Joonas Himmanen

REVIEW OF SEARCH ENGINE OPTIMIZATION TECH-NIQUES

TIIVISTELMÄ

Tiivistelmä: Tämän tutkielman tarkoitus on saada ymmärrystä hakukoneopti-
mointiin ja sen eri tekniikkoihin. Koska on hyvin tärkeää tärkeää olla korkealla
hakukoneiden kuten Googlen hakutuloksissa, nettisivut hyödyntävät erilaisia
tekniikoita nostaakseen näkyvyyttä hakukoneen tuloksissa. Moni näistä tekniik-
koista ei riko hakukoneiden palveluehtoja ja näitä tekniikoita kutsutaan valko-
hattu-tekniikoiksi ja tekniikat, jotka rikkovat näitä palveluehtoja kutsutaan
mustahattu-tekniikoiksi, jotka voivat olla hyvin tehokkaita mutta näiden käyttö
voi johtaa seuraamuksiin hakukoneiden toimesta. On myös tekniikoita, jotka
putoavat niin sanottuun harmaaseen alueeseen näiden kahden tekniikan väliin
ja tästä syystä näitä kutsutaan harmaahattu-tekniikoiksi. Tutkielma käy läpi
näitä tekniikoita sekä niiden hyödyt ja haitat nettisivuille.

Avainsanat: Hakukone, hakukoneoptimointi, SEO, valkohattu, mustahattu,
harmaahattu.

ABSTRACT

Abstract: The purpose of this thesis is to get an understanding on search engine optimization and its techniques. As being on top of the results on search engines such as Google can be extremely important, websites have been using different techniques to increase the ranking. Many of these techniques do not go against the terms of services of search engines and are considered as 'white hat' techniques but those that do go against the terms of services are so called 'black hat' techniques and while they can be very effective, search engines are known to give out sanctions for websites that use these techniques. There are also techniques that fall between the grey area of these two and they are called 'grey hat' techniques. This thesis aims to go through these techniques, their benefits and drawbacks for the websites.

Keywords: Search Engine, Search engine optimization, SEO, White Hat, Black Hat, Grey Hat

FIGURES

TABLE

Table of Contents

1   INTRODUCTION

In the Web 3.0 internet, everyone is trying to gain their own share of something. Whether it is money, clicks, views or prestige, it is an endless competition of trying to one up your competitors, even if you are simply making social media posts of amusing pictures. Everyone is trying to increase their reach, their voice and their influence. To achieve this, multiple techniques and tactics to help you achieve these goals have been made. One of these is search engine optimization (SEO), although in today's world it does not necessarily have to be focused on just search engines and so it could be called algorithm optimization, but for the sake of consistency I will mainly refer to the search engine part in this thesis.

Search engine optimization is a host of methods and techniques of trying to boost your visibility and ranking within the search engine results (Karmakar & Hazari, 2021). Nowadays this is mainly done for Google but of course can also be implemented in other search engines such as Yahoo and Yandex but many of the same ideas are implemented on other services like Facebook, Twitter, Reddit etc., pretty much anywhere where there is some kind of search function, including online shops. It is extremely important for companies to use these techniques to elevate your position in the search results, especially on Google. There is a saying that if it is on the second page of Google, it does not exist and in a way, this is true (Zilancan, 2015). The first results are deemed more trustworthy, accurate and official. In a way it is free advertisement for you, although in most cases it is definitely not free. Notable techniques to do search engine optimization are keywords, tagging, indexing, content updates and so on. But there are also many other techniques that may be a bit more dubious than others.

As search engine optimization is so important for the success of your company, many are willing to do anything to achieve higher ranking within the results. There are techniques that can be very effective in increasing your place in the search results, but they may also be breaking the terms of services of the search engine provider you are using or in some cases, even illegal. These are called the grey hat and black hat techniques, white hat techniques being the ones that are allowed by the service provider (Malaga, 2008). Grey hat techniques may not necessarily be against the terms of service but there could be for example moral ambiguity how the technique is used, or it is not known if the provider would allow the technique being used as in many cases these techniques are not used publicly. (Swati & Ajay, 2013) Black hat techniques have been deemed to be against the terms of service and in some cases may even be illegal. The usage of these techniques can lead to results that make the techniques be very efficient but there are risks involved when using such techniques.

The terms "White hat" and "Black hat" originate from old westerns where bad guys wore black hats, and the heroes wore white hats (Kapersky). They have later been used when talking about coding, hacking and other IT related topics. A black hat hacker for example is someone who uses his skills for malicious intentions and a white hat hacker would be someone who tries to fix things. A grey

hat fits somewhere in between, for example hacking networks on your own to find vulnerabilities but not doing anything malicious with the information.

This thesis aims to study what search engine optimization techniques are used and what implications they have. In this thesis, I will briefly go through the history of search engines and how they work, what is search engine optimization in general, explain what white hat, grey hat and black hat search engine optimization and examples of what kind of techniques can be used and finally, go through what kind of benefits and risks the use of these techniques have. In this thesis, I will answer the following questions:

- *What is white hat, grey hat and black hat search engine optimization?*
- *What are the benefits and drawbacks of using such techniques?*

With these questions answered, you will be able to get the understanding of what is search engine optimization, what are the different types of hats in search engine optimization and what benefits and risks the usage of such techniques can have. The thesis will be done as a literature review. Majority of sources are from scientific papers and conferences but there are also some news articles and expert sources as for some specific topics there are limited amount of actual scientific research done.

2 SEARCH ENGINE AND SEARCH ENGINE OPTI-MIZATION

In order to understand what search engine optimization is and what methods of search engine optimization there are, we first must understand what exactly a search engine is and how search engines work. In this chapter, I will briefly go through what is a search engine and how they work and then go through what is search engine optimization in general.

2.1 Search Engine

Search engines have existed since the beginning of the internet and post-web search engines like Yahoo, Aliweb and Altavista started to appear in the mid 1990's (Alexander, 2014), with Yahoo being a major player in the early days of the internet (Encyclopaedia Britannica). However, since the early 2000's, the search engine market share started to be dominated by Google and has become the de facto search engine to be used, with 92.47% market share as of June 2021 (Statista, 2022). Google's success is due to having a more advanced search engine than its competitors which used a popularity based ranking system. Instead of only using popularity as the main factor, Google created a robust set of algorithms that rank the search results based on multiple different criteria, such as
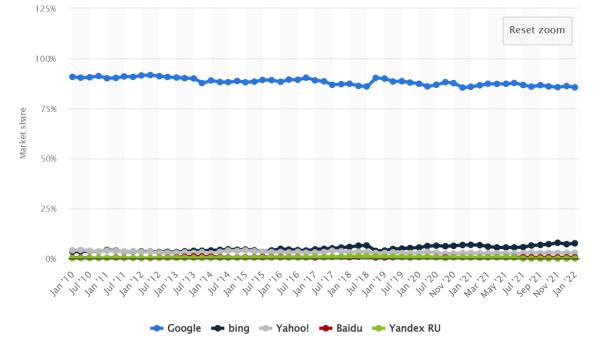

Figure 1. Search engine market share 2010-2022. (Statista, 2022).

keywords, context, links on other sites and relevancy (Heitzman, 2017). Due to

Google's market dominance and success, it has become the main focus of anyone wanting to do search engine optimization. Search engine optimization can of course be done on other search engine platforms such as Bing, Yahoo and Yandex, and there is also platform specific search engine optimization on sites like YouTube, Amazon and eBay but the main focus of every company will be Google. Because of this, I will focus mainly on how Google's search engines work, although most other search engines also use the same basic principles so the same information can mostly be used on other platforms too.

There are several ways search engines can work. Human-powered search engines used to be the most common way of making a search engine (Bradley, 2008). Human-powered search engines relied on website listings submitted by the website maker and the editors of the search engines then ranked the sites according to certain criteria. These are mostly no longer in widespread use, mainly due to the abundance of newer algorithm and crawler-based search engines as those are much more efficient with handling all the data of the ever-expanding internet, however they can be used in small scale things and in closed systems.

Currently the most popular type of search engine are the crawler-based search engines which, while not the inventor of the crawler-based search engine, was popularized by Google and is used to this day (Roth). Web crawlers are essentially bots that crawl through the web, going through websites, indexing and cataloguing the content which is then presented as a search result based on different criteria and algorithms (Google). Crawler-based search engines work on a few principles: Crawling, Indexing and ranking.

Crawling uses web crawlers, sometimes called spiders or robots, to go through the web and websites. Web crawlers are essentially bots that constantly crawl through websites and their content (Shrivastava, 2018). Web crawlers go through the ever-updating list of websites and begin to crawl through them based on automatic decisions of algorithms. Through the same algorithms, the crawler will know how many times, how deep into the website it should go and how often the site should be crawled. There are ways to help crawlers to go through the site like with the usage of sitemap files and I will go through them in a later section about white hat search engine optimization techniques.

Search indexing uses the data provided by crawlers and websites and makes a record of all the websites (Roth). It helps the algorithm by providing categories for keyword terms, pictures and videos, html, text and CSS files and it also grades the quality of the page which the engine can then use to connect and associate with other data such as the search term, user history, location, web provider and time of the day. Currently Google crawls through more than 100.000TB of webpage data so such algorithms are essential for the search engine to be able to function (Google).

Ranking sorts through the info in the index and then ranks them based on search algorithms and data points such as popularity, backlink quality, mobile friendliness, when the page has been updated, engagement, page speed and user experiences (Su, et al. 2014). These algorithms are often the bread and butter of a company, so they are mostly kept secret so only the engineers at Google know

how the ranking algorithms actually work. However ranking algorithms often work how a human would rank a website as page speed, readability and density of the keywords are metrics that humans also value (Roth).

Due to the prevalence of smartphones, search engines may cater different search results to them (Matta, et al. 2020). One of the reasons for this is the different user experience and purpose of use, whereas the search results for the desktop users might include results that are more in-depth, search results on mobile phones can be based on their faster consumability. The mobile phone specific search engine algorithms also rank the websites based on their mobile readability.

2.2   Search Engine Optimization

When doing search engine optimization, the main purpose is simply to increase your ranking on the result pages of search engines (Gudivada, et al. 2015). There are numerous reasons for this but one of the most important reasons is that often if your page past the first search engine result page (SERP), your site does not exist in the minds of your target audience. According to Gudivada et al, over 75% of users never go past the first search engine result page and the top 3 results get the majority of passthrough, over 60%. Search engines are the primary channel for finding info on the web and by having higher spots, you can increase both traffic and trust for your site as there is a reason why most of the time the more shadier sites are never ranked highly (Grappone & Couzin, 2010). Search engine optimization is not just beneficial for the website owners but also for the customers and visitors of the website as often good search engine optimization leads to the most accurate, up to date and correct information which can be crucial whether it comes to a buying decision or finding the most reliable sources of information (Roslina & Shahirah, 2019). While search engine optimization is technically not marketing, it can be part of a wider marketing strategy (Flying V Group, 2021). This is due to its cost-efficiency as while larger search engine optimization marketing strategies can cost tens of thousands of dollars per year, for similar scaled traditional marketing campaigns the prices can go up to as high as millions of dollars for a single 30 second slot on a network television. Especially for smaller budgets search engine optimization can be extremely cost effective as you can do basic but effective search engine optimization all by yourself.

Generally, search engine optimization can be split into two categories, on-page and off-page search engine optimization and for the purposes of this thesis I will use the same categorization when explaining the different techniques of both white and black hat search engine optimization (Matta, et al. 2020). On-page search engine optimization focuses on the internal components of the website and its pages (Zilancan, 2015). The main focus is the content of the page, keywords, internal links, use of tags and structure of the pages. Off-page on the other hand focuses on the external matters of the website such as external links, social media and anything that is done outside of the website.

3 SEARCH OPTIMIZATION TECHNIQUES

In this chapter, I will go into the different types of search engine optimization, white hat, black hat and grey hat search engine optimization techniques. And finally, we will conclude with what are the benefits and drawbacks of such techniques and provide few examples of these consequences from using such techniques.

3.1 White Hat SEO

Search engine optimization is an extremely important part of a company's marketing strategy. The higher you are in the search results, the more clicks and visibility you get so companies aim to use different techniques to increase their ranking. White hat search engine optimization techniques are techniques that are considered to be ethically correct to use and also follow the rules and guidelines set by the different search engines such as Google (Roslina & Shahirah, 2019). Using these techniques will not lead to your website getting sanctioned by the search engines although using them inefficiently or badly can sometimes actually affect your ranking in the search results in a negative way. As the search engines change their algorithms and guidelines constantly, the techniques can also change and some techniques that were effective a month ago can now actually hurt your ranking, so the field of search engine optimization is constantly changing, and updates are almost mandatory. However, even if some techniques are better than others, websites almost always use multiple techniques at the same time to keep up with the tough competition (Karmakar & Hazari, 2021). In the following sections, I will go through the most common search engine optimization techniques separated to the two categories mentioned earlier, on-page and off-page techniques. There are some that could be categorized differently from just these two like paid search engine advertising but as technically they are off-page techniques, they will be categorized as such for the sake of this thesis.

3.1.1 On-page Techniques

On-page search engine optimization techniques are techniques that alter the content of the page itself. One of the most important parts is of course the quality of the content itself (Mittal, et al. 2018). Not only is having good content on the page beneficial to the users but search engines are built to promote higher quality content which is based on for example structure of the page, user engagement and keyword use, latter of which is a very common way of search engine optimization, and it is called keyword optimization.

Using the correct keywords in the site content can have a tremendous impact on the search engine result rankings (Mittal, et al. 2018). These of course depend on how the content on the site has been built but can sometimes be run through optimizer software to identify the best keywords, but it is often manual

work. Best keywords always match the content of the site and sometimes using mismatched keywords can have negative effects on the website's ranking. Generally the amount of keywords out of all the words within the page should be 5-10% (Karmakar & Hazari, 2021). This however is not limited to only the pure content. Vast majority of keyword optimization is done within html tags. Title and heading tags are of course important and often one of the first things that the crawler notices (Mittal, et al. 2018). Pictures and videos are an important part of websites but as crawlers cannot identify or rank them by itself, alt tags are used to describe the content the media has. These are also used by the visitor when they hover a mouse over the picture so they will also be able to read the description and for this reason it might not necessarily be a good idea to fill in as many keywords as possible but to keep it descriptive and readable. Meta tags or meta descriptions on the other hand give a description of the whole webpage. Typically these are only readable within the html code where the crawlers can access them so they are built specifically for search engines and have keywords to reflect this.

In order for the search engine crawlers to be able to know where to go and what to do within the webpage, robots.txt is created (Karmakar & Hazari, 2021). This page almost always includes sitemap XML file which has links to all the pages on the website, or at least the ones the website wants the search engine crawlers to access. These often have a sitemap_index.xml file that includes subsections of the site, such as different language and regional versions of the same website (Valk, 2021). Like mentioned, this guides and helps the crawlers to find the correct pages within the website (Google). Crawlers have set amount of "allowance" for how long they spend in one site so by helping them find the correct pages faster and more efficiently, it can actively increase the ranking of the site within the SERP rankings (Roth). Robots.txt also has other uses such as allowance rules for the crawlers. You disallow a crawler from crawling certain pages or elements and by specifying these rules for a certain user-agent, these rules can be specific to specific search engines. While most search engines follow the same set of rules, they may also follow their own sets of rules, for example Yandex's crawlers understand Clean-param rule that is used when the content of the site is not affected by tags (Yandex) . In the figure 1, all other user-agents (search engines) are allowed to crawl the full site, but Yandex has been disallowed from crawling /*ref= and /*symbol= elements and also given clean-param rules for layout and callback. Host directive is used by Yandex to tell the search engine in what form

to show the website in the search results (Valk, 2021). Sitemap guides the crawlers to the actual sitemap xml-file that contains a list of pages on the website.

```
User-Agent: *
Allow: /

User-agent: Yandex
Disallow: /*ref=
Disallow: /*symbol=
Clean-param: layout
Clean-param: callback

Host: https://www.examplesite.com
Sitemap: https://www.example.com/sitemap_index.xml
```

Figure 2. Example of using Clean-param and Disallow rules.

In addition to sitemap telling where the crawlers should go, another important part of search engine optimization are internal links (Mittal, et al. 2018). These are links inside the website that link one webpage to another one within the website. Good internal linking structure is of course important for search engine optimization purposes, but they are also the same links used by the users of the site. In addition to internal links websites typically have external links too but these are considered off-page techniques which will be explained in the next section.

As the search engines rank pages differently for mobile phone users, a good way to gain higher ranking for mobile phone results is to make the website mobile phone friendly (Schubert, 2016). This can be done by making the whole website easier to browse on mobile phones or have a different site layout for mobile phones. The content itself could be adjusted to suit mobile phone users for example by making the content to be able to be consumed faster as generally the mobile phone users prefer a more casual approach to the content and website design (Roth). While the website could be adjusted to suit both mobile and desktop users, depending on the website and the target audience of the website, the whole website could be made to cater only mobile users.
.

3.1.2   Off-page Techniques

One of the first things you usually do when doing search engine optimization strategy is a proper competitor analysis which can be considered an off-page search engine optimization technique (Mittal, et al. 2018). This is typically done by studying the keywords of other similar sites and trying to gain better ranking in certain keywords. You might not be able to get the best ranking with keywords such as "Mobile phone repair" but by targeting specific keywords such as model

and part numbers you might be able to hit niches that will bring you more customers, such as "mobile phone repair Exynos 1280" which would target phones with specific processors. Competitor analysis can of course be used for pretty much any aspect of the search engine optimization strategy such as by adjusting the content itself and having certain links on the site.

In addition to on-page internal linking techniques, there are also many off-page techniques that are used in search engine optimization. One of them is backlinking in which other external websites link back to your website. The search engine uses backlinks based on the website and page the link is coming from, depending on whether they are relevant to your site and the quality of the sites is high (Jain & Dave, 2013). There are several ways of getting backlinks, common way being through partners and sometimes these are done through paid backlinks which could be considered a grey hat method. Of course, if you own multiple websites with relevant content, you should of course backlink your own sites (Mittal, et al. 2018). Websites also often manage link exchange so they backlink to each other's page. As poor quality backlinks, whether due to poor quality of the site or irrelevant content, may not increase your ranking or in some cases hurt your ranking, strategy for backlinking should be built with care.

As social media has become an almost symbiotic part of everyday life, it is also a vital part for search engine optimization strategies (Karmakar & Hazari, 2021). This of course includes having a social media presence in places like Twitter, Facebook, LinkedIn, Instagram and YouTube. All of these can be used for backlinking purposes, but they can also work as advertising and promotional tools. They allow for more visibility and engagement with your target audience and with quality content they can bring new audiences to your website or at least spread brand awareness (Mittal, et al. 2018). Having a good social media presence can however be very hard to manage but depending on the site and the content, it might be almost inevitable in today's world.

3.2 Black Hat SEO

Whereas white hat techniques are considered ethical search engine optimization, black hat techniques are often considered to be the other side of the coin. While normally black hat techniques are not illegal, they are often considered unethical (Swati & Ajay, 2013). Search engines aim to bring the most accurate and quality content to the users by using different algorithms and requirements. Sites that use black hat techniques are usually of lower quality than those that do not and by using these techniques, you are essentially pushing lower quality content to the users of the search engines which is the reason search engines do not allow such techniques to be used and often sanction sites for using black hat methods (Malaga, 2008). Some of the white hat techniques can be turned into black hat techniques by using the technique in a way that is deemed to abuse the system, for example backlinking is a common white hat technique but by spamming backlinking farms the technique can become a black hat technique (Duk, et al. 2013). In this section, I will go through some of the most common black hat search engine optimization techniques.

3.2.1   On-page Techniques

One of the most common ways of black hat search engine optimization techniques is usage of invisible content (Duk, et al. 2013). As the name suggests, the pages have content that is invisible to the user but visible for web crawlers and robots. The purpose is to insert content that doesn't necessarily have anything to do with the content but helps with increasing the ranking of the website (Malaga, 2008). In most basic form this is achieved by using for example white text on a white background or very small font, so it is nearly impossible to see for the user. They could also be hidden outside of the viewable area of the page or behind images and videos. Nowadays it is done using cascading style sheets (CSS) which allows hiding of the elements completely (Swati & Ajay, 2013). Links especially are often hidden with this technique as they help the ranking more than just keywords.

With a similar aim to invisible content, the purpose is to have as many keywords as possible on the webpage which attracts search engine spiders and can increase the ranking in SERP. These keywords are often hidden and stuffed in the elements and tags (Zuze & Weideman, 2011). Often the information these keywords have are of low quality and of course do not benefit the users in any way. It can be seen as similar to having hundreds of different tags and keywords for example on eBay listings and YouTube videos, although the latter has ways to identify unnecessary keywords (Abidin & Brown, 2019).

Cloaking techniques are often known for the case of the car maker BMW who used this black hat technique in 2006 (Solihin, 2013). One of the cloaking technique's purposes is to create pages specifically for search engine crawlers and redirect actual users automatically to the correct page (Wang, et al. 2014). The purpose of this is to fill the fake page with keywords, links and content. The content used in these pages are often regarded as poor content but are specifically chosen due to them affecting the ranking in SERP in a positive way. Cloaking can also be used to target different search engines. As all search engines can have vastly different algorithms and criteria how they rank pages for serp purposes, a site that uses cloaking technique has a page for each different search engine and its web crawlers. It does so by checking the IP address of the crawler and based on it, redirects the crawler to the correct page which is specifically built according to the criteria of that specific search engine and its algorithms (Malaga, 2008). Cloaking can also be used in a bit more malicious way where the search engine crawler is guided to the page indicated by the search results, but the user is sent to completely different page, such as gambling sites, pornographic sites or sites with the aim to scam the user in some way, these are also sometimes called doorways (Wang & Ma, 2006). This is obviously highly unethical and, in some cases, could be against the law, depending on the content of the site.

Different kinds of blogs are also used for search engine optimization purposes, such as the blog-ping technique. The purpose of blog-ping technique is to attract search engine spiders by continually pinging with blog updates. In this technique, the website has hundreds of blogs and then constantly pinging the

blog servers with info that the blog has been updated which in turn attracts search engine spiders and allows for the site to receive higher SERP ranking. This method can also be used as a white hat technique but of course the purpose is to ping only legitimate blogs and less frequently (Pohjanen, 2019).

One of the more effective ways of search engine optimization are the different linking techniques. For white hat search engine optimization techniques, there are numerous proper ways of doing this such as backlinking. However as always, there are also techniques that are of a more darker shade. As search engines use links as one of the criteria's when it comes to ranking the web pages, link spamming is a very effective way of achieving results (Swati & Ajay, 2013). The purpose is to have as many good quality links in the site as possible which obviously helps with the ranking. Essentially any time a site has links that have a sole purpose of increasing the SERP ranking and does not bring value to users, it is considered link spamming. These are often hidden in different ways, often outside of the visible area of the site.

Another way of using links as a black hat technique are link farms (Karmakar & Hazari, 2021). These are pages that have a sole purpose of having hundreds of links to sites that might be completely unrelated to the site in question or to its content. These are made automatically with optimizer tools. They might create a new link farm or use an existing one. These can also be part of linking schemes where you buy or sell links in linking pages and often work similarly to link farms.

3.2.2 Off-page Techniques

Typically, while black hat search engine optimization techniques are against the search engine's terms of service, they are not necessarily doing anything illegal. However, there are some techniques and practices that can be against the law, some of them being more malicious than others. The most common way of this happening is copyright infringement by stealing content from other websites (Swati & Ajay, 2013). Automatic optimizers use this to generate pages by copying content from other pages based on keywords suitable for the site, often making slight changes like use of synonyms so it won't be as similar (Gudivada, et al. 2015). This is especially targeting high ranking webpages, for the obvious reason. Sometimes this is hidden inside already existing pages but sometimes it is simply copying the whole page and using the already existing search engine optimization tricks within the page, boosting the site's SERP ranking. As this is blatant copyright infringement, it is of course against the law, but it can even actively hurt the site where the content was stolen from as often search engines penalize sites due to duplicate content (Malaga, 2008).

There is however another more direct way of affecting your competitors' ranking, known as bowling techniques, where the purpose is to make your competing sites' ranking to drop so you can take its place. One of these techniques is extensive linking to the competitors' site from sites that the search engine might consider bad (Malaga, 2008). For example, linking to the site from gambling sites, sites with illegal material, adult oriented sites or from known scam sites. These

might affect the ranking of the site due to penalties and in some cases, it might lead to a ban. This is of course good for the site that uses such techniques as it drops your competitors' from ahead of you, increasing your sites' rank. In some jurisdictions this might not technically be illegal but of course this is considered highly unethical technique.

Another example of bowling technique are HTML injections. Its purpose is to change the content of competitors' sites with an HTML injection in a way that penalises the sites' ranking (Malaga, 2008). These can vary to bad or extensive use of keywords, duplicating keywords or sometimes even blatantly using badly obfuscated black hat search engine optimization techniques which can end up in a complete ban from the search engine. As HTML injections are considered a type of hacking, these methods are illegal in the majority of jurisdictions and of course, considered unethical.

And finally, there is guestbook spamming which is a simple off-page linking technique (Malaga, 2008). In guestbook spamming the user of this technique essentially finds pages with some sort of comment section, typically with automated tools, and uses that to submit a comment with the link. The crawlers then crawl these comment sections and are forwarded to the site in the link, allowing for better SERP ranking.

3.3   Grey Hat SEO

As generally in life, almost nothing is black and white and the same applies to search engine optimization. For some techniques the line of licit and illicit is clear but for some it is harder to know what is allowed and what is not, what constitutes fair use of links and what is counted as spamming. Grey hat optimization is the unclear line between black and white, and the answer whether it is illicit or not is often that we don't know (Matta, et al. 2020). The definition of grey hat optimization itself is not clear either, sometimes it is used for specific techniques but sometimes it is used for websites itself when the site uses both white hat and black hat methods, often the site having content of decent quality while also having faked content that is there to fool the search engine crawlers. There are actually relatively little studies into grey hat search engine optimization techniques itself, often just as a passing mention in different papers and articles (Raiber, et al. 2013).

The use of grey hat techniques is questionable or unethical but not always against the rules. Most of the techniques are the same ones that are used in white and black hat search engine optimization but just somewhere in the middle so it is actually hard to explain what exactly constitutes a grey hat method until search engine specifies where the line goes so listing them in a same way as white and black hat techniques would be counterintuitive (Savoska, et al. 2016). There are some techniques that are more grey hat specific, being questionable but not necessarily against the terms of services of search engines. One of them being paid services such as purchasing links and followers. Paid links are very common as it allows for smaller sites to easily gain access to links for higher quality sites. This technique is of course a bit questionable as it could promote sites with lower
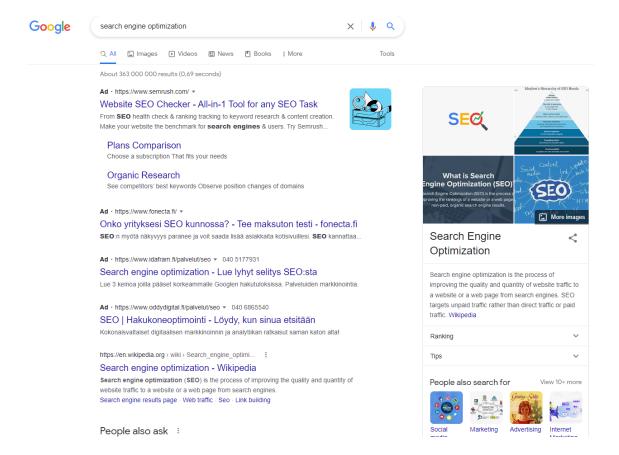
Figure 3. Google ads on search results page on April 8th 2022

quality content to higher SERP ranking and is often even considered to be against the terms of service of the search engine which in that case would make it a black hat technique. Buying followers on social media is almost always against the terms of service but if the purpose is to use the increased reach of the social media platform to promote your website, technically the website has not necessarily broken the terms of services of the search engine. Another highly debated topic are the paid spots on top of the search results (Hallam, 2019).For example, on Google you can purchase search result-like advertisements that are shown with specific search keywords and they are often above the first actual results. As these are obviously allowed and endorsed by the search engines, this would make the practice a white hat technique, but this would raise the question of whether this is ethical practice as the practice could put lower quality content at the top of the results. While these are technically advertisements and not actual search results, because the advertisements are based on the search keywords and are displayed on top of the results, you could argue that these are in fact part of the search results (Bruceb Consulting, 2020). As shown on the example picture, on full-hd display screen zoom set at 100%, by default you see only one actual search result without scrolling down with four of the first spots used by advertisement. These kinds of things are the reason defining grey hat search engine optimization is a complicated subject and shows that there is room for further research regarding it.

Technique	White Hat	Black Hat
Content quality and accuracy	X	
Keyword accuracy	X	
Using tags	X	*
Creating a sitemap, robots.txt	X	
Internal links	X	*
Mobile friendliness	X	
Competitor analysis	X	*
Back linking	X	*
Social media	X	
Invisible content		X
Keyword spamming		X
Cloaking techniques, doorway pages		X
Guest book spamming		X
Blog-ping techniques		X
Link spamming		X
Link farms		X
Paid links	*	X
Stealing content		X
Bowling techniques		X
HTML injections		X

Table 1. List of the mentioned search engine optimization techniques. The use of character * means that the technique can also be used in this way but is more commonly used for the other type of technique. Grey hat techniques are not included due to the ambiguity.

3.4 Benefits and Drawbacks

As in anything, whenever there are benefits in your actions, there are also drawbacks, and it always comes to balance. In regard to search engine optimization, it comes to balance, to whether the benefits outweigh the drawbacks, the risks. As we know now, the benefits of using search engine optimization outweigh the drawbacks of it, not doing at least the most common ways of optimization would generally be a very bad business decision. Yet sometimes businesses and site owners do this knowingly. In this final section, we will go through some of the benefits and drawbacks of doing search engine optimization in general and also for each hat of search engine optimization.

3.4.1 General Benefits and Drawbacks

One of the main reasons why websites do search engine optimization is its cost efficiency, especially when compared to more traditional types of advertising such as TV, newspaper or online advertising (Flying V Group, 2021). A lot of exposure for the potential users or customers are done within search engines and the vast majority of the clicks are from the first page of the search results which is the reason search engine optimization is so important. Search engine optimization is not cheap by all means, small companies wanting to have a proper search engine optimization strategy may have to pay over $1,000 per month and larger enterprises may have to expect to pay millions of dollars (Brown, 2021). One of the reasons for its efficiency is that its organic. While the cost search engine optimization is not calculated with Pay Per Click (ppc) like for example Google Ads would be, comparatively the dollar cost of per user gained with search engine optimization is much lower than with typical advertising campaign where you typically pay approximately $1 per click but which may not lead to a staying customer (McCormick, 2022). Especially during the Covid-19 pandemic, the cost of online advertising has risen tremendously due to it becoming more prevalent as people have spent more time online than previously (Loeb, 2021). Unlike online advertising, search engine optimization also has the cheapest option, free (Brown, 2021). As search engine optimization is often just managing the content and the code of the site, this is something anyone can do by themselves to some extent, especially as there are numerous free online guides to explain how to do search engine optimization. Many sites also have free tools to audit your website to see what problems or errors there might be that prevent a higher ranking (Varagouli, 2022). Some services such as WordPress and Squarespace also do some search engine optimization automatically to the websites created on their platforms. However due to the everchanging algorithms of search engines it is often a cat and mouse game that the vast majority of people cannot participate in for extended periods of time.

There are however some detriments to search engine optimization. Sometimes of course a site may choose to not do search engine optimization as it is deemed unnecessary, one reason being that it is only for internal or specific use

and better ranking is not relevant to the website. But doing search engine optimization rarely has any negative impact on the site itself unless done badly or with the use of black hat techniques, it is the fact that to have a successful site, doing search engine optimization is a necessity and not optional (Khedkar & Khedkar, 2021). In addition to more traditional marketing techniques, companies now have to allocate more resources into search engine optimization, whether it is by paying a company to do it or hiring people to do it within the company. While search engine optimization is more efficient than traditional marketing techniques, it is not cheap and typically you have to constantly update your optimization strategies as the methods and algorithms change. This is one of the reasons many smaller businesses do not do extensive search engine optimization and sometimes don't do it at all as it can feel like just another endless money sink (Snider).

3.4.2   White Hat Benefits and Drawbacks

In this section and the sections about black hat techniques, I will go through why and what kind of benefits and drawbacks there are using just the specific type of techniques over the others. The benefits of using just white hat techniques mainly comes to the ethical reasons and the fact that using white hat techniques will not lead to sanctions by the different search engines. Companies are typically very protective over their brand image and while using unethical search engine optimization techniques is not as obvious as for example a company using underaged labor in 3rd world countries, it can leave a black mark on a company's reputation (Gudivada, et al. 2015). Websites talking about search engine optimizations are still talking about BMW's use of black hat techniques despite it happening in 2006 and has tarnished the reputation of BMW, at least when it comes to ethical marketing strategies. Using white hat techniques will mean your website won't face any sanctions or in the worst-case scenario, blacklisting by the search engines. Being blacklisted by Google for example could become a death sentence for a company as in many people's eyes, if it is not on Google, it does not exist (Reputation Defender, 2022).

The drawbacks for using only white hat techniques are often the same as mentioned in the general drawbacks, search engine optimization costs money. Every website is using the same techniques as others and when two websites are fighting over the same keywords, it often comes to which company is using more money and becomes the endless cat and mouse game of following the whims of the algorithms (Verbolia, 2019). This can be very exhausting, especially for smaller companies that do not have resources to keep up with the competition. This is why many websites turn to alternative ways of increasing their ranking on the search results.

3.4.3   Black Hat Benefits and Drawbacks

The benefits of black hat techniques are a response to the negative side of white hat techniques. While the users of white hat techniques are limited to fighting

over the same techniques that answer the same whims of the algorithms, black hat techniques widen the available scope. By using black hat techniques, you don't necessarily have to spend as much resources as those that use only white hat techniques. You can do the bare minimum of white hat optimization and by applying additional techniques in the form of black hat techniques and doing the same bare minimum on them, you can achieve the same results if not better than by just using white hat techniques. Techniques like link farms and keyword spamming are extremely effective as those are the things search engines look for when ranking the pages. The main benefit of the use of black hat techniques is increased efficiency in terms of cost and time (Swati & Ajay, 2013). Depending on the purpose of the optimization, there are other benefits like being able to cloak the actual purpose of your site or being able to keep your own site clean but using the techniques to attack a competing site in a way that may lead them to being sanctioned by the search engines in some way.

The drawbacks of using black hat techniques, however, are very apparent. There is a reason why the majority of sites do not use these techniques; They go against the terms of service of search engines and breaking these terms may lead to sanctions. Typically, these sanctions lower your ranking on the search results but in some cases the whole website could be blacklisted from search engines all together which for many websites is a death sentence (Gudivada, et al. 2015). One of the more famous case of this happening is when BMW.de and Ricoh.de used cloaking techniques to lead crawlers to a different site than normal visitors (Segal, 2011). This ended up in Google giving BMW's site a ranking of zero which essentially deleted the site from Google for a period of time. Back in 2006 there were some other prominent search engines but Google being the main one it still hurt BMW's business and marketing strategy a lot, granted this only affected the German BMW.de page. Google also sanctioned German office-equipment Ricoh in the same way (Malaga, 2008). J.C. Penney was also sanctioned for the use of black hat techniques, albeit the punishment was short-lived (Segal, 2011). In 2011 it was found out that J.C. Penney was using a link farm to have thousands of sites of irrelevant content link back to JCPenney.com. This caused J.C. Penney to be at the top results of a vast number of different keywords, even if it had nothing to do with J.C. Penney. This was found out and Google issued sanctions against J.C. Penney and the ranking of the site fell for a few months. Since then, this process has become more automated due to more efficient algorithms, so the search engines identify the use of these techniques more frequently. This means that while the benefits of using black hat techniques are real, they are also often very short lived so as with white hat techniques, it ends up being another endless cat and mouse game to stay ahead of the search engine's algorithms.

3.4.4   Grey Hat Benefits and Drawbacks

As with the techniques itself, the benefits and drawbacks are a middle ground between white and black hat techniques (Raiber, et al. 2013). Grey hat techniques are often less efficient than black hat techniques, but this also means it is less likely the site will get sanctioned. The real benefit of using grey hat techniques is

when the technique is not technically against the terms of services or could be interpreted to not be against them, it is the ambiguity of grey hat that is its strongest point. In the case your site does get sanctioned, you might be able to appeal against the decision based on this ambiguity. But as the terms and algorithms change constantly, it is not the perfect solution either. What is a grey hat technique today, might become a black hat technique tomorrow in the eyes of search engines.

4   CONCLUSION

The purpose of this thesis was to get an overview of the search engine optimization landscape and its techniques through a literature review. The field of search engines and search engine optimization has changed a lot in the past 15 years and will keep changing so reviewing the industry of search engines should happen constantly. In this thesis I went through what is a search engine, what is search engine optimization, what are the different hats of search engine optimization and then dove into the specific techniques and their benefits and detriments. For this thesis, I asked the following questions:

- *What is white hat, grey hat and black hat search engine optimization?*
- *What are the benefits and drawbacks of using such techniques?*

After this, I dove into what is a search engine and how it works. We learned that search engines are not a new phenomenon, they have been around since the early 1990's but caught prevalence after the mid 1990's with the introduction of such search engines as Yahoo and Altavista. However, the search engine industry has been dominated by Google since the early 2000's and has continued to do so. The way search engines work has of course evolved over the years and the current search engines work by employing automatic web crawlers that go through the website, index its content and then rank the page and the site according to complicated and often hidden algorithms.

After the introduction, I go through what search engine optimization in general is and what are the different hats of search engine optimization. We learn that search engine optimization is an implementation of different techniques that allow the webpages to be ranked higher in the search engine results. This is deemed crucial for businesses as most users do not look at the results after the first few results or especially after the first page of results. It has become an important part of the wider marketing strategy of websites.

I then go through the meaning of the different hats in search engine optimization. White hat search engine optimization techniques are techniques that are considered ethically correct to use and do not break the terms of services of the search engines. These techniques include things as content management, backlinking, correct use of tags and social media marketing. Black hat techniques are the other side of the coin, these are techniques that are considered unethical and typically go against the terms of services of the search engines. Black hat optimization uses techniques such as cloaking, hidden content, keyword stuffing, link spamming and the sites associated with these techniques are generally considered to be of low quality. Grey hat search engine optimization techniques are techniques that are either between black hat and white hat techniques or it is not fully known if the technique is against the terms of services. These techniques can often go under the radar due to the ambiguity and while they might not technically break the terms of services, they are generally considered to be on the unethical side.

I then go through some of the benefits and drawbacks of using the specific techniques over the others. Using white hat techniques is considered to be ethically correct and doesn't pose the risk of being sanctioned by the search engines. However, while white hat techniques are very efficient compared to traditional marketing, it can become an endless resource sink that has to be adjusted constantly. Black hat techniques on the other hand can be more efficient and requires less resources compared to white hat techniques. However, the use of black hat techniques can lead to extreme sanctions by the search engines, the worst of them being blacklisted completely from the search results which can be a death sentence for a website. Grey hat techniques can avoid the drawbacks of the black hat techniques to some degree while enjoying more efficiency over the white hat techniques. However, using grey hat techniques is not risk free either and due to changing algorithms, what was deemed to be okay the previous day, becomes against the terms of services the next day.

While making the thesis I noticed the lack of research into the grey hat techniques. As mentioned by Raiber, et al. (2013), grey hat techniques are often just briefly mentioned in different researches and there is very little actual research into the subject of grey hat techniques specifically. This could indicate that there could be a need for more in-depth research into this and it is an avenue I could choose to go into when preparing for the master's thesis.

5   SOURCES

Karmakar, R., & Hazari, A. (2021). A Study Note On Search Engine Optimization Techniques. The University of Burdwan

Zilancan, J. (2015). Search Engine Optimization. CBU International Conference Proceedings, ISE Research Institute, vol. 3(0), pages 506-510, September.

Malaga, R. (2008). Worst Practices in Search Engine Optimization. Communications of the ACM, December 2008, Vol. 51 No. 12, Pages 147-150

Swati, P., & Ajay, S. (2013). Search Engine Optimization: A Study. Research Journal of Computer and Information Technology Sciences, Vol. 1(1), P.P 10-13, February 2013.

Kapersky Lab. What is a Black-Hat hacker? https://www.kaspersky.com/resource-center/threats/black-hat-hacker (Accessed on February 27th 2022)

Alexander, D. (2014). The Easy Guide to SEO. Booktango.

Encyclopaedia Britannica (2022). Yahoo! https://www.britannica.com/topic/Yahoo-Inc (Accessed on April 8th 2022)

Statista (2022). Worldwide desktop market share of leading search engines from January 2010 to January 2022. https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/ (Accessed on April 8th 2022)

Heitzman, A. (2017). How Google Came To Dominate Search And What The Future Holds. Forbes. https://www.forbes.com/sites/forbesagencycouncil/2017/06/05/how-google-came-to-dominate-search-and-what-the-future-holds/?sh=5f93b9183872 (Accessed on April 6th 2022)

Bradley, P. (2008). Human-powered Search Engines: An Overview and Roundup. 30-January-2008 Publication: Ariadne Issue 54.

Roth, L. How Search Engines Work: Everything You Need to Know to Understand Crawlers. Alexa. https://blog.alexa.com/how-search-engines-work/ (Accessed on April 4th 2022)

Google. Get Your Website on Google. https://developers.google.com/search/docs/basics/get-on-google (Accessed on April 1st 2022)

Google. The Basics of How Search Works. https://developers.google.com/search/docs/basics/how-search-works (Accessed on April 1st 2022)

Google. How Search Organizes Information. https://www.google.com/search/howsearchworks/crawling-indexing/ (Accessed on April 1st 2022)

Google. Create a robots.txt File. https://developers.google.com/search/docs/advanced/robots/create-robots-txt (Accessed on April 1st 2022)

Google. Learn About Sitemaps. https://developers.google.com/search/docs/advanced/sitemaps/overview (Accessed on April 1st 2022)

Shrivastava, V. (2018). A Methodical Study of Web Crawler. Journal of Engineering Research and Application. Vol. 8, Issue 11 (Part -I) Nov 2018, pp 01-08.

Su, A., & Hu, Y., & Kuzmanovic, A., & Koh, C. (2014). How to Improve Your Search Engine Ranking: Myths and Reality. ACM Trans. Web 8, 2, Article 8 (March 2014)

Matta, H., & Gupta, R., & Agarwal, S. (2020). Search Engine Optimization in Digital Marketing: Present Scenario and Future Scope. 2020 International Conference on Intelligent Engineering and Management (ICIEM).

Gudivada, V., & Rao, D. (2015). Understanding Search-Engine Optimization. Understanding Search-Engine Optimization. East Carolina University.

Grappone, J., & Couzin, G. (2010). Praise for Search Engine Optimization: An Hour a Day. Wiley Publishing.

Roslina, A., & Shahirah, M. (2019). Implementing White Hat Search Engine Technique in E-Business Website. University of Malaysia Pahang.

Flying V Group, 2021. How SEO Costs Compare to Traditional Advertising. https://www.flyingvgroup.com/blog/2021/01/05/how-seo-costs-compare-traditional-advertising/ (Accessed on March 12th 2022)

Mittal, M., & Kirar, N., & Meena, J. (2018). Implementation of Search Engine Optimization Through White Hat Techniques. International Conference on Advances in Computing, Communication Control and Networking (ICACCCN2018).

Yandex, 2022. The Clean-param Directive. https://yandex.com/support/webmaster/robot-workings/clean-param.html (Accessed on April 8th 2022)

Valk, J. (2021). The Ultimate Guide to Robots.txt. Yoast. https://yoast.com/ultimate-guide-robots-txt/ (Accessed on April 8th 2022)

Schubert, D. (2016). Influence of Mobile-friendly Design to Search Results on Google Search. Procedia - Social and Behavioral Sciences, Volume 220, 31 May 2016, Pages 424-433.

Jain, A., & Meenu, D. (2013). The Role of Backlinks in Search Engine Ranking. International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 4, April 2013.

Duk, S., & Bjelobrk, D., & Čarapina, M. (2013). SEO in e-commerce: balancing between white and black hat methods. Polytechnic of Zagreb.

Zuze, H., & Weideman, M. (2011). Keyword stuffing and the big three search engines. Online Information Review Vol. 37 No. 2, 2013 pp. 268-286, Emerald Group Publishing Limited.

Abidin, C., & Brown, M. (2019). Microcelebrity around the globe: Approaches to cultures of Internet Fame. Emerald Publishing.

Solihin, N. (2013). Search Engine Optimization: A Survey of Current Best Practices. Grand Valley State University.

Wang, D., & Der, M., & Karami, M., & Saul, L., & McCoy, D., & Savage, S., & Voelker, G. (2014). Search + Seizure: The Effectiveness of Interventions on SEO Campaigns. University of California, George Mason University.

Wang, Y., & Ming, M. (2006). Detecting Stealth Web Pages That Use click-Through Cloaking. Strider Search Ranger Report - Part 4. Microsoft Research, Microsoft Corporation.

Pohjanen, R. (2019). The Benefits of Search Engine Optimization in Google for Businesses. University of Oulu.

Raiber, F., & Collins-Thompson, K., & Kurland, O. (2013). Shame to be Sham: Addressing Content-Based Grey Hat Search Engine Optimization. Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval.

Savoska, S., & Naka, N., & Ristevski, B. (2016). Search Engine Optimization on PHP based web pages in practice. Emerging Research and Solutions in ICT 1(1):45–58.

Hallam (2019). Google's own advertising: competition, monopoly and fairness. https://www.hallaminternet.com/unethical-google-advertising-monopoly/ (Accessed on April 7th 2022)

Bruceb Consulting (2020). Google, Microsoft, And Unethical Behavior. https://www.bruceb.com/2020/01/google-microsoft-and-unethical-behavior/ (Accessed on April 7th 2022)

Brown, N. (2021). How Much Does SEO Cost? The True Price of SEO Services in 2022. https://accelerateagency.ai/how-much-does-seo-cost-pricing (Accessed on April 7th 2022)

McCormick, K. (2022). How Much Does Google Ads Cost in 2022? WordStream. https://www.wordstream.com/blog/ws/2015/05/21/how-much-does-ad-words-cost (Accessed on April 7th 2022)

Loeb, W. (2021). The Rising Costs Of Digital Advertising Will Force Spending Shifts. https://www.forbes.com/sites/walterloeb/2021/08/04/rising-costs-of-digital-advertising-will-force-spending-shifts/?sh=5043444c5a7c (Accessed on April 7th 2022)

Varagouli, E. (2022). How to Perform an SEO Audit in 18 Steps. Semrush. https://www.semrush.com/blog/seo-audit/ (Accessed on April 8th 2022)

Khedkar, C., & Khedkar, A. (2021). Search Engine Optimization: An Essential Part of the Marketing Toolkit. Vidyabharati International Interdisciplinary Research Journal 12(2), 106-109.

Snider, S. (2022). Why People Are Afraid of SEO? https://www.seo.com/blog/why-people-are-afraid-of-seo (Accessed on April 8th 2022)

Reputation Defender (2022). If you're not in Google, you don't exist. https://www.reputationdefender.com/blog/orm/if-youre-not-google-you-dont-exist (Accessed on April 7th 2022)

Verbolia (2019). SEO measures bringing Return on Investment. https://www.verbolia.com/seo-measures-bringing-return-on-investment/ (Accessed on April 8th 2022)

Segal, D. (2011). The Dirty Little Secrets of Search. New York Times. https://www.nytimes.com/2011/02/13/business/13search.html (Accessed on April 4th 2022)