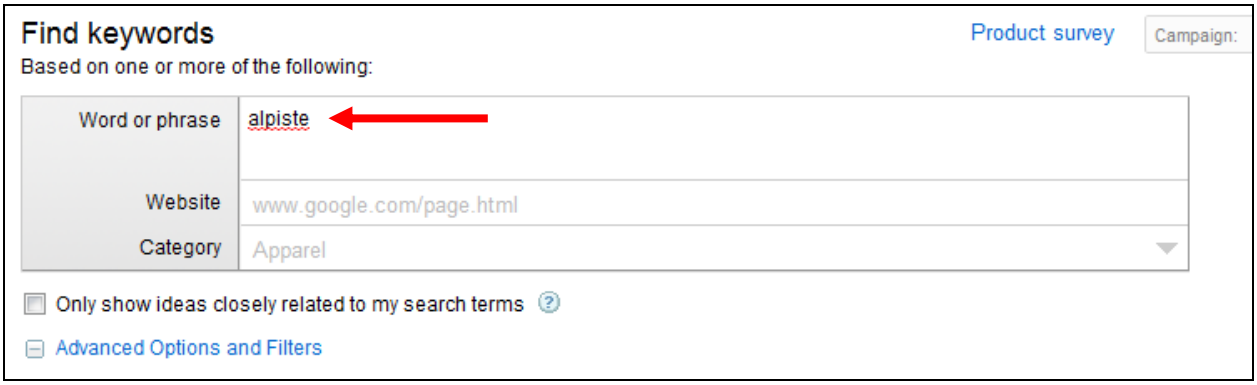


Figure 4.9: Start of keyword research in Google's Keyword Suggestion Tool (KST).

I then proceeded and selected *Advance Options and Filters* as seen in Figure 4.10. This provided me with more research options. Because I was focusing on the Spanish word *alpiste*, I selected *All Countries* and *Spanish* for the *Locations and languages* criteria, respectively. Also notice the three available options for *Match Types* on the left-hand side: **Broad**, **Exact** and **Phrase**. This is Google's definition for *Match Types*: "the traffic volume that appears in the statistics table for a specific match type is an approximation of the traffic that a keyword gets in a month on Google with that match type" [60].

- **Broad match**: This option includes the sum of the search volumes for the specified keyword, related grammatical forms, synonyms and related words.
- **Phrase match**: This option includes the sum of the search volumes for all terms that include the keyword.
- **Exact match**: This option includes only the search volume for the exact keyword as it is written.

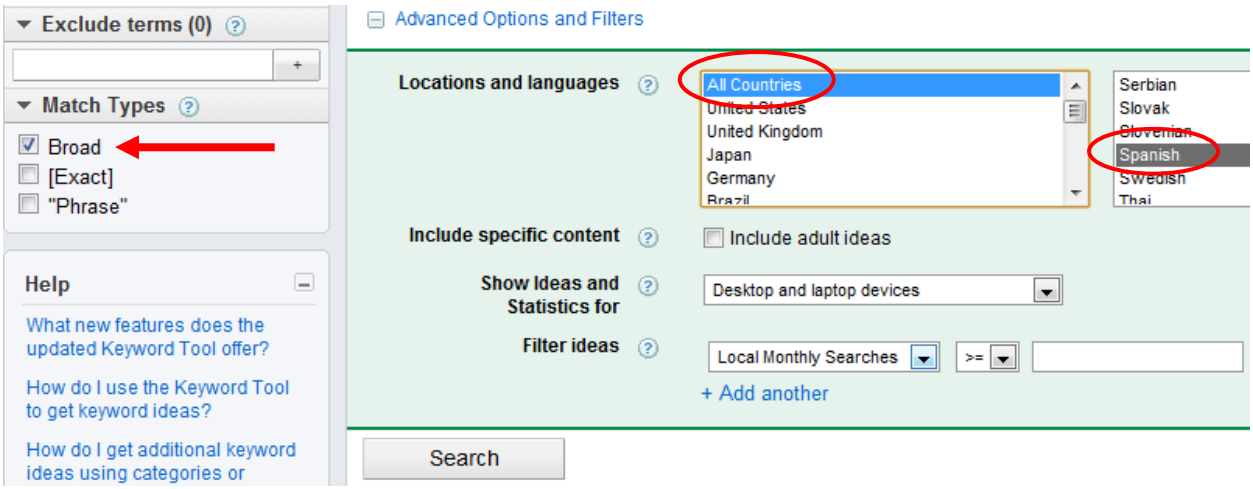


Figure 4.10: Keyword research Advance Options in KST.

At this early stage of the keyword research process, the focus is not to look for any specific keywords; you're brainstorming for keyword ideas and analyzing the results from Google's Keyword Suggestion Tool (KST). Thus, I selected **Broad** from the *Match Types* option to get all keyword variations such as synonyms, related terms and even misspellings of the *alpiste* keyword. Once I submitted my query, KST returned me a list of keyword variations related to *alpiste*; I then sorted the results by *Local Monthly Searches* in descending order to get the keywords with the highest volume searches at the top.

As you can see from Figure 4.11, at the very top of the results, KST provides you the search volume stats of the keyword you're researching. In this example, you can see the monthly search volume stats for the initial keyword I used: *alpiste*. Notice that because I selected *All Countries* for the **Location** criteria, the search volume stats for **Local Monthly Searches** and **Global Monthly Searches** is the same.

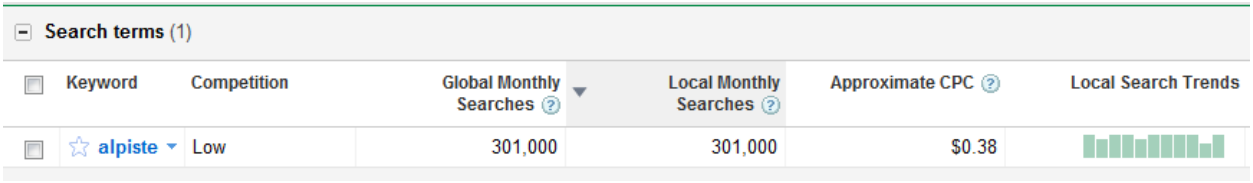


Figure 4.11: Search volume stats for *alpiste* keyword.

Other stats that I selected to view are *Competition*, *Approximate CPC* and *Local Search Trends*. The *Competition* indicator gives me a rough estimate of how competitive would be to rank for the keyword selected; the *Approximate CPC* indicator gives me the approximate cost-per-click (CPC) that I would pay Google every time users clicked on my paid listings (if I were to bid on the *alpiste* keyword), and *Local Search Trends* provides me with bar graphs showing the search trends of the last eleven months.

Figure 4.12 shows the keyword list (using the **Broad** view option) of all keyword variations of alpiste in addition to the key column indicators previously stated. One interesting thing to notice is that most of the keywords returned have low competition.

Keyword	Competition	Global Monthly Searches	Local Monthly Searches	Approximate CPC	Local Search Trends
el **alpiste**	Low	301,000	301,000	$0.38	
que es el **alpiste**	Low	301,000	301,000	$0.36	
que es **alpiste**	Low	301,000	301,000	$0.38	
alpiste alpiste	Low	301,000	301,000	$0.36	
propiedades **alpiste**	Low	33,100	33,100	$0.32	
propiedades de **alpiste**	Low	33,100	33,100	$0.32	
propiedades del **alpiste**	Low	33,100	33,100	$0.30	
alpiste propiedades	Low	33,100	33,100	$0.31	
alpiste de leche	Low	33,100	33,100	$0.41	
alpiste leche	Low	33,100	33,100	$0.41	
leche **alpiste**	Low	33,100	33,100	$0.41	
leche de **alpiste**	Low	33,100	33,100	$0.41	
leche del **alpiste**	Low	33,100	33,100	$0.38	
la leche de **alpiste**	Low	33,100	33,100	$0.39	
beneficios de **alpiste**	Low	22,200	22,200	$0.36	
beneficios del **alpiste**	Low	22,200	22,200	$0.35	
beneficios **alpiste**	Low	22,200	22,200	$0.37	
alpiste beneficios	Low	22,200	22,200	$0.35	

Figure 4.12: Keyword variations and their volume stats for alpiste.

Again, this *Low Competition* indicator gives you a rough idea of how much advertisers are bidding on the specified keyword through Adwords, Google's PPC (Pay-Per-Click) program. Next is brief explanation of PPC, and the keyword and bidding process.

PPC is an online advertising model used to drive visitors to sites through paid listings that appear on the search results whenever users search for certain keywords. Advertisers pay Google

a determined amount every time a searcher clicks on the paid listings; the amount paid per click is the Cost-Per-Click (CPC). In the case with Google Adwords, advertisers bid on keywords they have decided are relevant to their users and website.

For example, Google will display ads when a search query matches an advertiser's chosen keyword list. Such advertisements are known as sponsored links or paid listings, and as seen in Section 2.2.4, these paid listings appear to the right or above the organic search results.

Again, a *Low Competition* indicator is not provided for SEO purposes, it's more used for PPC advertising on the search engine. As you can see from Figure 4.13, when I typed *alpiste* in Google, you can see the paid listings to the right of the organic results; these paid listing are from advertisers who bid on the keyword *alpiste* and wanted their ads to be displayed

From Figure 4.13 below, we can assume that these advertisers are paying Google approximately $0.38 each time a user clicks on any of their paid listings. What was interesting to see was that there were no paid listings at the top of the organic results, only to the right.

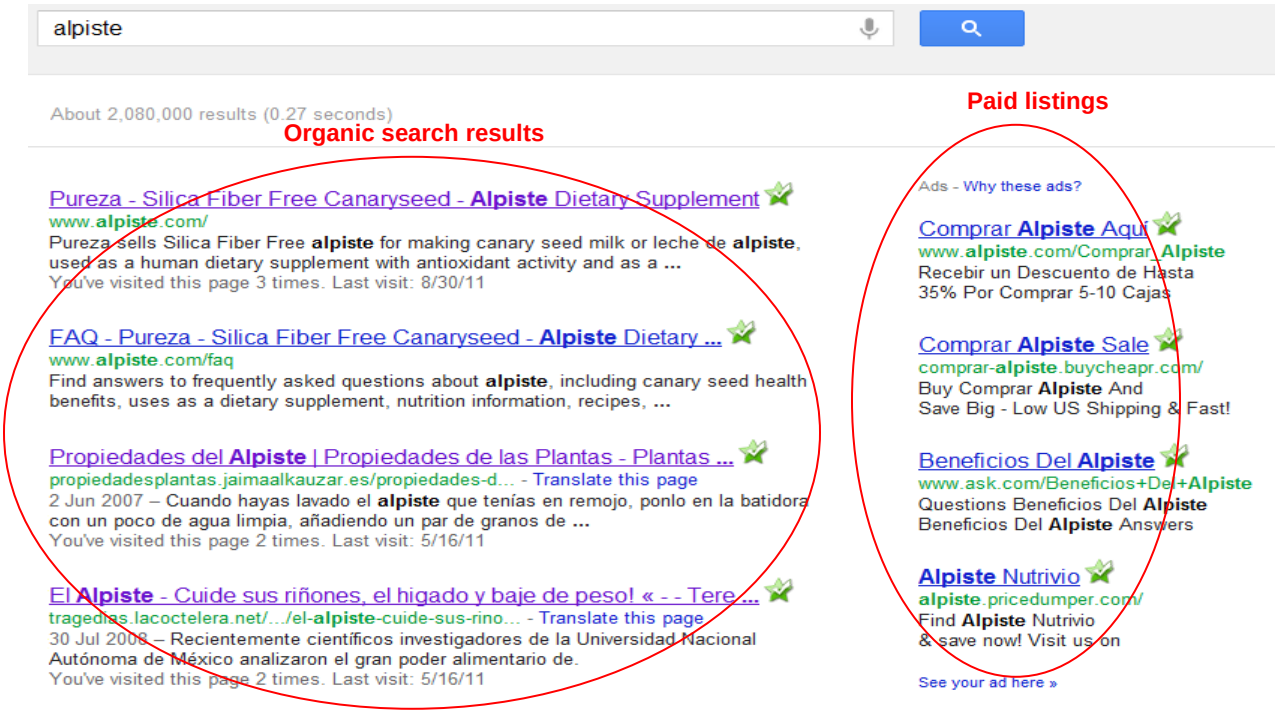
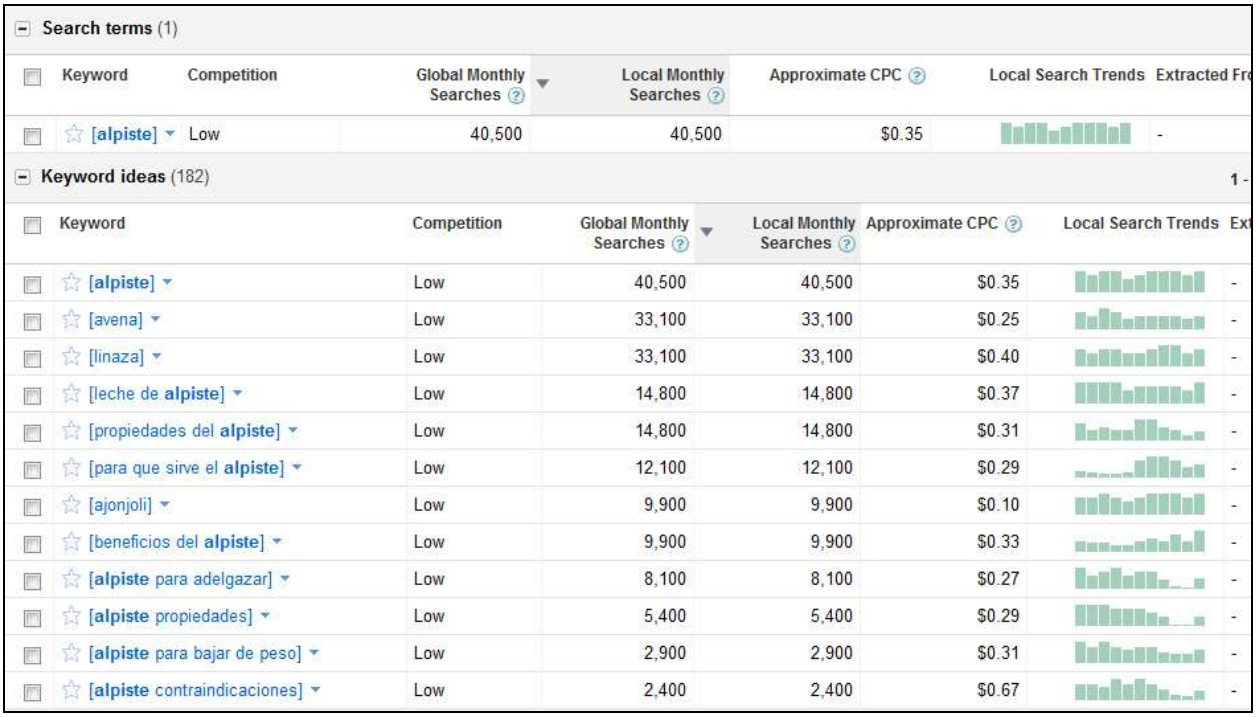


Figure 4.13: Organic results and paid listing results for the *alpiste* keyword.

In continuing with the keyword research process, then I switched to **Exact** from the *Match Types* criteria, I received a different set of keywords from Google. But what's interesting

to note is that the *Monthly Search Volume* has now decreased dramatically. In **Broad Match**, the keyword *alpiste* was shown to have an average of 301,000 monthly searches, but when I switched to **Exact Match** it dropped to 40,500. This is expected because **Broad Match** search volume includes all related terms, synonyms, and related grammatical forms; whereas **Exact Match** will only include search volume for the specified keyword as it is typed.



Search terms (1)

Keyword	Competition	Global Monthly Searches	Local Monthly Searches	Approximate CPC	Local Search Trends	Extracted Fr
[alpiste]	Low	40,500	40,500	$0.35		-

Keyword ideas (182) 1 -

Keyword	Competition	Global Monthly Searches	Local Monthly Searches	Approximate CPC	Local Search Trends	Ex
[alpiste]	Low	40,500	40,500	$0.35		-
[avena]	Low	33,100	33,100	$0.25		-
[linaza]	Low	33,100	33,100	$0.40		-
[leche de alpiste]	Low	14,800	14,800	$0.37		-
[propiedades del alpiste]	Low	14,800	14,800	$0.31		-
[para que sirve el alpiste]	Low	12,100	12,100	$0.29		-
[ajonjoli]	Low	9,900	9,900	$0.10		-
[beneficios del alpiste]	Low	9,900	9,900	$0.33		-
[alpiste para adelgazar]	Low	8,100	8,100	$0.27		-
[alpiste propiedades]	Low	5,400	5,400	$0.29		-
[alpiste para bajar de peso]	Low	2,900	2,900	$0.31		-
[alpiste contraindicaciones]	Low	2,400	2,400	$0.67		-

Figure 4.14: Exact Match results for *alpiste* keyword.

Even though the monthly search volume for *alpiste* dropped to 40,500 when switched to the **Exact Match** option, this high search volume meant there was a significant interest in this keyword as seen from Figure 4.14 above. In addition to the *alpiste* keyword, people were also searching for other keyword variations such as *propiedades del alpiste*, *beneficios del alpiste*, and *alpiste propiedades*; all of these having 14800, 9900, and 5400 monthly searches respectively.

So for an effective keyword research using Google's KST, it's always good to look at different *Match Types* (i.e. **Broad**, **Exact**, **Phrase**) to brainstorm for keyword ideas and see what people are typing into the search engine; but using the search volume stats from the **Exact** option

will give the SEO engineer a more accurate number of the true monthly search volume and the keywords the users are typing into Google.

This concludes the process of keyword brainstorming, discovery and analysis; the next section will go further into more keyword search analysis and keyword selection for the SEO strategies.

4.3.3 Keyword Selection for SEO

As explained previously in section 4.3.2, once I had spent adequate time looking at all the keyword variations and keyword statistics such as monthly search volumes and competition, I needed to choose a keyword that would be the subject of my SEO experiment. It's critical to spend adequate time going through the keyword research and discovery process because some keywords cannot be found in any other way.

You cannot randomly choose keywords that you "think" users will search for to get to your site; a systematic approach has to be followed in order to discover the keywords that will be most effective in any SEO strategy. It's important to know the exact keywords users are typing in the search engines to find information online. The worst case scenario would be to optimize a website around keywords that people are not using.

For this research, the target keyword selected for the SEO implementation on the experimental site was *beneficios del alpiste,* which translates to "alpiste benefits" in English. The reason why it was chosen and the keyword research analysis will be explained next.

Having selected **Exact Match**, **All Countries** and **Spanish** as the criteria in Google's KST, Figure 4.15 shows the top keyword results based on the average monthly search volume. And in looking at the returned keyword results, there seemed to be low competition. But as explained before, the **Low** competition indicator applies more to PPC.

According to KST, my target keyword (*beneficios del alpiste*) had global monthly search volume of 9,900; and in looking at the **Local Search Trends** bars in Figure 4.15, its search volume has remained relatively constant with a slightly increase in the last four months.

Keyword	Competition	Global Monthly Searches	Local Monthly Searches	Approximate CPC	Local Search Trends
[alpiste]	Low	40,500	40,500	$0.35	
[leche de alpiste]	Low	14,800	14,800	$0.37	
[propiedades del alpiste]	Low	14,800	14,800	$0.31	
[para que sirve el alpiste]	Low	12,100	12,100	$0.29	
[beneficios del alpiste]	Low	9,900	9,900	$0.33	
[alpiste para adelgazar]	Low	8,100	8,100	$0.27	
[alpiste propiedades]	Low	5,400	5,400	$0.29	
[alpiste para bajar de peso]	Low	2,900	2,900	$0.31	

Figure 4.15: Top keywords (by monthly search volume) in Exact Match.

As expected, the keyword *alpiste* had the highest monthly search volume. For my research, I was looking for a keyword that contained at least three words and had a decent amount of search volume. I felt that the keyword *beneficios del alpsite* (which means "alpiste benefits") was good choice because it implied that people were searching for information on the health benefits of *alpiste*. Furthermore, this keyword seemed like a good selection for creating pages with content about the benefits of *alpiste*. Once I had selected my target keyword, I needed to do further research to validate my selection.

Once the target keyword was selected, I turned to *Google Insights* to perform further research and look at the worldwide keyword search trend. Figure 4.16 shows the worldwide popularity of *beneficios del alpiste*. You can see that prior to 2009 there were few searches, but since March 2009, the popularity of the keyword started to increase and it has remained popular ever since. Again, this was good to know because I wanted a keyword that was popular among searchers and one that would drive traffic to the experimental site.

Figure 4.16: Worldwide search trend for *beneficios del alpiste* keyword.

Continuing with the keyword research, Figure 4.17 below shows the regional interest of the selected keyword. As you can see, only four countries were shown as having the most interest for *beneficios del alpiste*, the highest being Mexico with a normalized value of 100 and the lowest being the United States with a normalized value of 3.

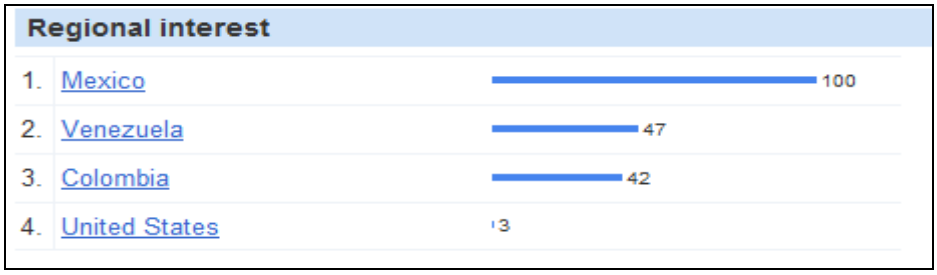


Figure 4.17: Regional interest for *beneficios del alpiste* keyword.

Analyzing further, you can see in Figure 4.18 how *Google Insights* also displayed **Top searches** and **Rising searches** related to *beneficios del alpiste;* these are additional ideas for popular keywords to consider and it further validates high interest in the *alpiste* search query. Furthermore, you can also see that the target keyword appeared in both, **Top searches** and **Rising searches**; this was also good to know because it further confirmed its popularity and for *beneficios del alpiste*, the target keyword.

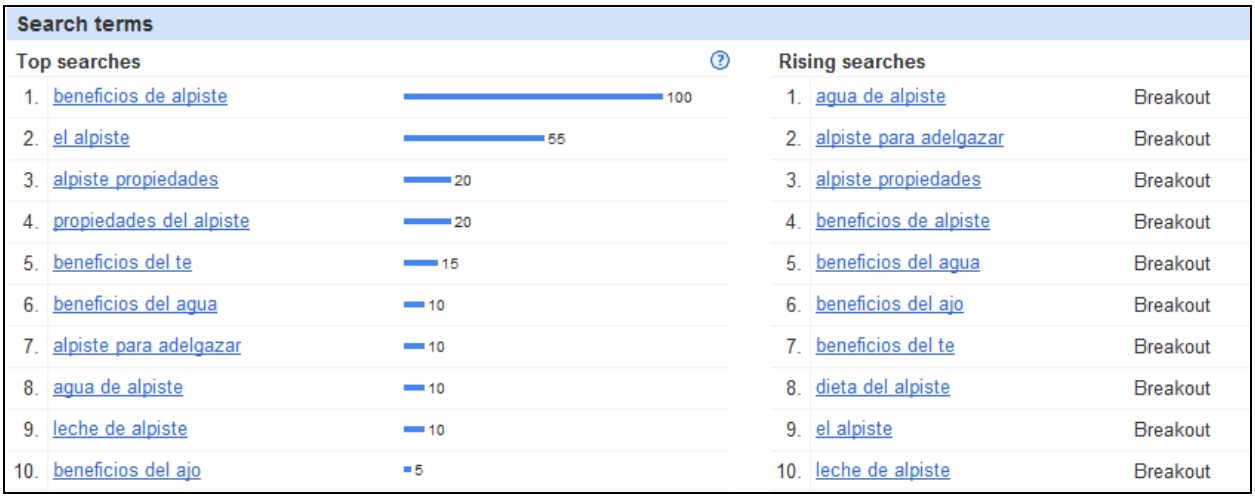


Figure 4.18: Top searches and Rising searches related to *beneficios del alpiste* keyword.

Once I completed the previous step, I went to Google and entered *beneficios del alpiste* into the search bar to see the total number of pages that were indexed (i.e. indexed pages are those that exist and are listed in the database of the search engine). You can see from Figure 4.19 that there were about 106,000 pages indexed for the search query *beneficios del alpiste*.

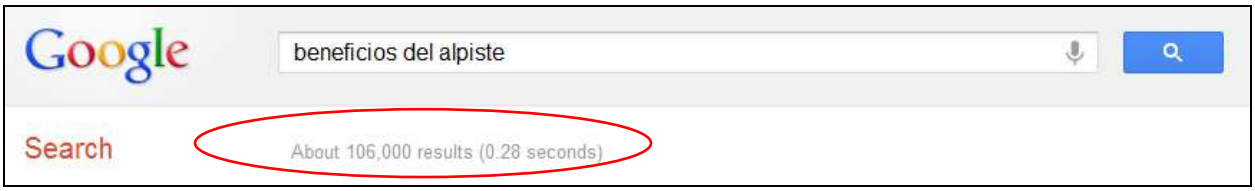


Figure 4.19: Number of indexed pages related to *beneficios del alpiste* keyword.

This number of page results may seem high; it indicates the number of web pages that Google has identified as relevant to *beneficios del alpiste*, but taking into account the size and how fast the Internet is growing by the day, this number is relatively small. Also, the 106,000 pages can also be considered my competition; these are the pages my experimental site would be competing with to achieve placement in the first page of Google.

It's important to note that the above steps were performed for the target keyword **only** (*beneficios del alpiste*). As you can see, the keyword research process can be time consuming and it requires data analysis. Although there are tools that can automate many of these tasks, going through the keyword data still requires human intervention for data analysis.

And as stated before, it's crucial to complete the keyword research before doing any SEO; going through this process you can discover whether optimizing for any selected keyword (or set of keywords) is feasible or not. Moreover, you can uncover new keywords, or you may find keywords that you never thought your target users were searching for; in this case, the target users are the type of visitors you want to attract to your website through the SEO strategies.

Once I had validated and verified my target keyword for optimizing the experimental site, I needed to install a tracking system on the site to begin gathering data and start measuring the effectiveness of the SEO strategies. Installing a tracking system from the start would create a baseline for measuring the effectiveness of the SEO.

4.4 Data Collection

The tracking tool for website data collection is Google Analytics. This is a "free service offered by Google that generates detailed statistics about the visitors to a website" [46]. This is the most widely used traffic analysis tool in the world, currently used by 55.6% of all the websites [16]. It's a very powerful tracking tool in that it provides you with detailed data on how visitors are arriving at your site, what keywords are driving the most visitors, what countries are sending the most visitors and much more detailed statistics for analysis which I will cover in the next sections.

4.4.1 Google Analytics installation

Setting up Google Analytics on the site was quite simple. Since this is a free service offered by Google, all you need to do is have an existing account with them, such as Gmail. I have a Gmail account and so in order to setup the tracking tool on *casadelalpiste.com* I had to first login to Google Analytics, go the Account Administration panel, select New Account and fill the required fields as seen in Figure 4.20 below.

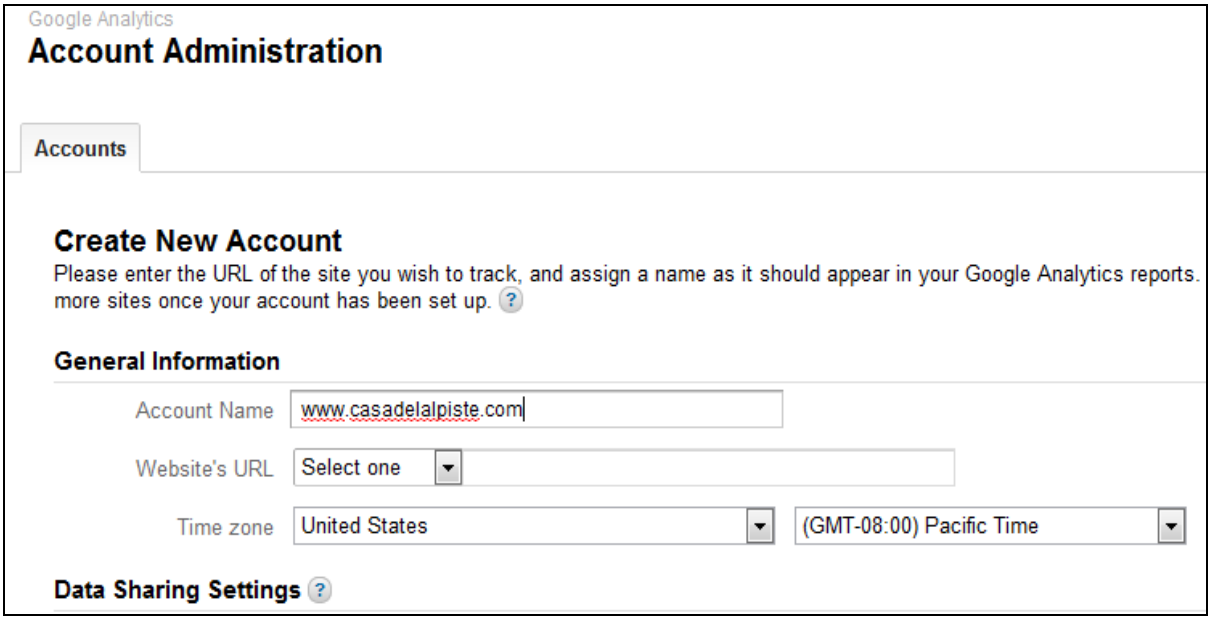


Figure 4.20: Google Analytics configuration for *casadelalpiste.com*

Once all the fields were filled out and I agreed to their Terms of Service, a new window was provided with a Javascript code. You can see a sample of the code in Figure 4.21.

Copy the following code, then paste it onto every page you want to track immediately before the closing </head> tag.

```
<script type="text/javascript">

 var _gaq = _gaq || [];
 _gaq.push(['_setAccount', 'UA-22089619-1']);
 _gaq.push(['_trackPageview']);

 (function() {
  var ga = document.createElement('script'); ga.type = 'text/javascript'; ga.async = true;
  ga.src = ('https:' == document.location.protocol ? 'https://ssl' : 'http://www') + '.google-analytics.com/ga.js';
  var s = document.getElementsByTagName('script')[0]; s.parentNode.insertBefore(ga, s);
 })();

</script>
```

Figure 4.21: Google Analytics Javascript code for tracking traffic stats in *casadelalpiste.com*

This Javascript code needed to be copied and pasted before the </head> tag (in the HTML code) of every web page that needed tracking analysis. You can see this how I implemented this on the experimental site in Figure 4.22; it's a screen-capture of how all the site's pages have the Google Analytics tracking code within the HTML source code.

```
35 <script type="text/javascript">
36         var analyticsFileTypes = [''];
37         var analyticsEventTracking = 'enabled';
38 </script>
39 <script type="text/javascript">
40         var _gaq = _gaq || [];
41         _gaq.push(['_setAccount', 'UA-22089619-1']);
42         _gaq.push(['_trackPageview']);
43
44         (function() {
45                 var ga = document.createElement('script'); ga.type = 'text/javascript'; ga.async
46                 ga.src = ('https:' == document.location.protocol ? 'https://ssl' : 'http://www')
47                 var s = document.getElementsByTagName('script')[0]; s.parentNode.insertBefore(ga,
48         })();
49 </script>
50 </head>
51 <body class="custom">
```

Figure 4.22: Google Analytics Javascript tracking code placement in HTML pages.

Once Google Analytics was configured properly and installed, the experimental site (*casadelalpiste.com*) was now ready to start gathering key web statistics. The next step was to start developing web pages of content and start the SEO implementaion.

4.5 On-Page SEO Strategies

As discussed in chapter 2, on-page SEO is the set of techniques that the SEO engineer can apply to resources on-page (composition of web pages, use of correct tags and keywords) that have measurable impact on the way the page or set of pages being optimized is ultimately ranked by Google. These techniques are in contrast to off-page techniques which are discussed in the following Section.

The right time to start on-page SEO is after the obvious first steps have been completed: initial keyword research, website setup (web hosting, domain name, etc…) and Google Analytics configuration for data collection. At this point, though, there's no actual content, so there are no pages to be optimized or rank for. So now the critical content must be developed and uploaded. Even some of the earlier steps can be positively influenced by knowledge of good SEO practices (as discussed earlier in section 4.2.2 on web hosting decisions). But the real on-page SEO techniques begins with the web page content creation of the site and page optimization.

On-page SEO is what the web developer does on the website to influence the rankings on the search engine results hopefully in a positive direction. It includes many factors, but steps such as an on-going keyword research and a well planned out web strategy from the start is crucial.

This section will cover the specific on-page SEO strategies that were implemented on the experimental site. Before embarking on any off-page SEO strategies, which will be covered in section 4.6, you want to make sure that your on-page SEO has been implemented correctly. This is because this process is fully under your control; as owner of the site, you have total control of how the site is structured, designed and how SEO gets implemented. And when done correctly, on-page SEO can have positive effects on how search engines view the site, thus increasing your rankings when people search for the keyword you have selected and many others.

I will begin with the optimization of meta tags; these are HTML labels that are added to website pages. The use of these tags help the search engines correctly categorize web pages and how to treat each page when indexed. When updating the meta tags, you want to avoid duplication because this can negatively affect the site's rankings, so the goal is to have unique meta tags on each page. There are many meta tags, but the following will be the focus of this research and SEO experiment: description, keywords and robots meta tag.

I will also cover effective ways to optimize URLs, web page content, correct usage of heading tags, images, use of the *rel=nofollow* attribute, keyword placement, and sitemaps. Effective use of the previous on-page factors can greatly help your site achieve higher search engine rankings when implemented correctly.

4.5.1 Title tag

The title tag defines the page title and informs the search engine what the page is about. In other words, it describes the overall theme of your web page. Google recommends choosing "a title that effectively communicates the topic of the page's content" [19] so that its search ranking algorithm can better categorize it and know what the page content is about. Moreover, the title tag is what gets displayed in Google's search results as the hyperlink title; an example of the title tag in the search results will be given next.

Figure 4.23 shows the top three organic results for the keyword *vacation packages to paris*. In this example, you can see that these top search results pages have been optimized for *vacation packages to paris* because words from my search query appear in bold within the hyperlink title. Having the keywords in the title tag, which appears as the hyperlink in Google's search results, informs the searcher that if he or she clicks on the link, the resulting destination page will be relevant to the search query.

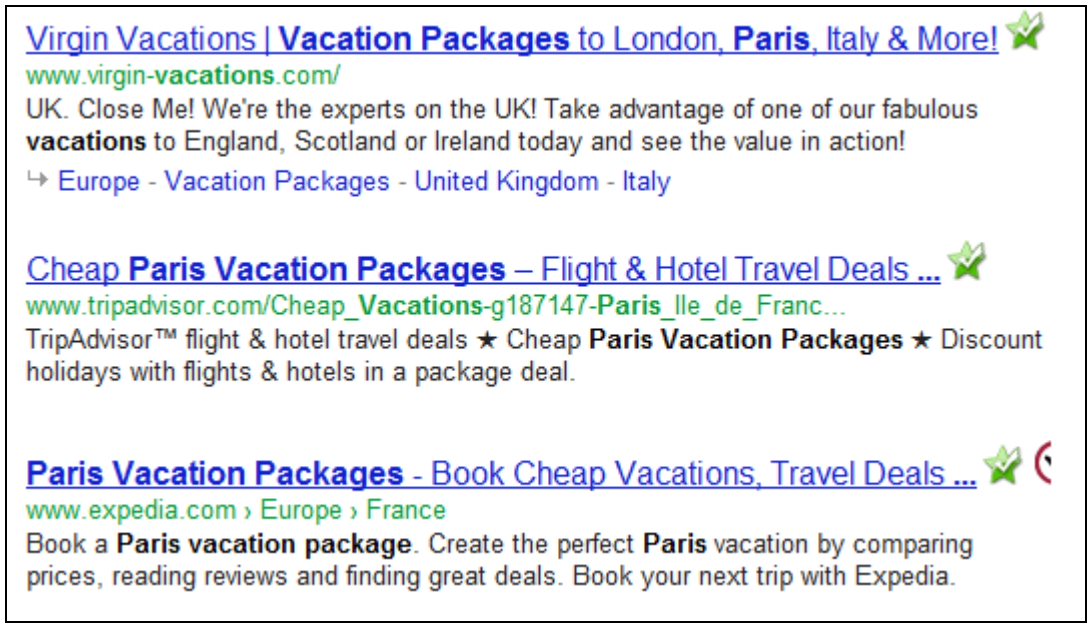


Figure 4.23: Top 3 organic results for the keyword *vacation packages to paris*.

When I click on the result No. 2 of the search engine results page, I'm taken to page that's optimized for the keyword *vacation packages to paris*. You can see in Figure 4.24 that the title tag appears at the very top of the web browser. This tells me this page is following correct SEO guidelines and therefore is being rewarded with top rankings for the keyword discussed.

Figure 4.24: Example of Title tag optimization for *vacation packages to paris*.

For the experimental site (*casadelalpiste.com*), the title tag optimization for my target keyword (*beneficios del alpiste*) was implemented as follows: once I decided on the page to be optimized for this keyword, I modified the HTML document by adding the keyword to the title tag; this process is done when coding and developing the web page. You can see in Figure 4.25 a snapshot of the HTML source code and how I included the keyword into the title tag.

```
<html xmlns="http://www.w3.org/1999/xhtml" dir="ltr" lang="en-US">
<head profile="http://gmpg.org/xfn/11">
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8" />
<title>Beneficios del Alpiste en el Ser Humano</title>
```

Figure 4.25: Title tag optimization for *beneficios del alpiste* keyword.

Also, in order to create an even more effective title tag, it's best to keep the length of the title tag to less than 68 characters (including spaces). This is because 68 is the maximum number characters that get displayed in Google's search results title. When the title length goes beyond the limit, Google will only display up to 68 characters followed by three dots to identify a continuation. You can see this scenario in Figure 4.26. The first result has a title that goes beyond 68 characters, the second result is less than the limit. It's interesting to see that the third one is exactly 68, which is definitely not just an odd coincidence.

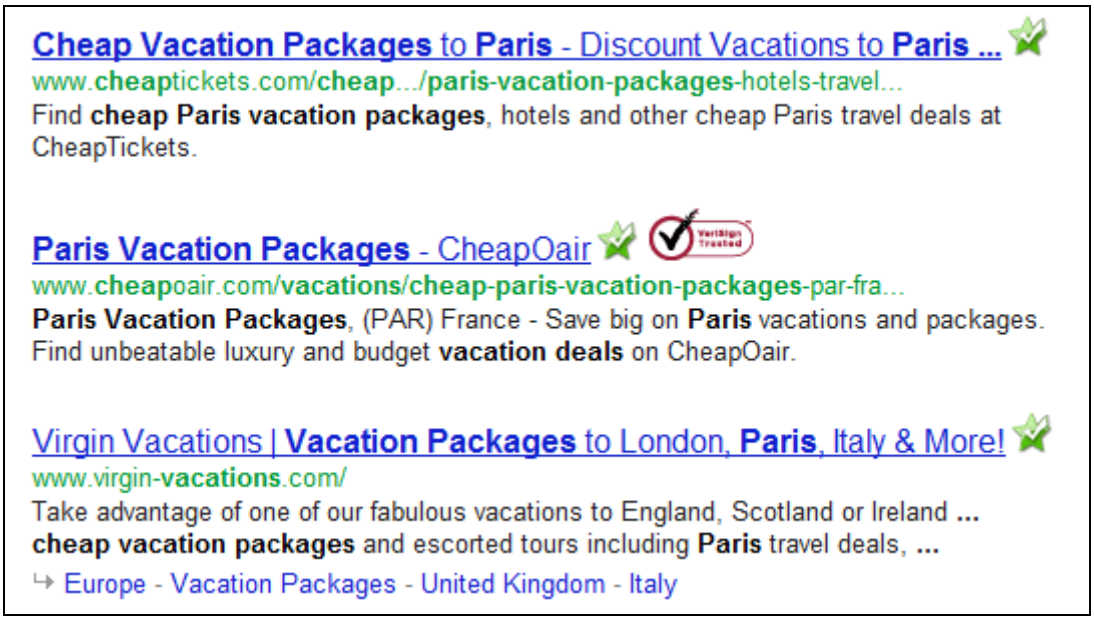


Figure 4.26: Top search results for *vacation packages to paris* keyword.

It's important to note that going over the maximum limit will not affect the page's ranking, this is only important when you want the searcher to read the entire message of the title tag and for aesthetics reasons. As long as the target keyword is included in the title tag, this is a first step to having an optimized page because having the keyword embedded into the title tag informs Google the main theme of the page.

To summarize this section for title tag optimization, below are the key factors to keep in mind when implementing this on-page SEO method:

- Avoid title tag duplication. If possible, each page on the site should have its own unique title tag.
- Always include the keyword in the title tag.
- Keep the title tag short. Do not go over the maximum limit, otherwise only a portion will be displayed in the search results (at the time of this writing the maximum is 68 for Google).
- Use short and descriptive titles to let users and search engines know what the page content is about.
- Avoid keyword stuffing. Do not overuse the keyword in the title tag; Google discourages of this practice. If found to be using these unapproved practices, your SEO rankings can be negatively affected.

4.5.2 Description meta tag

As the `description meta tag` implies, it provides a description of the web page or site. Google explains that "a page's description meta tag gives Google and other search engines a

summary of what the page is about" [19]. Also, this description is what appears (in most cases) in Google's search results, just below the title. I say in most cases because Google uses this description on the search results in most cases, but it can also include a snippet of text from anywhere on the page that it considers relevant to the search query; we will see an example of this next.

You can see the use of the `description meta tag` in Figure 4.27 when a search for *cheap school books* is performed. The words from the keyword are bolded or enhanced to give it its attention.

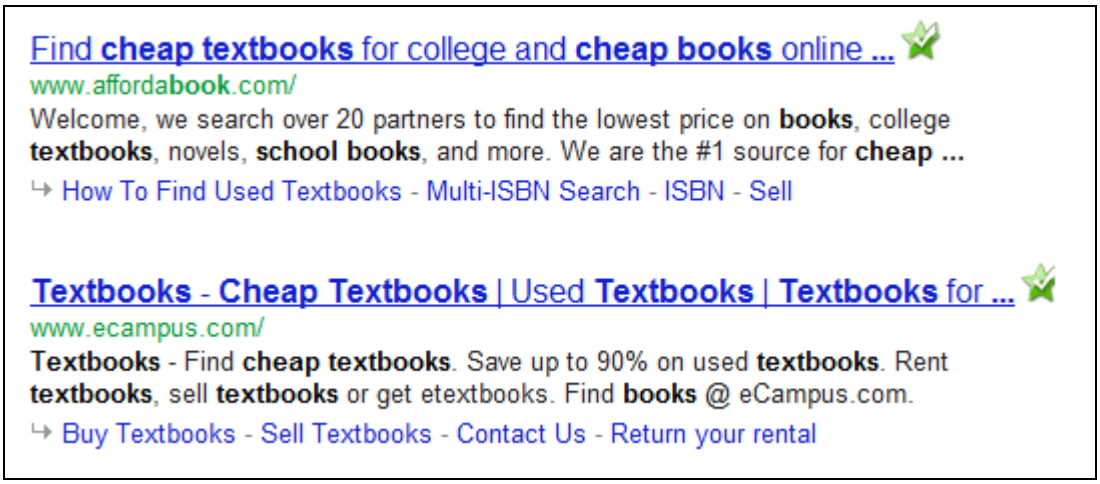


Figure 4.27: Use of the *description meta tag* in the search results.

The idea behind the `description meta tag` is to provide an accurate and compelling description of what users will find once they click on any of the search results and are taken to the destination page. And as explained previously, this description also helps search engines categorize and determine the theme of the page.

The key to having an effective and optimized `description meta tag` is to maintain its consistency with the overall theme of the page. This is what Google recommends because it enhances the overall user experience by letting users know exactly what they will find when they visit the resulting page, and it helps search engines determine the theme of the page.

In order to complete this optimization step, the SEO engineer needs to make sure that the target keyword is included (at least once) in the `description meta tag`, but keep from

overusing it. If excessive use of the keyword happens, the page runs the risk of getting penalized which can have a negative impact on the search rankings.

In Figure 4.28, you can see the optimization of the `description meta tag` as seen in the HTML source code of the page, notice how the target keyword (*beneficios del alpiste*) is included in the description text.

```
<meta name="description" content="Beneficios del alpiste en el ser humano y las
propiedades curativas del alpiste." />
<meta name="keywords" content="alpiste para que sirve, propiedades del alpiste
para adelgazar, dieta de alpiste, beneficios del alpiste en el ser humano" />
```

Figure 4.28: Optimization of the *description meta tag* for *beneficios del alpiste*.

As mentioned before, the text from the `description meta tag` is what Google displays in the search results page most of the time, but it may also "choose to use a relevant section of your page's visible text if it does a good job of matching up with a user's query" [19]. Even though I have included a short text in the `description meta tag` of the optimized web page, Google displays a snippet from the first paragraph of the page. You can see in Figure 4.29 how Google has determined not to include the text that I used for my `description meta tag` and instead selected to display part of the first paragraph. I'm supposing it might be because my description is too short; I really can't explain this.

Figure 4.29: Search results for *beneficios del alpiste*.

For this reason, it's very important that the first few lines of the first paragraph on the page be well-written and say something descriptive about the theme of the page; more detail on content optimization will be covered in Section 4.5.6.

When the reader quickly glances at the search results, the keywords in bold will be noticed first. The description text will catch the reader's eye around the edges of the keywords, and will most likely contain secondary words or visual cues that motivate the reader to choose what link as the most relevant to the original search.

So good description text will make your link stand out favorably to the reader from its surrounding links on the search results page. In fact, there is some overlap between the writing style for effective web pages and journalism's "inverted pyramid" style of writing that puts the article's most important points at the top in an extremely concise format, with longer detailed lists of facts pushed further down the page.

To summary this section on effective use of the `description meta tag`, below are a few best practices to consider:

- Use your target keyword in the description meta tag. The contents of the description meta tag gets displayed in Google's search results. By having a compelling and descriptive text with your keywords, searchers are more likely to click on your link and visit your page.
- Write unique description meta tags. Whenever possible, avoid having exact description meta tags on different pages of the website.
- Avoid using generic descriptions. You want to accurately summarize the page's content to inform the users, and Google, what the page is about.

4.5.3 Effective use of robots.txt

The purpose of `robots.txt` is to signal the search engines whether to crawl your site and what locations (i.e. pages or directories) should not be accessed. Search engines send their crawlers every so often to crawl and index the content of web sites. And there may be times where when you want to keep certain directories or folders from getting crawled (for security reasons). In order to setup the permissions to certain directories or pages, you must configure the `robots.txt` file using the settings as seen in Figure 4.30.

Notice how the property `User-agent` is set to the wildcard * symbol; this configuration instructs all spiders (i.e. *googlebot*, *msnbot*, *yahoo-slurp*, etc…) to crawl and index all pages in the site. Furthermore, notice that the `robots.txt` file must be located in the root directory of the web server.

Figure 4.30: Sample configuration of *robots.txt* file and its location.

If I had wanted to be more specific as to what web crawlers I wanted to grant access to crawl my site, I could have configured the settings of the `robots.txt` as seen in Figure 4.31 below. Using this configuration I have specifically stated that I'm only allowing Google, Yahoo! and Microsoft's web crawlers to index the entire site, and I have disallowed Baidu's web crawler (China's most popular search engine) from indexing my site. In addition, I have disallowed all access to the /`secure/` folder to all crawlers.

```
User-agent: Googlebot
Disallow:
User-agent: MSNBot
Disallow:
User-agent: Slurp
Disallow:
User-agent: baiduspider
Disallow: /
User-agent: *
Disallow:
Disallow: /secure/
```

Figure 4.31: Sample configuration of the *robots.txt* file.

To summarize this section on effective use and configuration of the `robots.txt` file, below are some best practices to keep in mind:

- The file must be named `robots.txt`.
- Place `robots.txt` in the root directory of the website.
- Make sure proper configuration is in place to allow or disallow selected directories to get crawled; having the incorrect configuration may prevent your site from getting indexed, which means none of your page will appear in the search results.

4.5.4 Optimized URLs

An effective SEO strategy is to define a website structure that is user readable and intuitive. This means organizing the pages of the website into directories or folders with readable and meaningful names, down to the names of the individual web page files. In doing this, not only will the site appear well organized to a human reader, "but it could also lead to better crawling of your documents by search engines" [19]. You want to have "search engine friendly" (i.e. easy to understand) URLs for users who might want to link to your pages; long and cryptic URLs will make it difficult for users to understand.

An optimized URL is one in which the URL text is self-explanatory and self-documenting. The text of the URL gives the reader a brief description of the theme of the page's content. An un-optimized URL in contrast, consists of a long directory path with technical and non-intuitive folder and file names, which may have meaning in the context of the server file system, but which mean nothing to a human reader.

Figure 4.32 shows an example of a non-intuitive and cryptic-looking URL. Users who come across these types of un-optimized URLs may not be able to know what the page content is about, so they will be less likely to visit or link to the page.

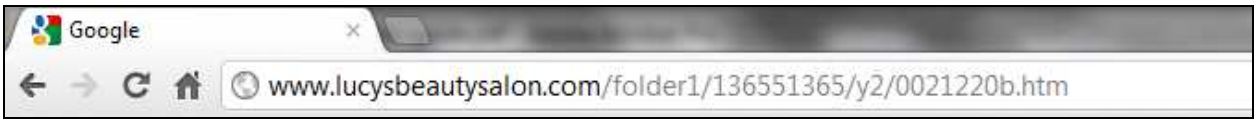


Figure 4.32: Example of a non-intuitive and cryptic-looking *URL*.

In Figure 4.33 you can see a better structured and highly optimized URL. Users who come across similar URLs such as this will have a better idea of what the web page is about. This is helpful to both users and search engines, because self-explanatory and self-documenting URLs provide more useful information about the page and its content.

Figure 4.33: Better structured and highly optimized *URL*.

Another important reason why you want a descriptive and optimized URL is because it appears on the search results, just below the title. The following example will illustrate its importance: I performed a search query in Google for the keyword *rare baseball cards for sale;* in Figure 4.34 you can see the top results. Notice how the keywords appear in bold within the URL; this occurs whenever they match the user's search query.

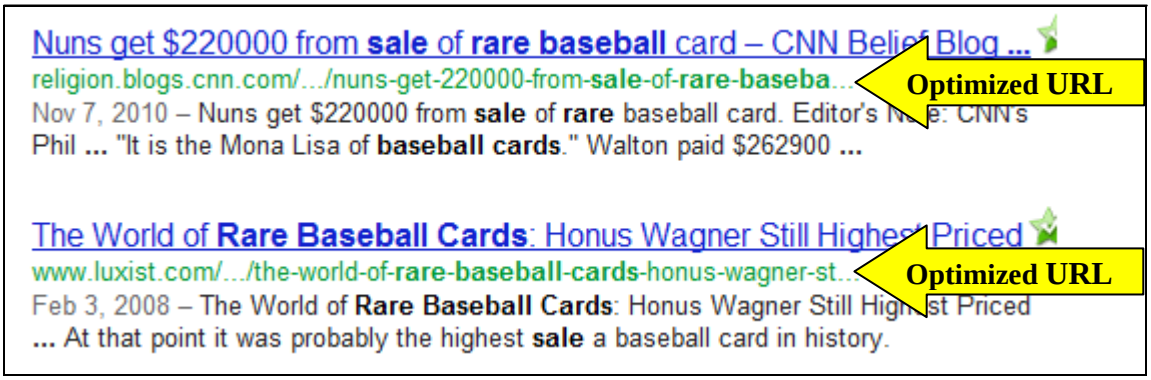


Figure 4.34: Top search results for the keyword *rare baseball cards for sale*.

This is important because you want to let the user know that if they click on the specified URL, they will find what they were looking for. Furthermore, if the URL has been optimized with the correct keyword, the searcher is more likely to click on the link.

Also, always provide one version of a URL to reach a web page. For example, avoid having the following scenario whenever creating URLs that takes users to the same location:

- domain.com/optimized-page.htm
- sub.domain/ optimized-page.htm
- www.domain.com/optimized-page.htm

Although all three URLs will take users to the same page, to the search engines these URLs are completely different. When this occurs, you're splitting "the reputation of that content between URLs" [19], which means all of these URLs are competing against each other in the search rankings.

If users are accessing the same pages through different URLs, Google advises that "you use a server-side 301 redirect. This is the best way to ensure that users and search engines are directed to the correct page. The 301 status code means that a page has permanently moved to a new location" [27].

In looking at the previous example of different URLs that take users to the same location, this is how to implement a 301 redirect: you first need to select one of those URLs as "your preferred (canonical) destination, and use 301 redirects to send traffic from the other URLs to your preferred URL" [27]; in order to apply the 301 redirect, you have to configure the `.htaccess` file. This is a distributed configuration file for websites that are hosted on servers running Apache that provides "a way to make configuration changes on a per-directory basis" [28]. Moreover, this file may contain one or more configuration directives and must be placed in the server's root directory. Figure 4.35 shows how the `.htaccess` file has been configured to force the use of **www.example.com** instead of **example.com**. In other words, this configuration sends users and search engines from **example.com** to **www.example.com.**

```
RewriteCond %{HTTP_HOST}   !^www\.example\.com [NC]
RewriteCond %{HTTP_HOST}   !^$
RewriteRule ^/(.*)         http://www.example.com/$1 [L,R]
```

Figure 4.35: Sample configuration of the *.htaccess* file.

Just to clarify, Google will not penalize your site if your site does not follow search engine friendly URL practices. Google is "good at crawling all types of URL structures, even if they're quite complex, but spending the time to make your URLs as simple as possible for both users and search engines can help" [19], which can also help in your SEO rankings.

For the experimental site, I made sure all the URLs were search engine friendly and that it each URL described what the web page content was about. In Figure 4.36 you can see the structure of the optimized URL (on the experimental site); notice how the target keyword (*beneficios del alpiste*) was embedded into the URL.

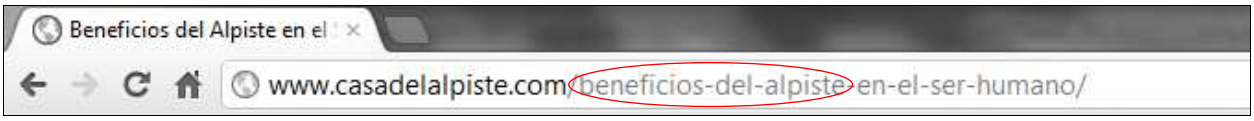


Figure 4.36: Optimized *URL* for the keyword *beneficios del alpiste.*

To summarize URL optimization, here are few key points to consider when structuring the URLs of your web site:

- Use descriptive words in your URLs. Incorporate the target keyword into the URL; users will remember them much easier and are more likely to link to your page.
- The URL gets displayed in the search results. Use the keyword in the URL to let search engines and users what the page is about.
- Create a simple site directory structure "that organizes your content well and makes it easy for visitors to know where they're at on your site. Try using your directory structure to indicate the type of content found at that URL" [19].
- Provide only one version of a URL to reach a page. Avoid having different versions of URLs access the same content, such as: `domain.com/page.htm` and `sub.domain/page.htm`. Although both URLs will take users to the same page, search engines view these two URLs as completely different, thus splitting "the reputation of that content between URLs" [19].

4.5.5 Content first

Search engines process web pages from left to right, top to bottom. So a strategic way to structure a web page is to place the most important text at the top. The first paragraph should contain your target keyword (or set of keywords) and it should describe what users will find if they continue reading. As discussed in Sub-section 4.5.2, Google uses the text from the `description meta tag` in the search results, but it may also use other text found on the page that it considers relevant to the user's search query. But in many cases, the text used comes from the first paragraph of the page as you will see in the next example.

Figure 4.37a, Figure 4.37b and Figure 4.37c show how I performed a Google search query for *rare baseball cards*. I took the top result; notice how the Google uses a snippet of text from the first paragraph in the destination page within the description of the Google results.

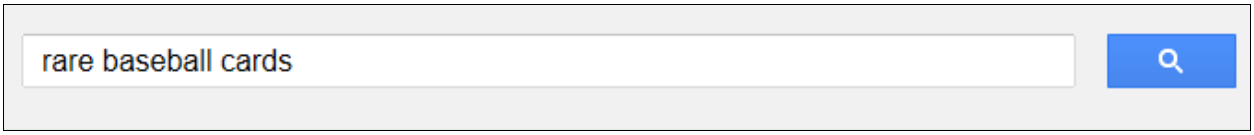


Figure 4.37a: Search engine form with the keyword *rare baseball cards*

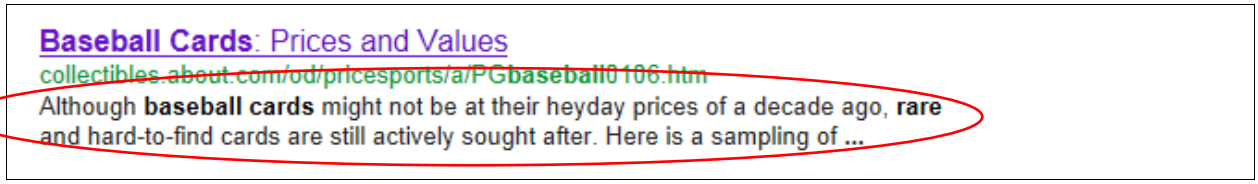


Figure 4.37b: Search results for the keyword *rare baseball cards*

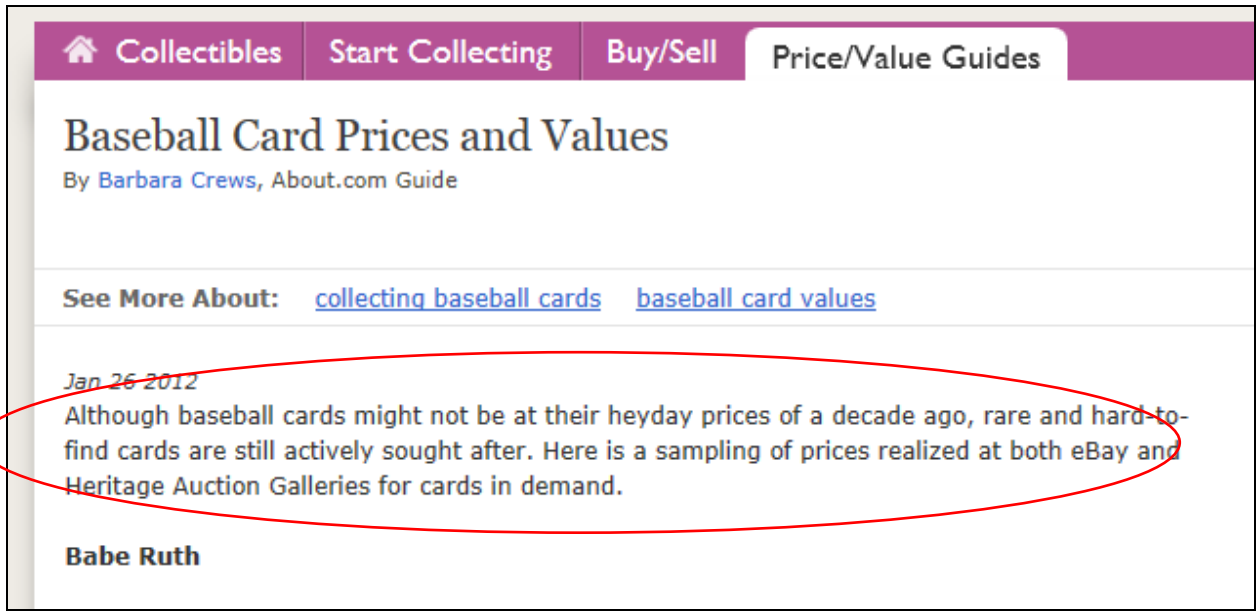


Figure 4.37c: Actual page in Google's top search results for the search query *rare baseball cards*

It's important to keep this optimization factor in mind because you want your users to read the full page. By providing the most important text and the most relevant keywords at the beginning of the first paragraph, you'll have a better chance of convincing users to continue reading further; moreover, users will be convinced that they've visited a relevant web page because the text in the page contains the keywords that were used. Also, this is part of the overall good user experience that Google is looking for, and when strategically applied, your page can be rewarded with higher rankings.

To summarize this section, here are few key points to remember:

- Use descriptive text and your keyword in the first paragraph of the web page you're optimizing.
- Know that Google uses the content of the `description meta tag` in the search results page, but it may also use other text from the page that it considers relevant to the search query. In most cases, the text used may be found within the first paragraph of the web page (as seen in the previous example).

4.5.6 Headings tags

Heading tags are used to emphasize important text and structure on the web page. They also inform the search engines what your page is about; the <h1> tag defines the most important heading and <h6> defines the least important heading, so it's important to include your target

keyword into the headings, most importantly within the <h1> tag. This can be effective because it helps establish information hierarchy within the page being optimized and it helps visitors understand the web page content and how it's organized; it's almost like writing an outline with main points and sub-points.

It's important to note that the <h1> tag displays a larger text format, <h2> displays it slightly smaller than <h1>, and <h3> smaller than <h2>, and so forth. Also, the ranking algorithm assumes that the heading tags establish the information hierarchy of the document; the web page author must therefore use the heading tags correctly to define that hierarchy.

When optimizing the <h1> heading, it's recommended to use only one <h1> in the web page and include your target keyword once; using more than one <h1> heading and repeating your keyword multiple times can raise a red flag for a black-hat technique known as keyword stuffing, which can negatively affect your search rankings.

Figure 4.38 shows part of the HTML source code where the <h1> heading was optimized for my target keyword (*beneficios del alpiste*). Note the HTML source code and how the page is viewed in the browser.

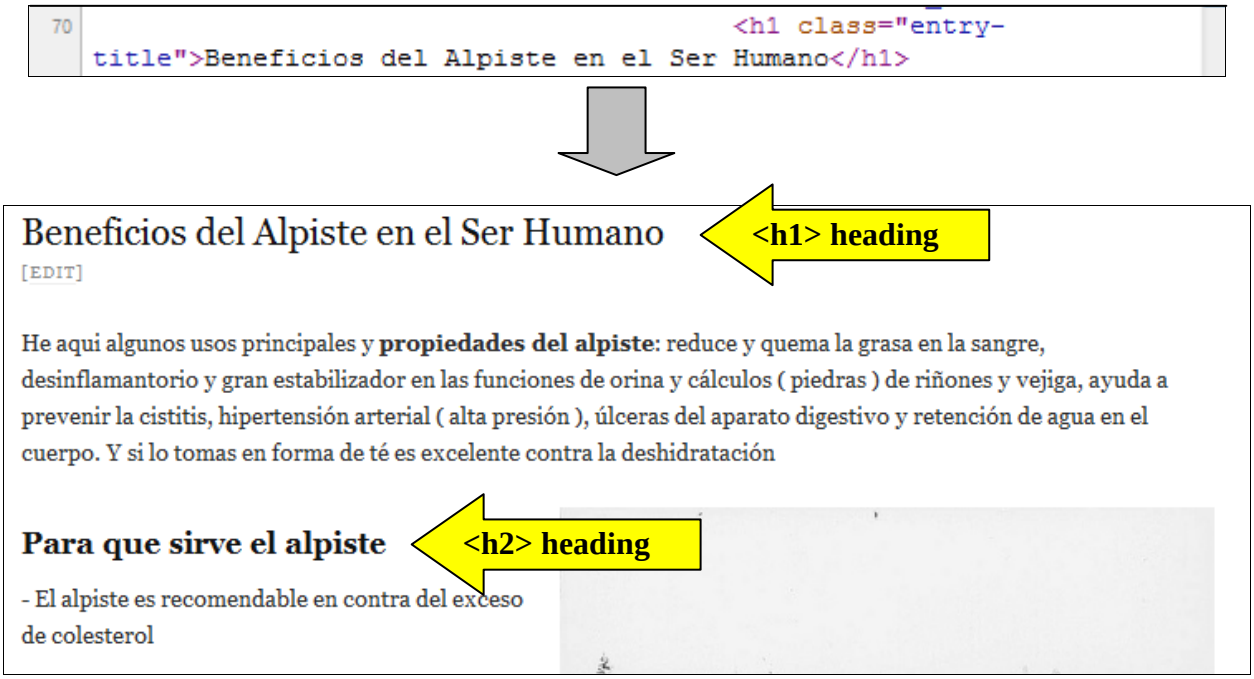


Figure 4.38: *H1* heading optimization for *beneficios del alpiste* and how is viewed in the browser.

To summarize the optimization of heading, consider the following key points:

- Insert your target keyword within the <h1> heading tag.
- Use only one <h1> heading per web page.
- Use the other headings (<h2>, <h3>,<h4>,<h5>, <h6>) to establish information hierarchy; you can also include your target keyword within any of the previous headings.

4.5.7 Images

The content of the modern Web is multimedia in nature, and images play a big role. Although "computer scientists have made great strides in visual semantics, human aids are still needed for computers to understand images on large scales" [20]. For this reason, whenever you use images in your web page, you need to make use of the image `alt` attribute to provide an accurate description of the image you're using.

The `alt` attribute from the `<img>` HTML tag specifies the alternative text of what the image is about in case the image doesn't load or cannot be displayed correctly. This short text description is helpful to the search engines and to people with disabilities so they can understand what's contained on the image.

In regards to SEO, Google takes the keywords from the `alt` attribute into account in their ranking algorithm. In addition, if an image is used as a link, "the `alt` text for that image will be treated similarly to the anchor text of a text link" [19]. So when optimizing your images, it's important to include your target keywords into the image `alt` description text.

Figure 4.39 shows the correct image optimization of the `alt` attribute; you can see how it's being optimized for the keyword: *vintage baseball card*. Also, it's recommended that you name your images with a descriptive name. And if possible, include your target keyword in the naming of the image as seen in Figure 4.39.

```
<img src="1926-vintage-baseball-card.jpg" alt="A 1926 vintage and rare baseball card.">
```

Figure 4.39: Image optimization example for the keyword *vintage baseball card.*

Google says that "optimizing your image filenames and alt text makes it easier for image search projects like Google Image Search to better understand your images" [19]. So whenever a

user is searching for images, it's possible that your optimized images may appear in the image search results, thus driving more users to your site.

Figure 4.40 shows the HTML source code of the optimized web page from my experimental site. Notice that I included the target keyword into the image `alt` attribute and I named the image with a descriptive name that includes the target keyword (*beneficios-del-alpiste.jpg*).

```
<p>
  <a href="http://www.casadelalpiste.com/beneficios-del-alpiste-en-el-ser-humano/">
       <img class="alignright size-full wp-image-574" title="beneficios-del-
       alpiste" src="http://www.casadelalpiste.com/wp-content/uploads/2011/03/beneficios-
       del-alpiste.jpg" alt="beneficios del alpiste" width="395" height="294"/>
    </a>
</p>
```

Figure 4.40: *alt* attribute and image optimization for the keyword *beneficios del alpiste*.

To summarize the image optimization process, below are a few key points to consider:

- Use short and descriptive filenames for your images. Instead of using "image.jpg" or "pic1.gif", use more descriptive image file names. Whenever possible, use your target keyword within the image file names.
- Use a short and descriptive text in the alt tag attribute; include your target keyword within the text.

4.5.8 Effective use of the "rel=nofollow" attribute

In a HTML anchor tag `<a>`, the `rel` attribute defines the relationship between the current page and the page being linked to by the anchor. Some possible values and their definitions can be seen in Figure 4.41 below.

Value	Description
alternate	An alternate version of the document (i.e. print page, translated or mirror)
stylesheet	An external style sheet for the document
start	The first document in a selection
next	The next document in a selection
prev	The previous document in a selection
contents	A table of contents for the document
index	An index for the document
glossary	A glossary (explanation) of words used in the document
copyright	A document containing copyright information
chapter	A chapter of the document
section	A section of the document
subsection	A subsection of the document
appendix	An appendix for the document
help	A help document
bookmark	A related document
nofollow	"nofollow" is used by Google, to specify that the Google search spider should not follow that link (mostly used for paid links)
license	Link to copyright information for the document
tag	A tag (keyword) for the current document

Figure 4.41: Some possible values for the `rel` attribute from the <a> tag

This attribute is commonly set to a value of "*nofollow*" to signal web spiders not to follow the link. Google states that "setting the value of the 'rel' attribute of a link to 'nofollow' will tell Google that certain links on your site shouldn't be followed or pass your page's reputation to the pages linked to." Google added this attribute to combat comment spam and to keep comment spammers from receiving reputation (or PageRank) from high quality sites.

It's common for sites to have sections where people can leave comments, and before the adoption of the "`rel=nofollow`" attribute, spammers took advantage by commenting on trusted sites with the sole goal of acquiring backlinks to their sites in order to gain higher rankings; similar to the JCP incident that was discussed in Sub-section 2.3.5.

Figure 4.42 shows a possible scenario in which a spammer leaves a comment on a trusted site with the hopes of getting reputation from backlinks.

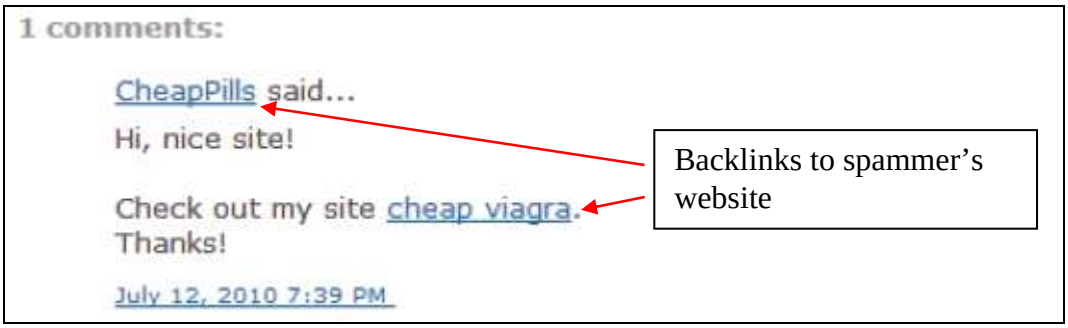


Figure 4.42: Backlinks to spammer's website.

So to combat this form of spam, Google introduced the "`rel=nofollow`" attribute; Figure 4.43 shows the correct use of the `rel` attribute.

```
<a href="http://www.shadyseo.com" rel="nofollow">Comment spammer</a>
```

Figure 4.43: Use of the "*nofollow*" value to combat spam.

Ineffective use of the `rel` attribute and unknowingly "linking to sites that Google considers spammy can affect the reputation of your own site" [19]; this can hurt your rankings. The definition of a "spammy site" varies, but it can be a site that contains commercial offers that the current site is tricked into linking to, even though it has no real relationship to the current site. The spammy site may also have monetary intentions to try to influence the search engine rankings in its favor by leaving spam comments.

Since I used Wordpress (a web-based CMS) for the development of the experimental website and because I disabled all public commenting, I didn't have to configure the `rel` attribute. Most web-based CMS platforms have a feature that automatically sets the value of `rel` to `nofollow` in order to protect your site's reputation, but you can also do this manually by configuring the web page's `robots meta tag` in the code. Figure 4.44 shows an example of `robots meta tag` configuration; this setting sets all the links coming out of the page to `nofollow`, so you can be sure that your page is not passing any reputation to the outgoing links.

```
<html>
<head>
<title>Brandon's Baseball Cards - Buy Cards, Baseball News, Card Prices</title>
<meta name="description=" content="Brandon's Baseball Cards provides a
large selection of vintage and modern baseball cards for sale. We also offer
daily baseball news and events in">
<meta name="robots" content="nofollow">
</head>
<body>
```

Figure 4.44: Example of *robots meta tag* configuration.

To summarize this section, below are key points to consider on the effective use of the "`rel=nofollow`" attribute:

- If your site has a section that allows public commenting, make sure the comments are configured to `rel=nofollow` as a precaution to combat spam.
- "If you're willing to vouch for links added by third parties (e.g. if a commenter is trusted on your site), then there's no need to use nofollow on links" [19], but having links to sites that Google considers spammy can hurt your site's reputation and rankings.

4.5.9 Keyword Placement

How and where you place your target keyword within the web page being optimized is important; the goal is to develop a highly relevant page (for search engines and users) by strategically placing your target keyword in key places of the web page. So when users submit a search query that includes your target keyword and arrive on your page, they'll consider your page to be relevant. To illustrate this point, I will show an example of high-relevancy next.

I performed a Google search for *rare baseball cards for sale;* Figure 4.45 shows a highly optimized page for my search query. You can see how my search query appears in the URL, in the heading and in the text body of the page. Any user who submits the search query *rare baseball cards for sale* and arrives at this page will most likely stay longer and read the entire page. Also, Google has also rewarded this page with the No. 1 position on the search results; you can conclude that this page is optimized for the pervious keyword.

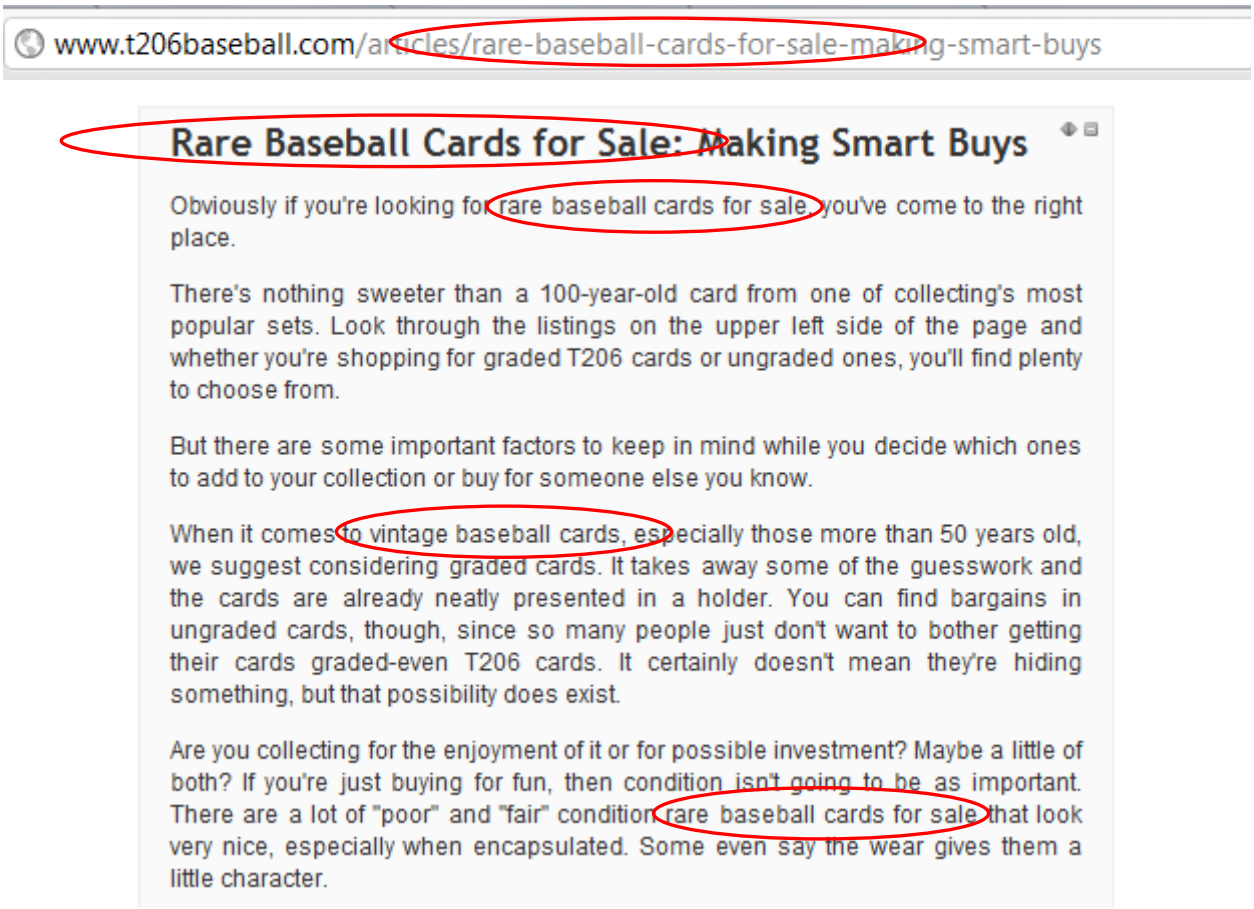


Figure 4.45: Highly optimized page for the keyword *rare baseball cards for sale.*

Figure 4.46a shows the HTML source code of the page; you can also see how the web developer has incorporated the keyword (*rare baseball cards for sale*) into key HTML elements: the *title tag* and the *description meta tag* (seen in Figure 4.46b and Figure 4.46c), which is what Google uses to display on the results page. This is not a coincidence; the page was optimized for the previous keyword.

```
<html xmlns="http://www.w3.org/1999/xhtml">

<head profile="http://gmpg.org/xfn/11">
<meta http-equiv="Content-Type" content="text/html;
charset=UTF-8" />

<title>Rare Baseball Cards for Sale: Making Smart Buys
| T206 Baseball Cards</title>
```

Figure 4.46a: Title tag optimization for *rare baseball cards for sale*

```
<meta name="description" content="Obviously if you're
looking for rare baseball cards for sale, you've come to
the right place. There's nothing sweeter than a 100-
year-old card from one of" />
```

Figure 4.46b: Meta description optimization for *rare baseball cards for sale*.

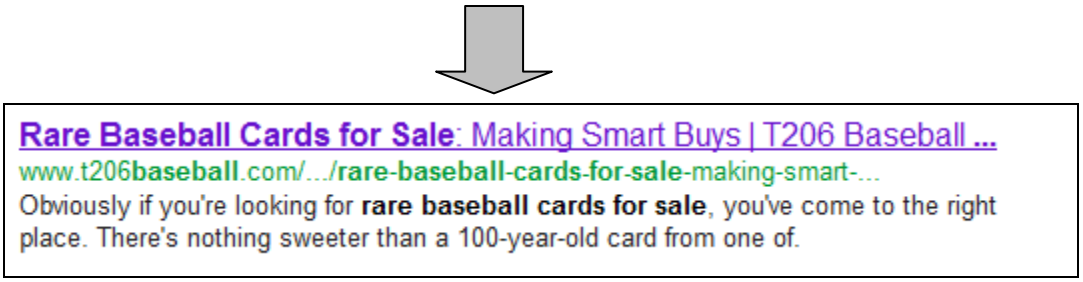
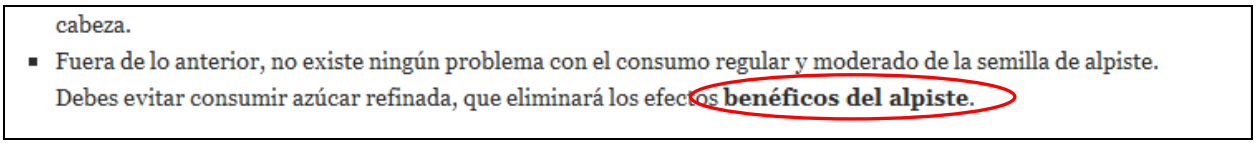


Figure 4.46c: Google search results for the keyword *rare baseball cards for sale*

When placing your target keyword on the page being optimized, it's important not to overuse the keywords; this applies to the visual text of the HTML document and to the HTML source code that's only visible to the search engine. Your page can get penalized if over use of the keywords is detected.

SEO experts use the concept of *keyword density*, which is defined as the number of times your target keyword appears on the web page as a percent of the total number of words. SEO engineers offer a heuristic to maintain a keyword density of 2 – 3%.

Figure 4.47 shows different snapshots of the experimental page that was optimized for the target keyword *beneficios del alpiste*. You can see how the target keyword is included in the URL, in the heading tags, within the body of the web page and to the placements that were implemented in the HTML source code (image and description meta tag).


cabeza.

- Fuera de lo anterior, no existe ningún problema con el consumo regular y moderado de la semilla de alpiste. Debes evitar consumir azúcar refinada, que eliminará los efectos **benéficos del alpiste**.

Figure 4.47: Key page elements optimized for the keyword *beneficios del alpiste*.

To summarize this section, below are guidelines for keyword placement and keyword use. Follow these steps to highlight your important keywords and increase page relevancy for the keywords you're trying to rank for.

- Place keyword in the title tag, description meta tag, image alt attribute and headings (as discussed in the previous sections)
- Place keyword in the first and last paragraph of the page.
- Bold your target keyword at least once on your page.
- Italicize/underline your target keyword at least once on the page.

4.5.10 Sitemaps

It's recommended to create an HTML sitemap for users and a XML sitemap for search engines; their differences and their uses be explained next.

The HTML sitemap "is a simple page on your site that displays the structure of your website, and usually consists of a hierarchical listing of the pages on your site" [19]. This page is visible to all users and contains links to all other pages on your website.

While this sitemap is mainly created for human visitors, search engine crawlers may also use it to achieve "good crawl coverage of all the pages on your site" [19]. In other words, crawlers may visit the sitemap and branch out from there to crawl all the pages contained in your site; this increases the possibility of getting most of your pages indexed.

Figure 4.48 shows an example of a site map from a website that sells vintage cards. You can see that this sitemap provides users with a quick way to access all pages of the site; it's another way to provide a great user experience which Google encourages.

Figure 4.48: HTML *site map* from a website that sells vintage cards.

Figure 4.49 below shows the site map created for the experimental site; these are links to all the pages. It's simple sitemap with only twelve links to the pages.

Figure 4.49: HTML site map from *casadelalpiste.com* (the experimental site).

The second type of sitemap is created using XML; this sitemap makes it easy for search engines "to discover the pages in your site" [19]. It's a way to tell Google what pages exists in your site so that they can get indexed. Google says it won't guarantee that they'll crawl or index all of the URLs, but they'll "use the data in your Sitemap to learn about your site's structure, which will allow us to improve our crawler schedule and do a better job crawling your site in the future" [21].

While having an XML sitemap may not directly affect your search engine rankings, it increases the probability of getting all of your pages crawled and indexed; once this occurs, the ranking algorithm will analyze and determine their relevancy for your target keywords.

In Figure 4.50 you can see a shortened version of the XML site map that I developed for the experimental site. Notice that the location of the XML sitemap needs to be uploaded on the root directory of the website.

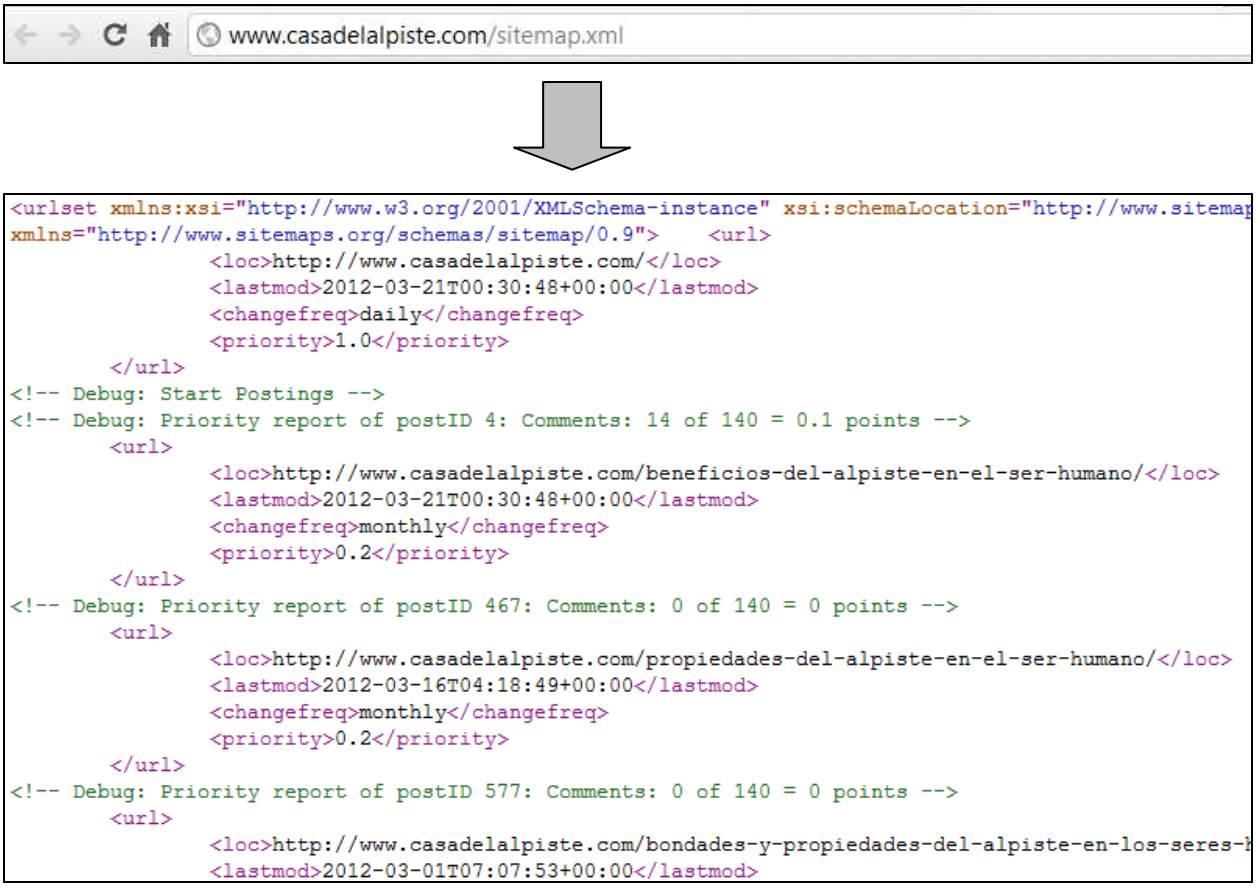


Figure 4.50: XML sitemap for casadelalpiste.com and sample of its XML source code.

In summary, sitemaps serve two purposes:

1. An HTML sitemap will make it easy for users to access all of your content from a single page. This provides a great user experience for the visitors, which in Google's view, this enhances the user experience and eventually can help your site's rankings.
2. With a XML sitemap, the search engine crawler will discover the pages in your site much faster. Once the crawler arrives to the sitemap, it can process all the links and index the

pages. Once this occurs, the search ranking algorithm can determine page relevancy and page rankings for any target keywords you may be optimizing for.

4.6 Off-Page SEO Strategies

This section covers the off-page SEO process, its definition and methods for effective execution. It is structured as follows: Section 4.6.1 discusses important concepts. Section 4.6.2 covers the importance of links and the anchor text. Section 4.6.3 describes link-building and why getting links in a gradual manner is important. Section 4.6.4 discusses the creation of valuable content for getting links. Section 4.6.5 covers article writing as a link-building strategy. Section 4.6.6 covers reaching out to potential partners for link-building. Section 4.6.7 discusses the strategy of getting backlinks from websites that are similar to yours. Section 4.6.8 provides a list of other link-building ideas or methods. Section 4.6.9 provides a list of riskier link-building methods. And finally Section 4.6.10 concludes the off-page SEO Section.

4.6.1 Background

Off-page SEO is primarily focused on acquiring backlinks from reputable sites in a gradual reputation-building process. Any technique that the SEO engineers can create that accomplishes this goal is by definition a candidate off-page SEO technique; it is an open-ended process and is driven by the SEO engineer's individual creativity. In contrast, on-page SEO is a more standardized set of technical procedures that can be studied and applied by anyone.

Off-page SEO is what gets implemented away from site as opposed to on-page SEO. And by definition, the SEO engineer will have less control, making these techniques more difficult to execute. Off-page SEO is more about building relationships with other sites through the development and promotion of attractive web content and reaching out to people who run similar websites in hopes of getting backlinks. This is part of the link building process.

The type of sites that link to your site, how they link to your site, and the way your content is shared on the Web are "all factors that can have a significant impact on your ability to" [19] achieve high rankings on the search engine results page.

For this reason, it's important have a solid on-page SEO implementation in place before embarking on any off-page SEO strategies. The site must be structured correctly by making it easy to navigate, and it must have content relevant to the set of keywords you're optimizing for.

These are a few important points to consider before approaching off-page SEO through link building.

Link building is not something that can or should be achieved over night; it's an on-going process that takes time. As mentioned before, most of the time it's something that's out of your direct control but there are certain tactics that SEO engineers can do to increase the possibilities of having other sites link to yours. Before covering the basics of link-building, there are a few important factors that need to be discussed in order to understand the importance of links Google's ranking algorithm.

As discussed in Section 2.3, the founders of Google published their research paper in 1997 called "*The Anatomy of a Large-Scale Hypertextual Web Search Engine*." In the paper, they presented their "prototype of a large-scale search engine" which eventually evolved into Google's present day search engine.

The Google founders described the importance of hyperlinks and the text within the link. They explained that "the text of links is treated in a special way" [22] in Google's ranking algorithm; they "associate it with the page the link points to. This has several advantages. First, anchors often provide more accurate descriptions of web pages than the pages themselves" [22]. You can see that links and the text that appears in the hyperlink occupied an important role in the development of the search engine and its ranking algorithm; this was at the heart of how its algorithm ranked pages and the importance it placed on links.

It is therefore crucial to remember that links and the text associated with it is very important; this will play a critical role on how you approach the link building process. Also, you will see that link building is more of an art than a science.

4.6.2 Importance of backlinks

As discussed, Google's ranking algorithm places high importance on the anchor text; it's "one of the strongest signals the engines use in rankings. If dozens of links point to a page with the right keywords, that page has a very good probability of ranking well for the targeted phrase that appears in the anchor text" [61].

You can see the effects of this in Figure 4.51; I performed a Google search for the keyword "*click here*." Notice how the page for *Adobe Reader* appeared in the No. 1 position in the search results. The reason Adobe achieves high rankings is because millions of people commonly use the anchor text "*click here*" on their websites and link to the page where *Adobe*

Reader can be downloaded. As a result, the *Adobe Reader* download page has tens of thousands of backlinks; thus, it ranks high for the keyword "*click here*."

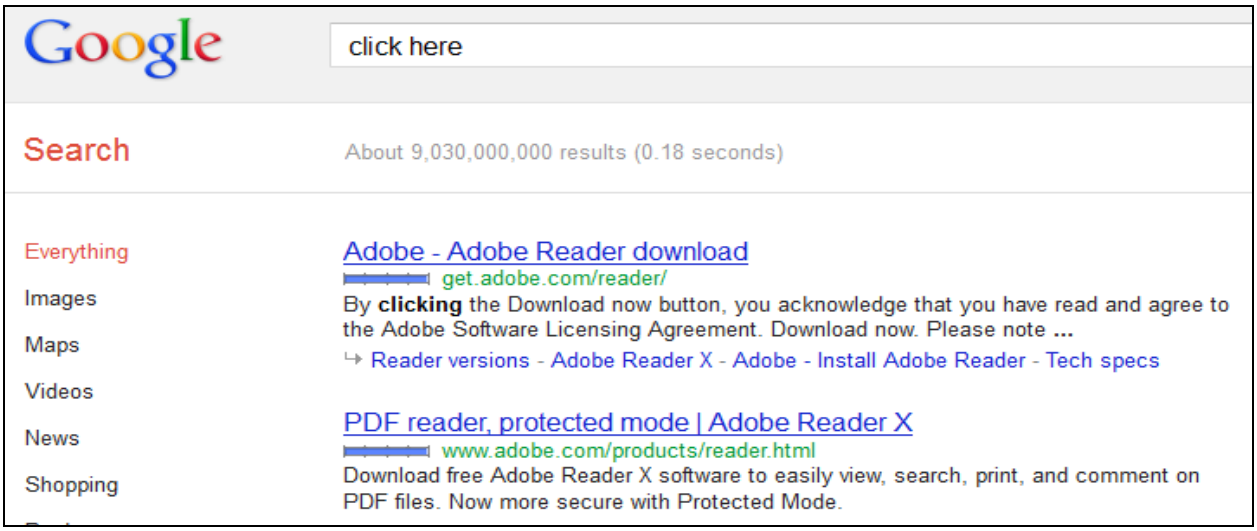


Figure 4.51: Adobe's website shown at No. 1 position for the keyword "*click here*"

So whenever you have the option to get a backlink from a high-authority website, always request keyword-rich anchor text; in other words, make sure that the link that points to your site has your target keyword in the anchor text. You will see an example of this next.

For the experimental website, I created profiles on reputable sites where users can publish content. And since I wanted to achieve high rankings for "*beneficios del alpiste*" (my target keyword), I created short articles about *alpiste* and included my target keyword within the anchor text that pointed to the page I optimized. Figure 4.52 shows the URL where I published the article; you can also see part of the article where I inserted the target keyword "*beneficios del alpiste*" and the HTML code of the backlink.

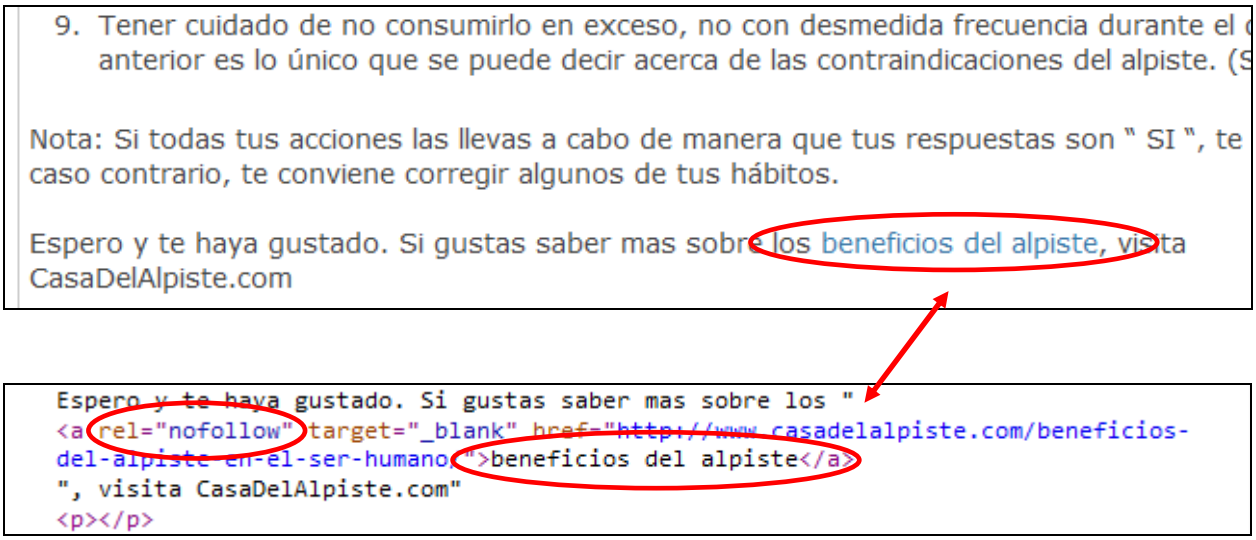


Figure 4.52: Actual backlink to *casadelalpiste.com*.

Notice that this backlink is *rel=nofollow*; as explained in Section 4.5.8, having this attribute in the anchor tag prevented the source site from transferring its *PageRank* to my experimental site. Even though this was the case, I wanted to test it and decided to continue with the creation of the backlink.

In the best case scenario, you're looking for as many links as possible in the format as seen below. Notice the "*rel=follow*" attribute; this is a signal to the search engine to follow the link and the website receiving the link gets PageRank:

```
<a href="http://www.OptimizedPage.com" rel="follow">Target Keyword</a>
```

In the above example, the *target keyword* (or a variation of the keyword) is the keyword that you hope your website achieved high rankings for. When it comes to off-page SEO, there is only so much you can do as far as how other sites link to yours. Therefore, getting links from trusted sites can help increase your site's reputation, and important metric used by Google and other search engines.

4.6.3 Importance of a gradual link-building process

Keep in mind that Google looks to see that backlinks are "gained gradually, as people discover your content through search or other ways and link to it" [19]. This means that the

acquiring backlinks should be natural and steady. Also, it is for this reason that SEO takes a lot of time and patience to implement.

Furthermore, it's an intentional part of Google's strategy for gradual reputation building that it not be quick or overnight. In fact, if a site were to acquire dozens or hundreds of backlinks overnight, Google would almost certainly consider this a red flag (spam) that most likely will get your site penalized.

Many unscrupulous web page developers might try to game the ranking algorithm by spreading many backlinks to their site across other web sites, for example via comment spam. But as its ranking algorithm continues to evolve, Google can easily detect that such links are spam and not organically developed over time from different users.

Such techniques are not a serious enough offense for Google to delist the site. In an extreme example, an adversary could create many backlinks to a competitor's site from undesirable secondary sites in an attempt to hurt the reputation (or ranking) of the competitor's site. Fighting such techniques requires tedious configuration to block links from specific IP addresses, but the net result may be that the site is still ranked lower than it would have been otherwise.

Having covered the importance of links, the next sections cover link-building strategies for effective off-page SEO, which can help in your rankings when implemented correctly.

4.6.4 Importance of quality content

As you create web pages, it's critical to understand that in order to get backlinks, you need to publish content that make people want to share, read and link to. For this to occur, your content has to provide value, be interesting and/or be unique. Furthermore, don't expect to develop pages of content and anticipate users to find it without any work on your part; you need to publish and promote the content where potential users will find it.

Google says that "effectively promoting your new content will lead to faster discovery by those who are interested in the same subject" [19]. Figure 4.53 provides a general idea of effective ways to promote your content. You can use a combination of online and offline promotion for your site as a way to get quality links to your site, which can help increase your site's reputation and rankings. Offline promotions can be implemented through traditional direct marketing approaches such as newsletters, posters and mailings for example.

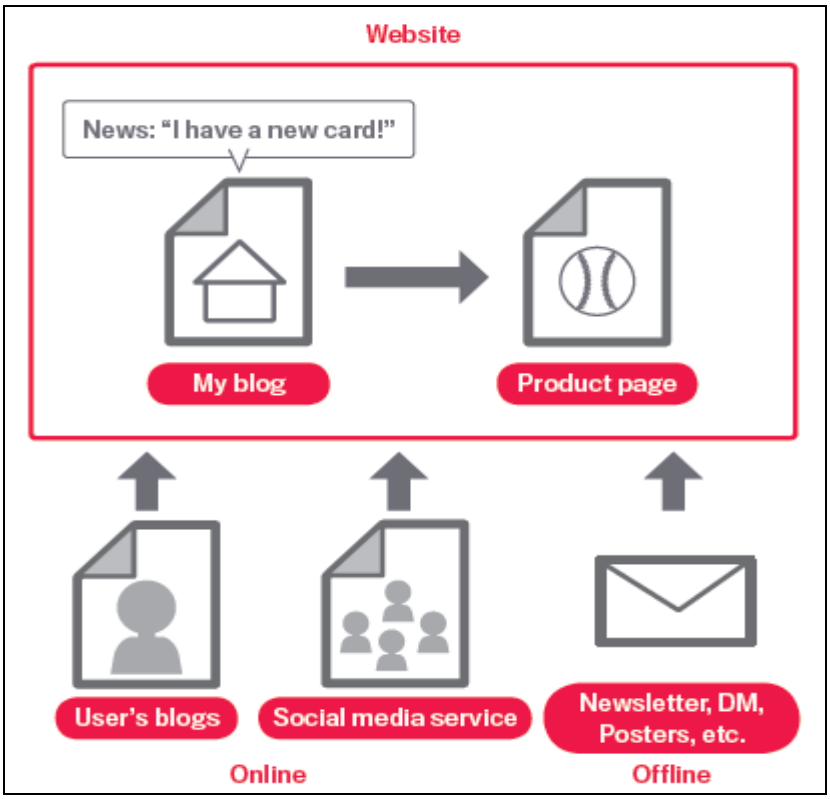


Figure 4.53: Effective ways to promote website content.

4.6.5 Writing articles to establish domain authority

Another strategy that can help your site get backlinks is by writing useful articles that provide value. In addition to providing valuable content on your website, you can also write valuable articles and submit them to trusted sites such as About.com, Wikipedia.org or NewYorkTimes.com. Getting a backlink from high-authority site such as these will greatly increase your website's reputation and help you achieve higher rankings.

Another approach is to submit articles to article directories. These are online directories that publish articles on any topic imaginable; from gardening ideas to dating tips. This approach works really well. However, you have to find reputable article directories with high *PageRank*. Submitting articles to these trusted sites is an effective way to get quality backlinks, which can help increase your site's exposure and search rankings.

For example, Figure 4.54 shows an article titled "*How to Install and Untangle Linux Virtualbox*" that was submitted to www.ezinearticles.com, an article directory, which has a PageRank of 6 (at the time of this of this writing). Authors that submit to these directories use the last paragraph to provide a link back to their site. This can also be seen in Figure 4.54.

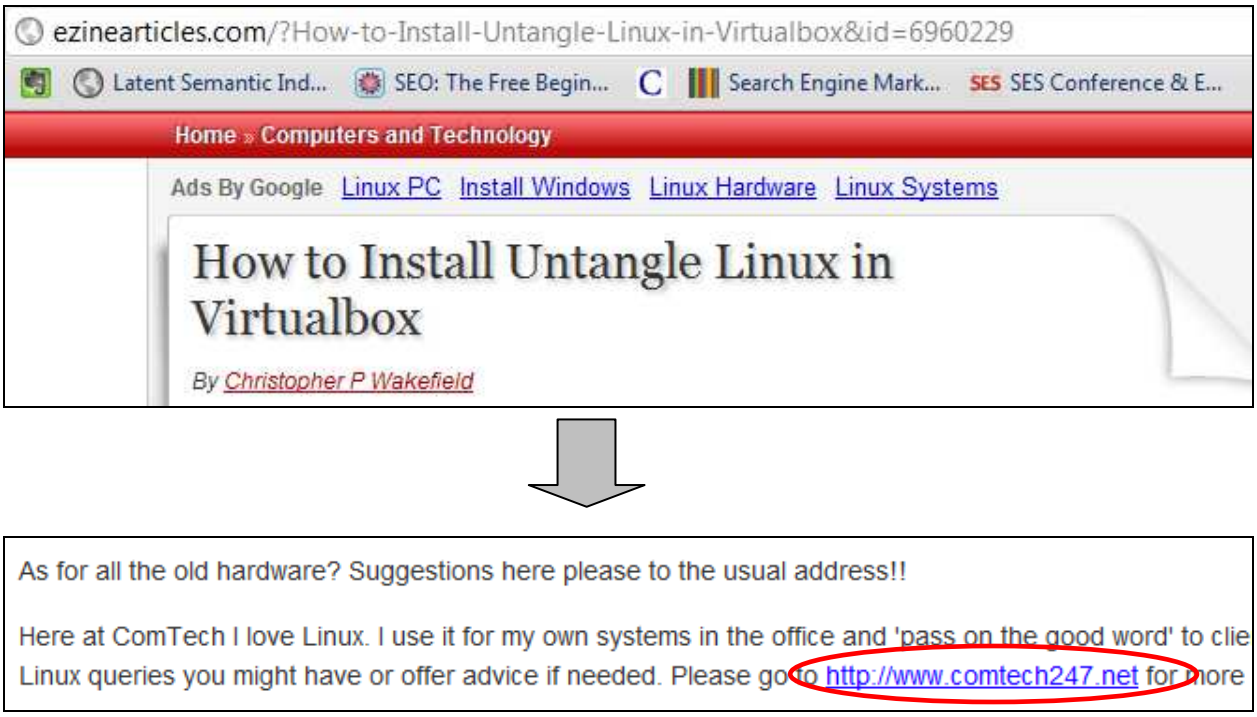


Figure 4.54: Actual backlink from an article directory to the author's site.

4.6.6 Personal networking to establish a reputation

Due to the size of the Web, there is a high probability that there will be sites that have content related to yours. Google suggests that you reach out to those in your site's community "that cover topic areas similar to yours. Opening up communication with these sites is usually beneficial" [19]. Contacting sites that are related to what your site is about is a great way to network, promote and increase your site's exposure.

For example, if you own an online baseball card store, you may contact a sporting goods store to see if it is interested in what your site offers. The site owner may be willing to link to your site and vice versa. The goal is to get as many backlinks as possible, but keep in mind that search engines place more importance to links from popular, trustworthy and authoritative sites. So always make sure that you approach the correct websites for backlinks.

4.6.7 Finding your website's natural affinity group

When possible, always get backlinks from sites that are similar or that are as topic-specific as your site; links from off-topic sites do not count as much as links from sites that have

content related to yours. For example, a site about weight training would benefit more by getting backlinks from a health-related site than links from a grocery store website.

4.6.8 Other link-building methods

The following list provides other link-building methods that can be strategically used in order to get backlinks:

- **Write guest articles** – Just like submitting to article directories, you can write articles and get them published on reputable sites similar to yours.
- **Contact business partners** – If you're part of a network or business group, get in contact with key people.
- **Do charity work and ask for a link in return** – If you do work for a non-profit organization, you may ask for a link to your site.
- **Ask friends and family** – This can help you if family members or friends have sites that have similar content as your site.
- **Participate in the comments of authoritative websites** – Leaving useful comments in authority sites (such as the New York Times) can help your SEO.
- **Attend industry events** – going to industry events are a good way to network and meet potential business partners to exchange links.

4.6.9 Riskier link-building methods

There are other, riskier methods, of getting backlinks. The following are some methods used by many SEOs but these are all strongly discouraged:

- **Buying links from link brokers** – There exist link brokers that sell or lease space on highly trafficked sites. This method can help in your SEO and in getting traffic, but this method is also not recommended. Google has advised all websites to avoid "purchasing links from another site with the aim of getting PageRank instead of traffic" [19].
- **Buy natural links** – These are links that seem as if they were organically acquired, but instead were paid for. If you have connections with someone at a high authority site, you can purchase a link (or links) from them. While this may help in your SEO and rankings, this practice is not recommended. Your site risks getting penalized for such procedures.

4.6.10 Summary

As you can see, Google places high importance on links, the type of sites that link to your site, and how they link to your site. Google mentions that the anchor text gets associated "with the page the link points to" [22]; so whenever possible, always try to get backlinks with your target keyword in the anchor text. This will increase page relevancy and it can help your site achieve higher rankings for your target keyword.

Furthermore, off-page SEO is mostly about link-building and it's an on-going process that takes time, patience and creativity. Link-building can be considered more of an art than a science. If you create content that's worthy to be shared, you will ultimately get lots of backlinks the natural way and in a gradual manner, just as Google recommends. Valuable and unique content is what attracts users and gets shared among the Web.

Finally, you can see that off-page SEO is less clearly defined than on-page SEO. It is not easy to create a standardized list of off-page SEO techniques. The goal of such techniques is to acquire backlinks from highly reputable sites. But the state of the practice today is that the SEO engineer must use his or her creativity and technical knowledge in order to be successful here.

Chapter 5 - Result Analysis

This chapter discusses and analyzes the results of the SEO strategies implemented. As explained in Section 4.1, early March 2011 marked the beginning of the development of *casadelalpiste.com*, the experimental site. Once the development had been completed, articles and pages were uploaded to the site. One of the pages was selected to become the experimental page where selected SEO techniques were implemented. This started the optimization process which continued until September 2011.

The next sections will analyze the metrics and the effectiveness of the application of the SEO strategies on the experimental site. The rest of the Chapter is structured as follows: Section 5.1 discusses "*Number of Visitors*", Section 5.2 discusses "*Pageviews*" and Section 5.3 covers "*First Page of Google*."

5.1 Number of Visitors

Number of visitors calculates the total number of users who visit your website and was used as a benchmark to determine the effectiveness of the SEO implementation. I decided to use *number of visitors* as a metric (to track and measure) because it tells you whether the optimized site is being found online. In the end, that's the goal of most SEO engineers; to increase a website's visibility by sending as many users to the website as possible.

Figure 5.1 shows a weekly trend of the number of visitors to the experimental website from March 1, 2011 thru April 30, 2011. Notice how the number of visits at the beginning of March was low and then started to increase towards the end of March. This increase in number of visits was most likely caused by the work I was performing on the site (implementing on-page SEO, uploading content, etc…) and Google Analytics was tracking and including my visits to the site.

So in order to prevent Google Analytics from counting my visits to the site, I configured the analytic settings so that my IP address was excluded from the tracking analysis. This configuration excluded my visits, and it would provide a true picture of the actual visits other than mine.

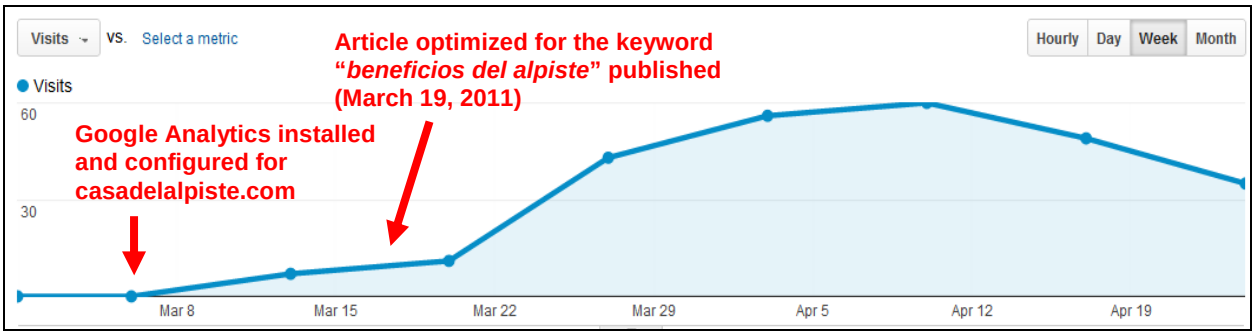
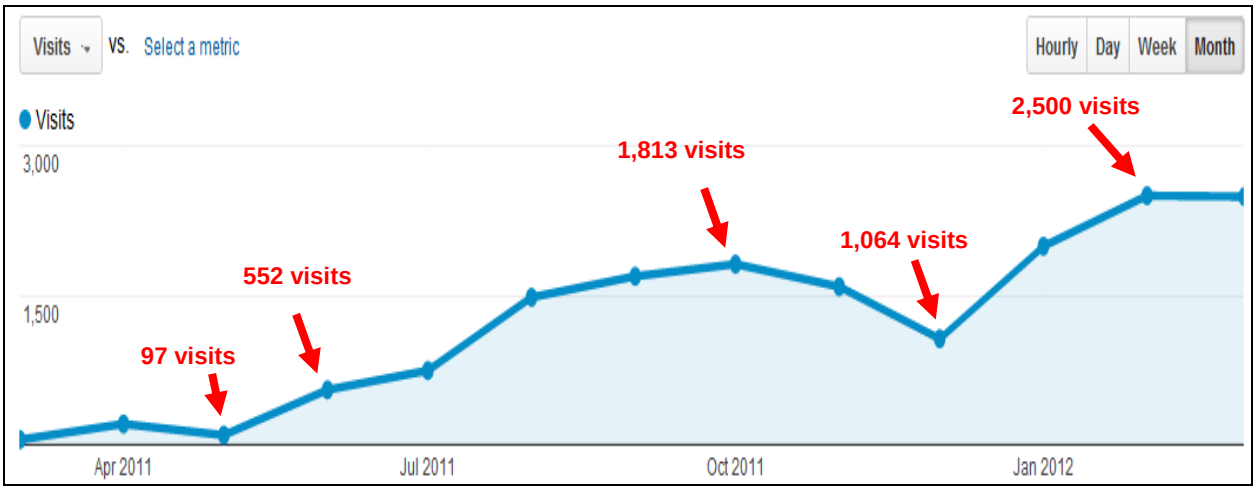


Figure 5.1: Weekly number of visitors to *casadelalpiste.com* (March 1, 2011 – April 30, 2011).

Three months passed before I started to notice an increase of visitors to the site. The number of visitors started to climb in June 2011; Figure 5.2 shows the monthly trend. You can see that in June 2011 the site received 552 monthly visitors, a 469% increase from the previous month.

From June 2011 thru October 2011, the total monthly visitors increased month to month; in June there were 552 visitors and by October 2011 this number had increased to 1813 monthly visitors. Then in November 2011 and December 2011, the number of monthly visits decreased to 1,587 and 1,064, respectively. I'm guessing this decrease in number of visitors was the result of the holiday season, because in January 2012 the number of monthly visitors increased to 1,992 and in February 2012 continued to increase to 2,500. The monthly trend in number of visitors is seen in Figure 5.2.

This surge in traffic at the beginning of the year coincides with the research performed in Google Trends (Section 4.3); for some reason people start searching for *alpiste* related keywords at the beginning of the year. This may also have to do with New Year's resolutions. People are looking for ways to lose weight and stay healthy, which are some of the benefits that *alpiste* offers.



Visits:	52	209	97	552	745	1,480	1,691	1,813	1,587	1,064	1,992	2,500	2,500
Month:	Mar-11	Apr-11	May-11	Jun-11	Jul-11	Aug-11	Sep-11	Oct-11	Nov-11	Dec-11	Jan-12	Feb-12	Mar-12
% Change:		302%	-54%	469%	35%	99%	14%	7%	-12%	-33%	87%	26%	0%

Figure 5.2: Monthly number of visitors to *casadelalpiste.com* (Mar '11 – Mar '12).

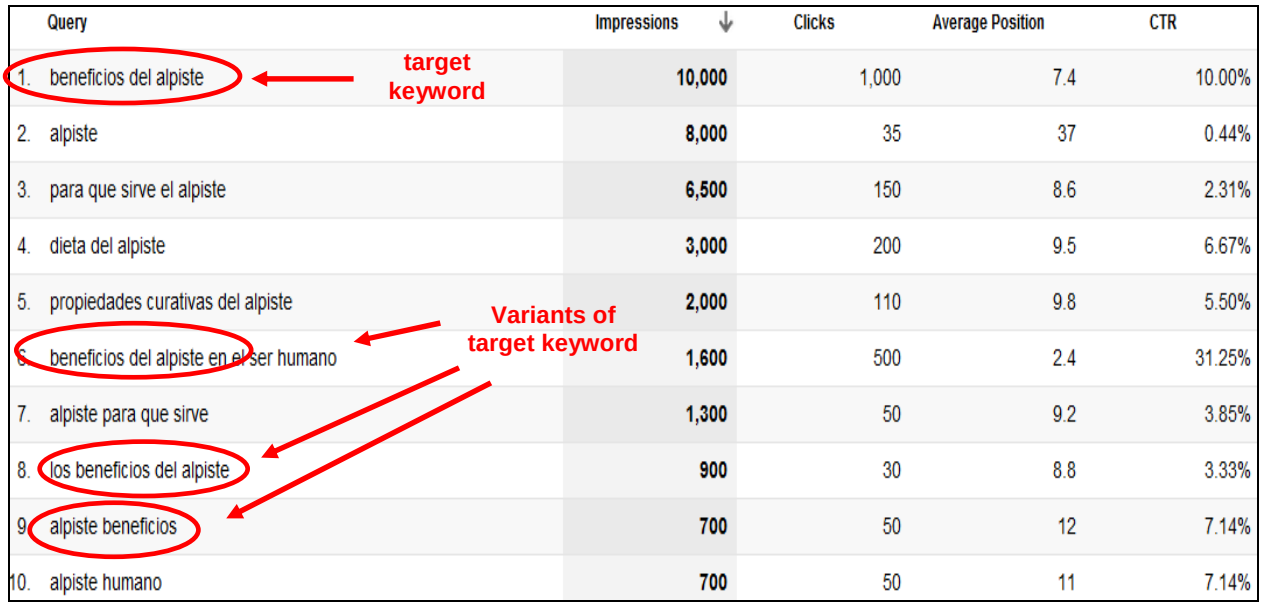
Figure 5.2 also shows the % change in the number of visits that takes place month-to-month. The biggest % change occurred in June 2011 with a 469% surge in the number of visits compared to May 2011. The only decrease in visitors occurred in November and December.

Furthermore, I noticed that as I continued to add more pages of alpiste-related content to the experimental site, the number of visitors continued to increase. I did not create pages and upload them all at once. Once I had published (on March 19, 2011) the page to be optimized for my target keyword (*beneficios del alpiste*), I uploaded about one page of content every other week.

As stated before, *beneficios del alpiste* was my target keyword and the focus of my SEO techniques. And not surprisingly, this keyword was responsible for bringing the most visitors to the website. Figure 5.3 shows the top ten keywords (by number of impressions) that sent the most users to *casadelalpiste.com*. This data was taken from Google Analytics and it provides an additional four key metrics; the definitions were taken from Google's SEO Reports help page [25]:

- **Impressions**: the number of times any URL from your site appeared in search results viewed by a user.
- **Clicks**: the number of clicks on your website URLs from a Google Search results page.

- **Average Position**: the average ranking of your website URLs for the query or queries. For example, if your site's URL appeared at position 3 for one query and position 7 for another query, the average position would be 5 ((3+7)/2).
- **CTR (click-thru-rate)**: calculated as Clicks / Impressions * 100



	Query	Impressions	Clicks	Average Position	CTR
1.	beneficios del alpiste	10,000	1,000	7.4	10.00%
2.	alpiste	8,000	35	37	0.44%
3.	para que sirve el alpiste	6,500	150	8.6	2.31%
4.	dieta del alpiste	3,000	200	9.5	6.67%
5.	propiedades curativas del alpiste	2,000	110	9.8	5.50%
6.	beneficios del alpiste en el ser humano	1,600	500	2.4	31.25%
7.	alpiste para que sirve	1,300	50	9.2	3.85%
8.	los beneficios del alpiste	900	30	8.8	3.33%
9.	alpiste beneficios	700	50	12	7.14%
10.	alpiste humano	700	50	11	7.14%

Figure 5.3: Top 10 keywords receiving the most impressions (Dec '11 – Mar '12).

You can see from Figure 5.3 above that the main target keyword is getting the most impressions, which is also responsible for receiving most of the clicks that eventually become visits to the site. Furthermore, notice the other search queries (variants to the target keyword), which taken together, are also responsible for sending a high number of users to the site. Look at the search query "*beneficios del alpiste en el ser humano*", which translates to "alpiste benefits in the human body"; this *long-tail keyword* has the highest click-thru-rate (CTR) with 31.25%. This is not surprising because the average position of one of my URLs is at No 2.4 in Google's search results page.

Cleary, the experimental site is getting most of the visitors from the target keyword; but as shown, there are other similar and related keywords that are also sending users to the site. This is great data for SEO engineers to uncover because they can use this information to create additional pages and optimize them using these newly-uncovered keywords. Google is directly telling you what keywords are already sending traffic to your site. You can now take this data and continue optimizing, which can lead to more visitors to the site.

Figure 5.4 shows the top 10 queries by total number of clicks from December 2011 thru March 2012. Notice how the main keyword and three of its variants are responsible for about 53% of all the top 10 clicks (visits) to the site.

	Query	Clicks	Impressions
1.	beneficios del alpiste	**1,000**	20.00%
2.	beneficios del alpiste en el ser humano	**500**	3.20%
3.	dieta del alpiste	**200**	6.00%
4.	para que sirve el alpiste	**150**	13.00%
5.	propiedades curativas del alpiste	**110**	4.00%
6.	beneficios del alpiste en los humanos	**60**	0.64%
7.	usos del alpiste	**60**	0.64%
8.	alpiste beneficios	**50**	1.40%
9.	alpiste humano	**50**	1.40%
10.	alpiste para que sirve	**50**	2.60%

Figure 5.4: Top 10 keywords receiving the most clicks (Dec '11 – Mar '12).

I continued to analyze further within Google Analytics to see what pages (URLs) were ranking the highest in Google's search results page; Figure 5.5 shows the top 10 pages receiving the most clicks. Not surprisingly, the page that I selected to apply the SEO strategies to is getting the highest number of impressions and the most clicks, which leads to more visitors.

	Landing Page	Impressions	Clicks
1.	http://www.casadelalpiste.com/beneficios-del-alpiste-en-el-ser-humano/	27,000	3,000
2.	http://www.casadelalpiste.com/para-que-sirve-el-alpiste-verdades-y-mitos/	12,000	500
3.	http://www.casadelalpiste.com/	4,500	400
4.	http://www.casadelalpiste.com/el-abc-de-la-dieta-del-alpiste-para-bajar-de-peso/	5,500	320
5.	http://www.casadelalpiste.com/propiedades-curativas-del-alpiste/	3,500	250
6.	http://www.casadelalpiste.com/5-usos-simples-del-alpiste/	1,000	150
7.	http://www.casadelalpiste.com/8-beneficios-del-alpiste-para-tu-salud/	2,000	150
8.	http://www.casadelalpiste.com/el-alpiste-para-consumo-humano/	900	90
9.	http://www.casadelalpiste.com/objeciones-sobre-el-alpiste/	600	90
10.	http://www.casadelalpiste.com/propiedades-del-alpiste-en-dos-minutos/	12,000	90

Figure 5.5: Top 10 URLs receiving the most clicks or visits (Dec ’11 – Mar ‘12).

I wanted to see what countries were sending the most visitors to the experimental site; Figure 5.6 shows the top ten countries that are sending the most visitors to *casadelalpiste.com*. As expected, the majority of the visitors came from places where Spanish is the official language. And it was interesting to see that the No. 2 country sending the most users was the United States.

	Country/Territory	Visits
1.	Mexico	3,453
2.	United States	2,792
3.	Colombia	1,764
4.	Spain	1,129
5.	Venezuela	1,093
6.	Peru	1,031
7.	Argentina	948
8.	Chile	642
9.	Ecuador	636
10.	Costa Rica	378

Figure 5.6: Top 10 countries/territories sending the most visitors (Mar ’11 – Mar ‘12).

To summarize this Section, you can see that there has been a continuous increase in the number of visitors since the website was first developed and when the SEO strategies were implemented. Looking at Figure 5.7 below, you can see the monthly stats in the number of visits. With the exception of November 2011 and December 2011, there has been an increase month-to-month.

Keep in mind that the experimental site was optimized for only one keyword (*beneficios del alpiste*), which is responsible for most of the visitors. Going through the same keyword research process that was covered in Section 4.3, I can uncover additional keywords for the application of SEO. And high rankings can also be achieved for other pages and more visitors can be sent to the site.

Visits:	52	209	97	552	745	1,480	1,691	1,813	1,587	1,064	1,992	2,500	2,500
Month:	Mar-11	Apr-11	May-11	Jun-11	Jul-11	Aug-11	Sep-11	Oct-11	Nov-11	Dec-11	Jan-12	Feb-12	Mar-12
% Change:		302%	-54%	469%	35%	99%	14%	7%	-12%	-33%	87%	26%	0%

Figure 5.7: Monthly number of visitors to *casadelalpiste.com* (Mar '11 – Mar '12).

5.2 Pageviews

The second metric to be discussed is *Pageviews*; according to Google, it's an "instance of a page being loaded by a browser. The Pageviews metric is the total number of pages viewed; repeated views of a single page are also counted" [26]. Anytime a page is served by the web server, it is counted as a *pageview*. I decided to include this metric because it's a strong indicator of good user experience, a key factor that Google takes into account in its search ranking algorithm. This may be another reason why Google Analytics tracks it.

This is what Google says about user experience: it's the "experience gained by a user through using products, services, etc. Emphasis is placed on providing an experience truly sought after by the user, such as 'enjoyment,' 'convenience' and 'comfort'"[19]. This includes making the site easy to navigate, providing a sitemap to make the pages easier to find, and creating unique and valuable content.

For example, when a user enters a search query into Google, he or she will have the option of clicking on any of the top search results. Once a click takes place, the user will be taken to a page that's relevant to the keyword used. If the page does not meet the user's expectations (i.e. it's not relevant), the following two scenarios can occur:

1. User will click the back button of the browser and look for another relevant to link to click, or
2. User will try a different keyword.

When any of the previous two scenarios happen, it can mean that the resulting page is not relevant to the user's search query, so the number of *Pageviews* will be low because the user will simply leave the site. In the case that the user does visit a relevant page and finds what he or she is looking for, there is a high probability that the user will stay on the page longer and continue clicking; thus, there will be a larger *Pageviews* value.

Figure 5.8 shows a weekly trend (March 2011 thru June 2011) of the number of *Pageviews* for the experimental site. Notice that throughout the month of March, the weekly *Pageviews* value remains relatively low, then it increases for most of April 2011. As stated before, the increased number of *Pageviews* were most likely due to my web development activities (page creation, on-page SEO and web development). But in late April Google Analytics was configured to exclude my IP from the tracking analysis.

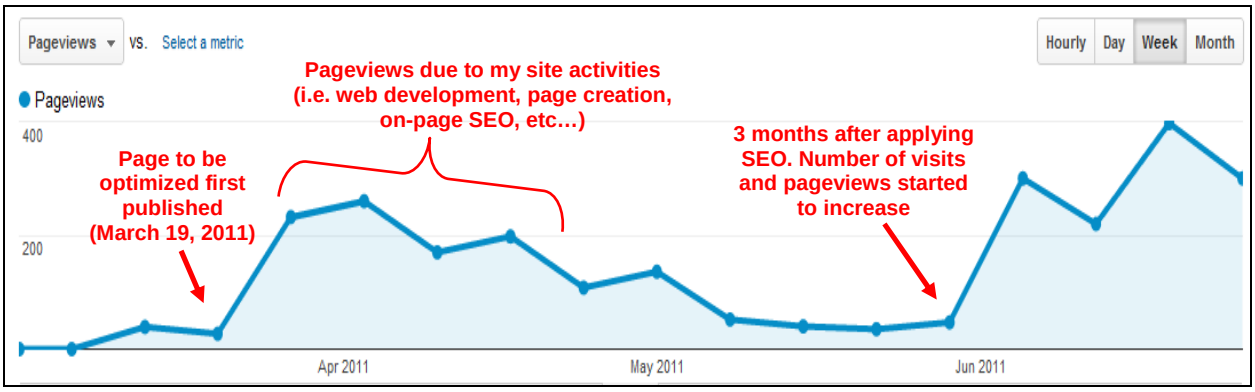
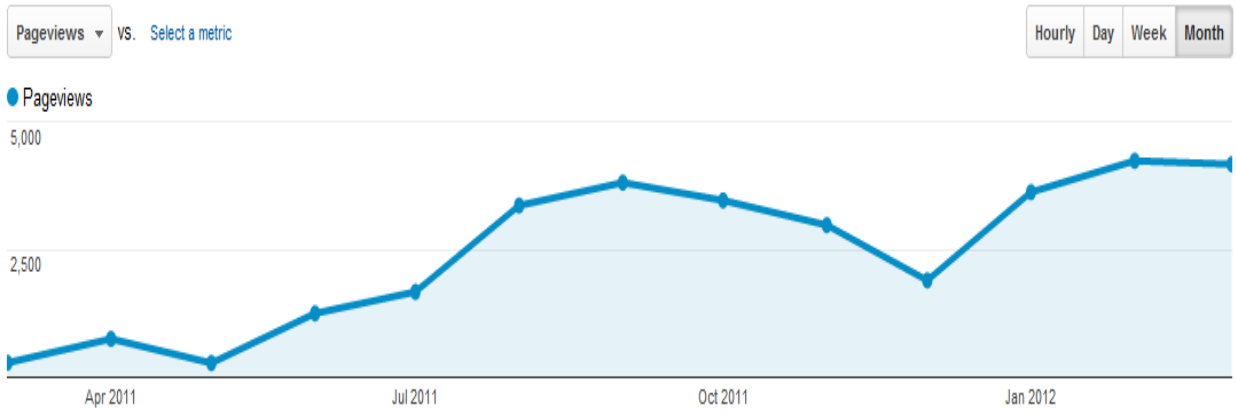


Figure 5.8: Weekly trend of number of pageviews for casadelalpiste.com (Mar '11 – Jun '11).

To analyze the *Pageviews* metric further, Figure 5.9 shows the monthly trend of the number of *Pageviews* for *casadelalpiste.com*. Similar to what occurred to the number of visits, the number of *Pageviews* had its biggest percent increase in June 2011 with a 349% increase from the previous month, 278 versus 1,249. Then, the number of *Pageviews* declined in October 2011 thru December 2011. Once the New Year begun, the number of *Pageviews* started to increase again.

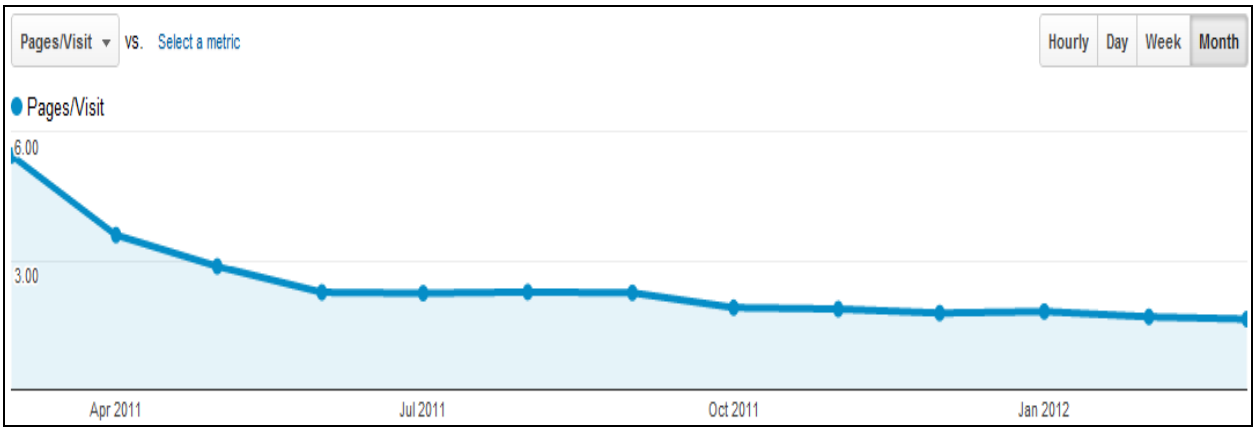


Pageviews:	283	751	278	1,249	1,670	3,359	3,807	3,459	2,976	1,895	3,621	4,233	4,166
Month:	Mar-11	Apr-11	May-11	Jun-11	Jul-11	Aug-11	Sep-11	Oct-11	Nov-11	Dec-11	Jan-12	Feb-12	Mar-12
% Change:		165%	-63%	349%	34%	101%	13%	-9%	-14%	-36%	91%	17%	-2%

Figure 5.9: Monthly trend of the number of *Pageviews* for *casadelalpiste.com*.

Figure 5.10 shows a chart with the two metrics previously discussed (*Pageviews* and *Visitors*) and the addition of another derived metric: "*Pages/Visit*"; this is the average number of pages viewed during a visit to the site. You can see that in March 2011 and April 2011 the *Pages/Visit* was at 5.4 and 3.6, respectively; the highest from any other months. Again, this was most likely due to my web development activities, and so these two months should be discarded from the tracking analysis.

As you look further into the months, from May 2011 thru March 2012 the average is close to 2.0 *Pages/Visit*. This means users are viewing (on average) only two pages from the experimental site. Obviously, I'm not happy with this low number; I was hoping for a higher *Page/Visit* value.



Pageviews:	283	751	278	1,249	1,670	3,359	3,807	3,459	2,976	1,895	3,621	4,233	4,166
Visits:	52	209	97	552	745	1,480	1,691	1,813	1,587	1,064	1,992	2,500	2,500
Month:	Mar-11	Apr-11	May-11	Jun-11	Jul-11	Aug-11	Sep-11	Oct-11	Nov-11	Dec-11	Jan-12	Feb-12	Mar-12
Pages/Visit:	5.4	3.6	2.9	2.3	2.2	2.3	2.3	1.9	1.9	1.8	1.8	1.7	1.7

Figure 5.10: Average number of *Pages/Visit* to *casadelalpiste.com*.

Again, the data from Figure 5.10 above is not a very good sign for a few reasons. This data is telling me that most of my visitors are visiting (on average) only two pages every time they visit. The number of *Visitors* has been increasing, as well as the number of *Pageviews*, but *Pages/Visit* has remained relatively the same (close to 2.0) since May 2011. This also means that I may need to optimize other pages of the site, making the site more engaging so that visitors stay longer and view more pages.

In summary, you can see there was a significant increase in the number of *Pageviews* since the development of the experimental site, followed by the application of the SEO strategies. Figure 5.11 below shows the *Pageviews* monthly stats from March 2011 thru March 2012. In mid March the on-page SEO was implemented; notice that three months later (June 2011) there was a significant increase in *Pageviews* and it continued to increase until October 2011. There was a slight decrease in October, November and December of 2011 and then it early in 2012, the *Pageviews* started to increase again.

Pageviews:	283	751	278	1,249	1,670	3,359	3,807	3,459	2,976	1,895	3,621	4,233	4,166
Month:	Mar-11	Apr-11	May-11	Jun-11	Jul-11	Aug-11	Sep-11	Oct-11	Nov-11	Dec-11	Jan-12	Feb-12	Mar-12
% Change:		165%	-63%	349%	34%	101%	13%	-9%	-14%	-36%	91%	17%	-2%

Figure 5.11: Monthly number of *Pageviews* for *casadelalpiste.com* (March '11 – March '12).

5.3 Reaching the First Page of Google

From the beginning of my research and development of the experimental site, the keyword selected for this research was *beneficios del alpiste*. And like any other SEO engineer, my goal was to achieve high rankings for my target keyword, preferably in the top 5 results on the first page of Google. With this in mind, this was also my goal from the beginning of this project, to get the experimental page to the first page of Google whenever a user typed the search query: *beneficios del alpiste*. Did I achieve my goal? I'll explain next.

As discussed in Section 4.3.3, my target keyword gets (on average) about 9,900 monthly searches; this is according to Google's Keyword Search Tool. Figure 5.12 shows the monthly stats this search query.

Keyword	Competition	Global Monthly Searches	Local Monthly Searches	Approximate CPC	Local Search Trends
[beneficios del **alpiste**]	Low	9,900	9,900	$0.33	

Figure 5.12: Monthly volume stats for the target keyword *beneficios del alpiste*.

Moreover, there are roughly 106,000 indexed pages that Google considers to be relevant to the keyword: *beneficios del alpiste*. This means that my page is competing against this number of pages. You can see this in Figure 5.13a below.

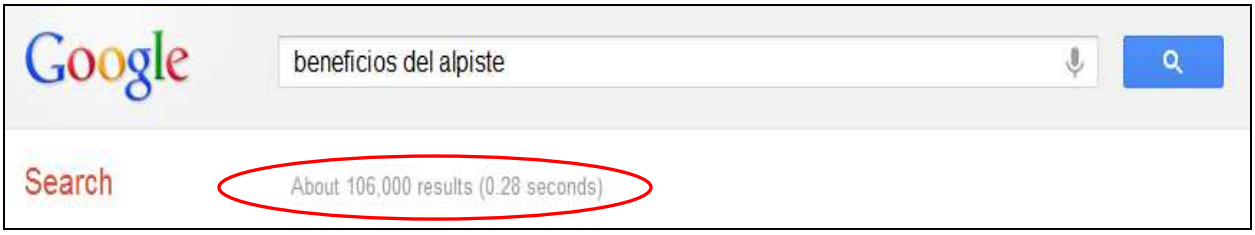


Figure 5.13a: Number of pages indexed for *beneficios del alpiste*.

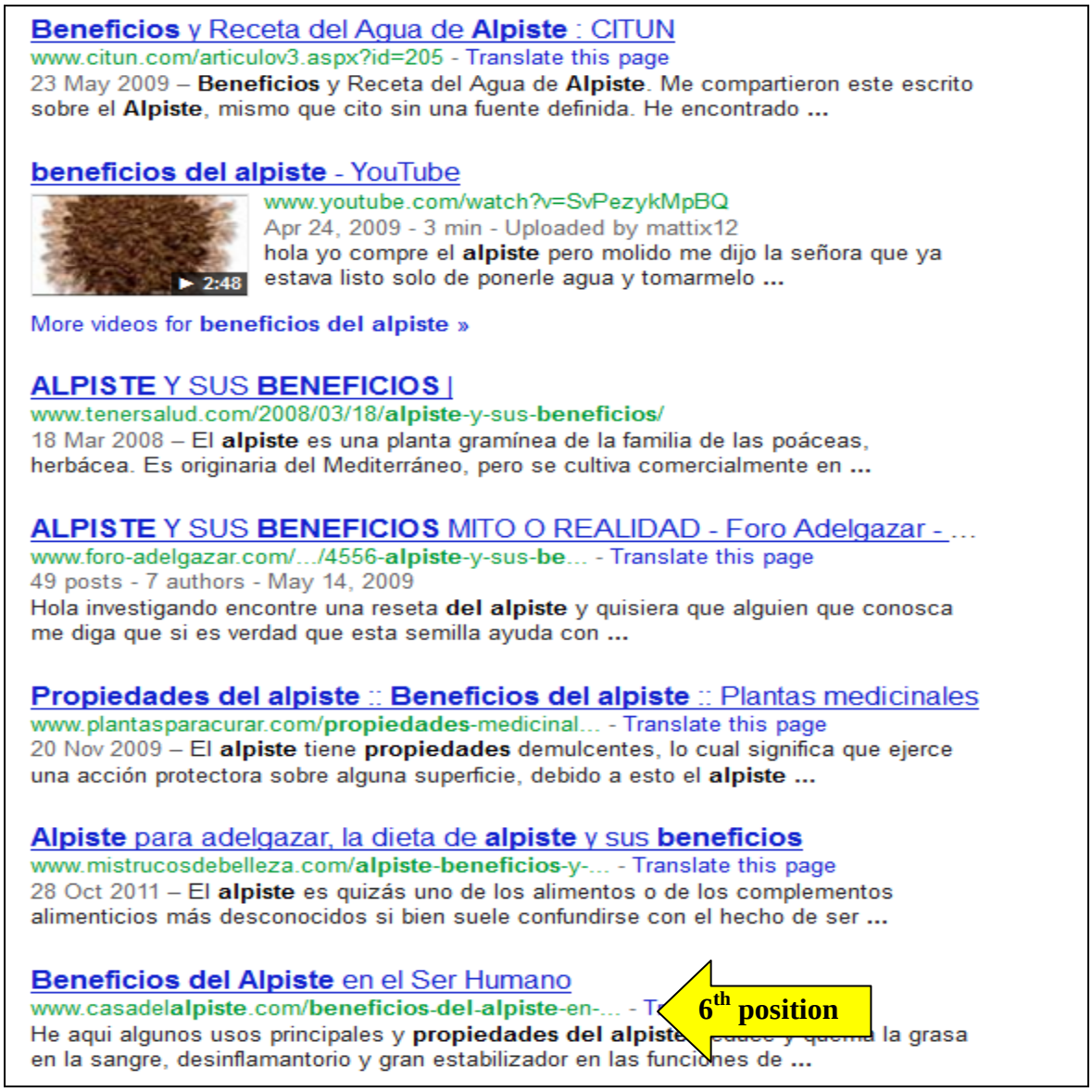


Figure 5.13b: casadelalpiste.com currently at position No. 6 for *beneficios del alpiste*.

I ran my target keyword (*beneficios del alpiste*) through Google and the optimized page appeared in position No. 6 of the search results. You can see this in Figure 5.13b above. While I managed to get the experimental page to the first page of Google, I'm still not in the top 5 results for the selected keyword. Clearly, the SEO strategies implemented had a positive effect on the high rankings. Could I have done more to improve the page's rankings? The answer is yes. Spending more time focusing on link-building to get high-quality backlinks will definitely help reach the coveted No. 1 position.

Keep in mind that the previous ranking results are not static. As discussed in Section 2.3, there are many factors that Google's algorithm considers when ranking pages. And with the constant changes to its ranking algorithm, website rankings also fluctuate. But if SEO gets implemented constantly, higher rankings can be maintained.

Chapter 6 - Conclusion

This chapter presents the conclusion for this thesis and it is structured as follows: Section 6.1 discusses key ideas and important terms covered in this research paper. Section 6.2 covers a summary of the results and finally section 6.3 provides directions for future research in the subject of SEO.

6.1 Summary

As the Internet continues to grow, the availability of information will also continue to increase dramatically. There needs to be order if people must find what they're looking for online. For this reason, search engines will continue to occupy a critical and "a prominent position in the online world" [2]. They will continue to be part of the everyday lives of most people and will only become more important and necessary so people can find information more easily. Below are few key statistics from recent studies [3, 48]:

- More than 80% of first visits to a website come from search engines
- More than 76% of worldwide searchers use Google
- 84% of Google users don't go past the second page of the search results
- "January 2002, 52% of all Americans used search engines. In February 2012 that figure grew to 73% of all Americans. On any given day in early 2012, more than half of adults using the internet use a search engine (59%)"

The previous statistics clearly show the importance of search engines for sending people to websites. These studies show how prevalent search engines have become in the lives of people by helping them find information online. Furthermore, they also show how getting to the first page of Google is critical to a website's visibility by continually attracting visitors to its pages. As a result, sites must do what is required to stand out from the competition in order to stay in business; this means to start paying attention to key processes such as SEO in order to increase their site visibility.

SEO methods will continue to grow in importance and research in this area is destined to grow as the Internet continues its growth. As discussed, Google has become an essential link in the process by which customers and business locate each other. And in many ways, SEO can help bridge that gap. As a result, SEO has gained much popularity in recent years by giving SEO

engineers a strategic process for editing web page “content and HTML to boost its relevance with the specific keywords” [3] which can lead to higher rankings and more visitors. The SEO strategies implemented in this paper will be presented next.

6.2 SEO Process

In this paper natural search engine ranking factors were discussed, implemented and analyzed. In addition, their effectiveness on Google’s search engine ranking algorithm were tracked and measured. The following list provides a high-level overview of the SEO process implemented in this research:

1. **Topic website research**: Perform market research to validate search interest on *alpiste,* the topic of this research. Confirm that people were searching for alpiste-related information online.
2. **Domain name selection and website setup**: Selection of the domain name, and the technical details of setting up the website and establishing a web hosting service. The consequences of the choices made at this stage may influence the effectiveness of the SEO techniques.
3. **Keyword research**: Effective keyword research is at the heart of SEO. Keywords can be considered to be the link that gets searchers to websites. For this reason, keyword research can be considered the most important part of the SEO process because it is in this phase that keywords are discovered and identified to be used in the SEO efforts (on-page SEO and off-page SEO). Selecting the wrong keywords (those that are not relevant to the theme of your site or those that people are not searching for), can negatively affect your SEO strategy.
4. **Data collection**: Setup a tracking tool for website data collection (such as Google Analytics). It’s important to gather detailed statistics about visitors to the site being optimized because it can provide you with useful data such as how users are arriving at your site, what keywords are responsible for sending the most visitors, what countries are sending the most visitors, etc… This is important because you want to continually track and measure your SEO progress and effectiveness.
5. **On-page SEO strategies implementation**: On-page SEO is the set of techniques that the SEO engineer applies to resources on-page (composition of web pages, use of correct tags and keywords) that have measurable impact on the way the

page or set of pages being optimized is ultimately ranked by Google. These set of on-page SEO strategies discussed in Section 4.5 are listed below:

- *Title tag*: It defines the page title and it communicates what the page is about to the search engines; therefore, the target keyword should be inserted within this tag. It describes the overall theme of the optimized web page. Moreover, the title tag is what gets displayed in Google's search results as the hyperlink title.
- *Description meta tag*: It provides a summary description of the web page and its contents. Also, this description appears (in most cases) in Google's search results, just below the title. For this reason, it's recommended to use your target keyword within the description.
- *Use of robots.txt*: The `robots.txt` file gives directions to search engines of what pages or directories should be crawled. Having this file configured correctly will make sure all your optimized pages get indexed.
- *Optimization of URLs*: An optimized URL is one in which the URL text is self-explanatory and self-documenting. The text of the URL gives the reader a brief description of the theme of the page's content; therefore, the target keyword should be within the URL.
- *Content first*: Search engines process web pages from left to right, top to bottom. A strategic way to structure the web page content is to place the most important text at the top and should contain your target keyword for optimization.
- *Headings tags*: These tags are used to emphasize important text and structure on the web page. They also inform the search engines what your page is about how it's organized. The <h1> tag defines the most important heading and <h6> defines the least important heading, so it's important to include your target keyword into the <h1> tag.
- *Images*: Use the image `alt` attribute to provide an accurate description of the image being used; if possible, use your target keyword in the description. The `alt` attribute from the <img> tag specifies the alternative text of what the image is about in case the image doesn't load or cannot be

displayed correctly. This short text description is helpful to the search engines and to people with disabilities so they can understand what's contained on the image.

- *Use of the "rel=nofollow" attribute*: In a HTML anchor tag <a>, the rel attribute defines the relationship between the current page and the page being linked to by the anchor. The *nofollow* value signals web spiders not to follow the link. In other words, it tells Google that your site is not passing its reputation to the page or pages linked to.
- *Keyword placement*: The placement of the target keyword within the web page being optimized is important. By strategically placing the target keyword in key places, the web page will be seen as relevant page (for search engines and users). This will also help boost the page's search rankings.
- *Sitemaps*: It's recommended to create an HTML sitemap for users and a XML sitemap for search engines. The HTML version provides users easy access to all of the key pages in the site. The XML sitemap makes it easy for search engines know the site's structure; it's a way to tell Google what pages exists in your site so that they can get indexed

6. **Off-page SEO strategies implementation:** Off-page is primarily focused on link-building. This is the process of acquiring backlinks from reputable sites in a gradual reputation-building process. And any technique that the SEO engineer can create that accomplishes this goal is by definition a candidate off-page SEO technique. Some off-page SEO strategies and key ideas discussed in Section 4.6 are listed below:
    - *Importance of keyword in the backlink*: Google's ranking algorithm places high value on the text that appears within the link. The text within the link gets associated with the page and describes the page it links to. For this reason, it's important to have the target keyword within the text of the backlink. Having the keyword in the text will help the site achieve higher rankings.

- *Importance of gradual link-building*: It's important to build backlinks in a gradual manner. The link-building process should be natural and steady. It is for this reason that SEO takes a lot of work and patience to implement Furthermore, it's an intentional part of Google's strategy for gradual reputation building that it not be quick or overnight. In fact, if a site were to acquire dozens or hundreds of backlinks overnight, Google would almost certainly consider this a red flag (spam) that most likely will get your site penalized. But if the site content is compelling, people can find it through search (or through other means) and link to it. If this occurs, the site owner has no control of the amount of backlinks that the site will generate, and certainly Google can detect that this backlinks weren't intentional.
- *Importance of quality content*: As pages are created and uploaded to the site, it's important to publish content that make people want to share, read and link to. When this occurs, the site has a higher probability of getting more backlinks. The content should be of high quality, unique and must provide value to the users. Furthermore, the content needs to be published and promoted where potential users will find it.
- *Writing articles to establish domain authority*: Writing articles and getting them published on other reputable sites, is a strategy that can help your site get backlinks. Getting an article published on trusted sites such as About.com, Wikipedia.org or NewYorkTimes.com and getting a backlink in return, will help increase your website's reputation and achieve higher rankings.
- *Personal networking to establish a reputation:* Reaching out to those in your site's community that "that cover topic areas similar to yours. Opening up communication with these sites is usually beneficial" [19]. Contacting sites that are related to what your site is about is a great way to network, promote and increase your site's exposure
- *Finding your website's natural affinity group:* Find websites that are related or cover similar topics as yours for potential networking

opportunities. Getting backlinks off-topic sites do not count as much as links from sites that have related content to yours.

As seen from the previous list of tasks, the SEO process is involved and it requires extensive keyword research. The experimental site (*casadelalpiste.com*), served as a platform for the trial application of a set of selected SEO techniques. These SEO experiments were evaluated by collection and analysis of real data from the website through the use of Google analytics. Furthermore, their effectiveness on the number of users who visit the site and search engine rankings were measured and analyzed. It was shown that SEO had positive influence on Google search rankings and on the volume of users who visited the site. The results of the research will be summarized next.

6.3 Results

The experimental site was launched in early March 2011. The on-page SEO was completed on March 19, 2011 for the page used in the experiment (off-page SEO was an ongoing activity from March ’11 thru March ‘12). The collected data shows that the implementation of selected SEO strategies had a positive effect on the following metrics:

1. ***Number of Visitors****:* In June 2011, three months after the application of SEO, the experimental site was receiving about 552 monthly visitors. By March 2012, it was getting about 2,500 monthly visitors. A 353% increase in number of visitors. Section 5.1 provides the details.
2. ***Pageviews****:* In June 2011, three months after the application of SEO, the experimental site was generating 1,249 monthly *pageviews*. By March 2012, it was generating 4,166 monthly *pageviews*. A 234% increase in number of *pageviews*. Section 5.2 provides the details.
3. ***Reaching the First Page of Google****:* As discussed in Section 5.3, Google determined that there were 106,000 pages relevant to the target keyword (*beneficios del alpiste*). Before the application of the SEO strategies, the experimental page was not ranking high on Google search results. At the time of this writing, the site has reached the first page of Google and is currently at position No. 6 in the search results for the keyword *beneficios del alpiste.*

The previous data results show the effectiveness of the SEO strategies for the target keyword. The following are two factors that were not achieved in this research:

1. ***Pages/Visit***: I decided to include this additional metric to determine how visitors viewed my website. If users found the website useful or interesting, there is a high probability that they would browse other pages of the site; thus, increasing the number of pages viewed in each visit. As discussed in Section 5.2, in June 2011 *Pages/Visit* was at 2.3 and by March 2012 it had decreased to 1.7. This metric has remained constant, at just below 2.0. This tells me that I need to add more quality content to make the visitors stay at the site longer. Or that there is a problem with the site structure, and I need to make it easy for users to access the other pages.
2. ***Getting to No. 1 Position in Google***: Even though I was able to get the experimental page to the first page of Google, I was hoping to get it to the No. 1 position for my target keyword. While this is possible to achieve, I will need to focus in more link-building activities.

6.4 Limitations of Research

This research used only one keyword (*beneficios del alpiste*) to test the SEO strategies discussed in section 4.5 and in section 4.6. As explained, the SEO process is time-consuming and it requires extensive keyword research for an effective SEO implementation. Although only one keyword was used for optimization, there was noticeable increase in number of visitors to the experimental site and an increase in Google search rankings.

Furthermore, the keyword used was not highly competitive (as discussed in section 4.3). This may have been the reason why reaching top rankings were achieved in a short time. For other, highly competitive keywords a longer time may be needed. Also, more involved SEO strategies may be needed such as link-building.

6.5 Future Research

As search engines continue to update and enhance their search ranking algorithms, SEO engineers must adapt and continually learn new SEO strategies. The results of this research confirm and extend results of earlier SEO research. This paper provided a thorough analysis and

step-by-step implementation of selected search engine optimization techniques that were shown to increase visibility, get more visitors and achieve higher rankings in search results for a general class of website. For this reason, it is hoped that this paper can be used as a guidebook for new SEO engineers and as a basis for later continued SEO research.

References

[1] Lewandowski, Dirk: *The Influence of Commercial Intent of Search Results on Their Perceived Relevance*. iConference (ACM), February: 8-11, 2011.

[2] Wang, et al : *An Empirical Study on Search Engine Optimization Techniques and Its Outcomes*. (IEEE), 2011.

[3] Zhu, et al : *Research and Anaysis of Search Engine Optimization Factors Based on Reverse Engineering*. (IEEE), 2011.

[4] Grzywaczewski, et al : *E-Marketing Strategy for Businesses*. (IEEE), 2010.

[5] Neshat, Hamed Sadeghi: *Ranking of New Sponsored Online Ads*. (IEEE), 2011.

[6] http://trends.builtwith.com/analytics/Google-Analytics (Date retrieved: 2/16/12)

[7] http://support.google.com/googleanalytics/bin/answer.py?hl=en&answer=99118 (Date retrieved: 2/16/12)

[8] Su, et al : *How to Improve Your Google Ranking: Myths and Reality*. (IEEE), 2010.

[9] Li-Hsing, et al :*The Application of Search Engine Optimization Techniques for Internet Marketing: An Example of the Motel Websites*. (IEEE), 2010.

[10] http://support.google.com/insights/bin/bin/answer.py?hl=en&answer=87284 (Date retrieved: 2/29/2012)

[11] http://support.google.com/insights/bin/bin/answer.py?hl=en&answer=87284 (Date retrieved: 2/29/2012)

[12] http://support.google.com/insights/bin/answer.py?hl=en-US&answer=94793 (Date retrieved: 2/29/2012)

[13] http://www.icann.org/en/about/governance/bylaws/bylaws-18mar11-en.htm (Date retrieved: 3/4/2012)

[14] http://en.wikipedia.org/wiki/Content_management_system (Date retrieved: 3/4/2012)

[15] http://wordpress.org/about/ (Date retrieved: 3/4/2012)

[16] http://w3techs.com/technologies/overview/traffic_analysis/all (Date retrieved: 3/7/2012)

[17] http://www.youtube.com/watch?v=AsSwqo16C8s (Date retrieved: 3/10/2012)

[18] http://www.majesticseo.com/reports/neighbourhood-checker (Date retrieved: 3/10/2012)

[19] http://www.google.com/webmasters/docs/search-engine-optimization-starter-guide.pdf (Date retrieved: 3/10/2012)

[20] Dover, Danny. Search Engine Optimization Secrets. Wiley Publishing, Inc., Indianapolis, Indiana. 2011

[21] http://support.google.com/webmasters/bin/answer.py?hl=en&answer=156184 (Date retrieved: 3/24/2012)

[22] http://infolab.stanford.edu/~backrub/google.html (Date retrieved: 3/24/2012)

[23] Page, Lawrence and Brin, Sergey and Motwani, Rajeev and Winograd, Terry (1999) *The PageRank Citation Ranking: Bringing Order to the Web*. Technical Report. Stanford InfoLab.

[24] Gyongyi, Zoltan et al: *Combating Web Spam with TrustRank*. Proceedings of the 30th VLDB Conference, Toronto, Canada, 2004.

[25] https://support.google.com/analytics/bin/answer.py?hl=en&answer=1308626 (Date retrieved: 3/27/2012)

[26] http://support.google.com/analytics/bin/answer.py?hl=en&answer=1006243 (Date retrieved: 3/27/2012)

[27] http://support.google.com/webmasters/bin/answer.py?hl=en&answer=93633 (Date retrieved: 3/31/2012)

[28] http://httpd.apache.org/docs/2.0/howto/htaccess.html (Date retrieved: 3/31/2012)

[29] http://www.computerhistory.org/internet_history/ (Date retrieved: 4/4/2012)

[30] http://www.netvalley.com/cgi-bin/intval/net_history.pl?chapter=1 (Date retrieved: 4/4/2012)

[31] http://www.w3.org/People/Berners-Lee/WorldWideWeb.html (Date retrieved: 4/4/2012)

[32] http://www.livinginternet.com/w/wi_mosaic.htm (Date retrieved: 4/4/2012)

[33] http://www.bloomberg.com/video/67758394 (Date retrieved: 4/4/2012)

[34] http://en.wikipedia.org/wiki/Web_browser (Date retrieved: 4/4/2012)

[35] http://www.internetworldstats.com/emarketing.htm (Date retrieved: 4/4/2012)

[36] http://www.google.com/publicdata/explore?ds=d5bncppjof8f9_&ctype=l&met_y=it_net_user (Date retrieved: 4/4/2012)

[37] http://en.wikipedia.org/wiki/Search_engine (Date retrieved: 4/4/2012)

[38] http://en.wikipedia.org/wiki/Lycos (Date retrieved: 4/4/2012)

[39] http://www.nytimes.com/2011/02/13/business/13search.html (Date retrieved: 4/5/2012)

[40] Dover, Danny: Search Engine Optimization Secrets. Wiley Publishing, Inc., Indianapolis, Indiana. 2011

[41] http://www.nytimes.com/1996/11/11/business/desperately-seeking-surfers.html?src=pm (Date retrieved: 4/5/2012)

[42] http://en.wikipedia.org/wiki/SEO (Date retrieved: 4/5/2012)

[43] http://searchenginewatch.com/article/2049695/Top-Google-Result-Gets-36.4-of-Clicks-Study (Date retrieved: 4/5/2012)

[44] http://en.wikipedia.org/wiki/Canary_grass (Date retrieved: 4/10/2012)

[45] http://en.wikipedia.org/wiki/Shared_web_hosting_service (Date retrieved: 4/11/2012)

[46] http://en.wikipedia.org/wiki/WordPress (Date retrieved: 4/11/2012)

[47] http://en.wikipedia.org/wiki/Google_Analytics (Date retrieved: 4/11/2012)

[48] http://www.pewinternet.org/Reports/2012/Search-Engine-Use-2012/Summary-of-findings.aspx (Date retrieved: 4/12/2012)

[49] http://en.wikipedia.org/wiki/Google_(search_engine) (Date retrieved: 4/14/2012)

[50] http://searchengineland.com/by-the-numbers-twitter-vs-facebook-vs-google-buzz-36709 (Date retrieved: 4/14/2012)

[51] http://en.wikipedia.org/wiki/Search_engine_results_page (Date retrieved: 4/14/2012)

[52] http://en.wikipedia.org/wiki/Pay_per_click (Date retrieved: 4/14/2012)

[53] http://en.wikipedia.org/wiki/Pagerank (Date retrieved: 4/14/2012)

[54] http://en.wikipedia.org/wiki/Web_crawler (Date retrieved: 4/14/2012)

[55] Zhang, et al : *The Research on Search Engine Optimization Based on Six Sigma Management*. (IEEE), 2011.

[56] http://en.wikipedia.org/wiki/Six_Sigma (Date retrieved: 4/14/2012)

[57] http://www.seomoz.org/article/search-ranking-factors (Date retrieved: 4/14/2012)

[58] http://en.wikipedia.org/wiki/Domain_Name_System (Date retrieved: 4/18/2012)

[59] http://helpmyseo.com/seo-tips/312-the-history-of-search-engine-optimization-and-the-optimization-of-your-website.html (Date retrieved: 4/22/2012)

[60] http://support.google.com/adwords/agency/bin/answer.py?hl=en&answer=25148 (Date retrieved: 4/22/2012)

[61] http://www.seomoz.org/beginners-guide-to-seo/growing-popularity-and-links (Date retrieved: 4/22/2012)