



UPPSALA
UNIVERSITET

Search Engine Optimization and the Long Tail of Web Search

Dennis Johansson

Uppsala University
Department of Linguistics and Philology
Språkteknologiprogrammet
(Language Technology Programme)
Bachelor's Thesis in Language Technology, 15 credits

June 16, 2016

Supervisors:
Mats Dahllöf, Uppsala University
Vincenza La Starza, iProspect

Abstract

In the subject of search engine optimization, many methods exist and many aspects are important to keep in mind. This thesis studies the relation between keywords and website ranking in Google Search, and how one can create the biggest positive impact. Keywords with smaller search volume are called "long tail" keywords, and they bear the potential to expand visibility of the website to a larger crowd by increasing the rank of the website for the large fraction of keywords that might not be as common on their own, but together make up for a large amount of the total web searches. This thesis will analyze where on the web page these keywords should be placed, and a case study will be performed in which the goal is to increase the rank of a website with knowledge from previous tests in mind.

Contents

Acknowledgements	4
1 Introduction	5
1.1 Purpose	5
1.2 Outline	6
2 Background	7
2.1 Search Engine Optimization	7
2.1.1 Best Practices Overview	7
2.1.2 Keywords and Site Data	8
2.1.3 Advanced Factors	8
2.1.4 SEO Tools	10
2.2 The Google Algorithm	11
2.3 Long Tail Searches	12
2.3.1 Defining the Long Tail	12
2.3.2 User Patterns	12
3 Method	14
3.1 Keyword Analysis	14
3.2 On-page Analysis	15
4 Results	17
4.1 Keyword Data	17
4.2 Website Data	19
5 Discussion	23
5.1 Keyword Phase	23
5.2 Website Phase	23
5.3 Future Research	24
6 Conclusion	25
Bibliography	27

Acknowledgements

I would first like to thank my primary supervisor Mats Dahllöf for pointing me in the right direction and for giving overall guidance with the thesis from start to finish. I would also like to express my gratitude to my secondary supervisor Vincenza La Starza and iProspect for helping me throughout the course of the work, and for welcoming me to their office and offering their resources. This thesis would not be possible without her. I am also grateful to Odd Jakobsson for his good advice and proof-reading.

1 Introduction

1.1 Purpose

Search engine optimization (SEO) is a vast subject with many different approaches. The foundation of this particular thesis lies in computational linguistics and information retrieval, which might prove to give a different perspective than what is usually seen in this subject. SEO can be summarized as the different methods that can be used to increase the ranking of a website in search engines such as Google and Bing. There are many companies that currently work in the field of SEO, one of them being iProspect (a part of the Dentsu Aegis network), and this study will be performed in co-operation with them.

One major aspect of SEO is the keywords (and key phrases) on websites, since well-chosen keywords increases the likelihood of users finding their way to the site, which in turn increases its rank in search engines. Relevant keywords can be found with various tools that fetch data from Google regarding factors such as its popularity, difficulty to rank with, as well as related keyword suggestions. Some more uncommon and less searched for keywords are gaining power in SEO, also known as "long tail" keywords. By including these keywords on a website by rewriting the content to match user needs, one might draw potential customers or users who are looking for something rather specific to the website. Although long tail keywords make up for a small fraction of the total web searches on their own, these words usually have a lower competition with other websites, and might give the website an extra edge in the world of search engines, as well as leading important users to the site. The journey from finding keywords to including them on your site is long and requires research, and this thesis will investigate whether or not it is worth it and how these keywords are ideally placed on the website, with the usage of various SEO tools from Google and Moz.com. The thesis will serve as an example of how companies or owners of websites can draw a specific and important audience to their website by optimizing their linguistic content. The conclusion of this thesis will therefore attempt to answer the following questions:

- How can one find the best keywords to use, and where should they ideally be placed on the web page?
- How important is the relation between Moz-scores and website ranking?

- Can we increase the rank of a real website by following our keyword optimization method?

1.2 Outline

In chapter 2 a number of important background aspects of the subject are explained: search engine optimization and its different parts, the evolution of the Google search algorithm and an explanation of the long tail. Chapter 3 presents the practical part of the work, which includes the gathering of keywords and website data, as well as the rewriting of website content. Chapter 4 will present the data gained from the previous research. The results will be analyzed in chapter 5, where the discussion of the different parts take place. The thesis then comes to its conclusion in chapter 6.

2 Background

2.1 Search Engine Optimization

Search engine optimization (SEO) is, in short, the various methods you use to make a website more visible in a search engine. There are many ways of achieving this, as well as many different aspects to keep in mind; everything from code structure and link networks to quality content. There are a number of typical best practices that should generally be followed, as well as more advanced details to keep in mind. The following two sections will briefly cover both parts, as well as describing the importance of keywords. The last section will cover the tools used in the later parts of the thesis.

2.1.1 Best Practices Overview

Google themselves share some essential guidelines for how optimization should be performed (Google, 2010) to gain high ranks. While it is not particularly in-depth, it does work as a proper guide for anyone developing a website or want an overview of the subject. Furthermore, the SEO company iProspect who has helped with this thesis, has their best practices which give more weight and perspective to the already available information.

First of all, what might be the most important factor to bear in mind is that the best optimization is done for the person behind the computer, and not solely for the search engine, meaning that it should be easy to understand and navigate, as well as look good. But to go over more technical aspects: the content on pages of the website should have good titles and tags, both for users and for Google; furthermore, the description meta tags provide detailed information about each page. URL's should be descriptive instead of a random sequence of characters to make things clear for the users. Good navigation techniques can also be helpful, for example breadcrumbs to easily revert to previous pages. Regarding sitemaps, there should be separate ones for users and search engines.

Links should not be hidden inside codes such as Flash or JavaScript, since it can possibly cause problems when different devices such as mobile phones try to handle these, as well as interfere with search engines. Also, organic links are usually the most important in contrast to paid ones. Keywords should be nicely mixed to attract different users searching with different queries, while wanting to reach the same information (synonyms and related keywords are therefore important). Links should have anchor text, and images

need metadata to make them reachable by crawlers. A robots.txt-file is also needed for crawlers like GoogleBot, to tell them which pages contain useful information. “Nofollow” can be used to counter spam and links to shady websites, since reputation and reliable links are some of the most important things in SEO.

2.1.2 Keywords and Site Data

“By researching the market’s keyword demand, you learn which terms to use in the content, target for SEO, but also more about your audience as a whole”, quoted from the iProspect keyword-analysis slide (La Starza, 2015). It emphasizes a focus on users rather than digging into the details when it comes to keyword values. Rather than the highest search volume-keywords, one must tailor the research according to relevance of the current page and the need of its users. Google’s autocomplete gives good information regarding the frequency of certain keywords, and some handy software tools take advantage of this by extracting them, as well as Google’s own Keyword Planner, which is a part of Google AdWords, a service offering statistics to help with advertisement and search patterns. The ideal main keyword placement is, according to iProspect, in the title, meta description, main heading, and body text. In addition to this, the supporting keywords should be placed in the sub headings and body text.

Considering that over 40 000 Google searches are performed per second, as well as 3.5 billion every day (Internetlivestats.com, 2016) in plenty of different subjects, it is important to map the groups’ search behaviour in regards of the specific subjects. What journey is performed from the point of getting familiar with a product or service to the point of conversion? After search behaviour is mapped it is time to also look at the queries and needs that exist and how they are recieved by visitors. For this part it is highly relevant to look at the various factors that measure the traffic and value of your site.

When sites develop during the years some pages might become outdated or duplicate, which gets in the way of the Google rank since it will confuse which site is more relevant than the other. It is, in other words, very important to keep every page fresh. Keeping the pages fresh is important for plenty of other reasons as well, for example when bringing out the full potential of some pages that does not recieve the rank it might deserve. Review, merge, delete and maintain when needed. Filling up content where there is room can be very effective when the major fixes has been performed, by performing the same search journey as earlier, perhaps the search need has changed or the user behaviours has been updated.

2.1.3 Advanced Factors

Since the previous part of the thesis scrapes the surface and summarizes practical applications of SEO, it might also be interesting to dive into the details. The well-established SEO site Moz.com, which also provides SEO tools used in this thesis, goes over such detailed information about both basics and technical aspects in their SEO guide (Fishkin and Moz, 2015). If one

truly wants to understand the subject, it might prove useful to dig deeper into technicalities. This section will cover many interesting and relevant details.

Quality content is important to make users stay or return to your site. This can be measured by looking at engagement metrics, which measures if users stay on a site or click the back button right away. Also, sites do get a natural high ranking due to the Google Panda update (see section 2.2), so it might be a proof on its own that the site provides good content. Also, the link pattern for the site can give good hints regarding the quality of the site, if good sites link to it, and it links to good sites, it probably is a good site on its own as well. Search intents is an important detail regarding the users search patterns. It includes three categories:

- Transactional Searches: The need to find local business or shop online straight away.
- Navigational Searches: People who already know what particular URL they are looking for.
- Informational Searches: Users who use search to gain information about a subject or a person, for example in the middle of a discussion.

Link signals are very important to SEO. The global popularity of a site gives information about their popularity and the quality of the links. There is also a topic-specific popularity, meaning that sites within a local topic matters more to each others. The overall quality of keywords in the anchor text can have an impact on the link value of a site. Rather than looking at a specific site regarding links, the most effective thing might be to look at the whole link neighborhood, meaning all of the in- and out links. A sign of quality is keeping the site and links maintained and fresh, and an extra bonus might be social sharing, which are quite different from normal links but bear a good value on its own.

How the links are built can vary. The natural (editorial) links are those naturally given by other sites because you offer quality content. One might also do it manually (outreach), where the links are spread via email and so on. A final option is the self-created, non-editorial link building, in which you place links in forum signatures, comments and various online profiles.

Search engine protocols has been previously mentioned, but are crucial. Sitemaps are the things that gives hints to the crawlers, and they can be in various formats such as XML, RSS and .txt. An important protocol is the Robots.txt, the file that allows or disallows crawling of the pages on the site. It has a number of various commands to help the crawlers. Protocols might also be simple code pieces, such as: Rel= "NoFollow" and Rel = "canonical", which tells the search engines which pages should be skipped and which ones should count.

2.1.4 SEO Tools

There are a large number of tools used for SEO regarding keywords. The ones presented in this section will be used in the practical part and has been recommended by iProspect themselves. "Keywordtool.io" offers a tool that provides suggestions for related keywords once a particular word has been inserted, many of them being long-tail words. This can be used to find out which words one can use on a website regarding a specific subject.

Another very popular tool is provided by Google themselves, and is called Keyword Planner, which as mentioned in section 2.1.1 is a part of their AdWord service. It is an SEO tool that provides relevant data for inserted keywords, such as "competition", which according to Google themselves mean: "The number of advertisers that showed on each keyword relative to all keywords across Google. Note that this data is specific to the location and Search Network targeting that you've selected. In the "Competition" column, you can see whether the competition for a keyword is low, medium, or high." (Google, 2016). It can be used to figure out which words are worth working with, since some are harder to include due to competition of advertisers. A third tool which will be used in this thesis is a tool from Moz.com (an SEO website which was mentioned in section 2.1.3) known as the "Keyword Explorer" which features data about:

- Search Volume: The amount of Google Searches per month.
- Difficulty: How hard it is to achieve a higher rank than the competitors on the first page of Google search.
- Opportunity: The relative click-through rate of organic web results for the keyword.
- Potential: A combined metric for finding a sweet spot between search volume, opportunity and difficulty.

Moz also has another good tool called the On-Page Grader. In this tool the user can insert a keyword and an URL, and get an overall grade for the site as a whole and the keyword usage. It analyses if the keyword can be found in the title, URL, meta description, H1, body and img alt positions. In addition to these metrics, it also provides data regarding other contributing factors such as:

- Accessibility to search engines
- Keyword Stuffing
- Tags
- Content
- Links

- Metadata

When the URL and keyword has been inserted, the tool gives the website a grade ranging from A-F, in which A fulfills all or most of the conditions while a site with an F-grade requires much optimization. It is possible to gain decent grade without actually including the keywords, since many of their conditions can be fulfilled anyways, but to gain a higher rank the keyword must be found in the relevant sections previously mentioned.

Regarding actual website ranking for various keywords, the site "web-position.com" can be used, to which you send the website and the keywords that are to be tested for their respective Google rank. This website outputs the highest ranks that the website currently has for these keywords, and on which page on the website that the keywords are ranking.

2.2 The Google Algorithm

Google has gone through some massive changes since its initial build and release, and in the subject of search engine optimization (SEO) one has to take every major change into account, which can be found on the Wikipedia page for the Google Search timeline (Wikipedia, 2016a), further explained on Moz.com (Moz, 2016). In the early years (1996 – 2002) the engine was released, internationalized and received some extra features such as Google Books and Google Scholar. Soon thereafter, in 2002, announcements were made by Google regarding updates to the search algorithm, and it has been ever changing since then. After this, a number of major improvements were added to the engine. However, from the perspective of SEO, the more recent updates might be of bigger interest.

In 2011, the Google Panda Update was released which works like a filter against poor quality sites such as those including spam, instead improving organic search. This was of course important for SEO, since keyword stuffing and similar techniques would thereafter lead to negative results. Another attempt to stop poor quality sites was done with the Google Penguin Update; its primary purpose is to work against sites with paid-for links or artificial link networks created solely to boost the Google rank. Google Hummingbird was released in 2013 and is a search platform that improves semantic search, since it pays attention to the meaning of the words in a query rather than just the words themselves. Some more recent updates are called Google Pigeon, improving local search, and the Mobile Friendly Update, offering better support for mobile search. In addition to all these major updates, smaller ones are constantly being made, and new components get added. One new component is called RankBrain (DeMers, 2015), and while it might not have a huge impact at this point, it likely will in the future. It is, in summary, a signal that semantically interprets queries based on previous knowledge, even if it does not "understand" it. With these things in mind, it might be more clear how and why search engine optimization is done. As the search engines change, SEO too

must change and adapt.

2.3 Long Tail Searches

2.3.1 Defining the Long Tail

The way that customers act has changed drastically since Internet commerce made its way into our homes, as mentioned in the book "Long Tail" (Anderson, 2006). As the typical "top hits" of sales has lost some of its impact, the concept of mass culture as a whole becomes less relevant. Today, the trend points toward a need for variety and options, as services offering many different options have become more popular. The old mass market has been shattered, and commerce is left with a large amount of fragments represented as niches. The old "80/20"-rule has changed (20 percent of products stand for 80 percent of sales), as almost any product can be relevant these days. Since the sales are done through the Internet, keeping more obscure products around won't hurt the provider and waste its resources, while there is a high likelihood that someone, somewhere will eventually get interested. And when that happens it's a win for both parts. This is the reason why companies like Amazon and Netflix can reach such great success; their supply can be relevant for such a varying amount of people without burdening the companies. In other words: Internet decreases inventory costs, making it easier to reach out to everyone.

The long tail takes its name from a distribution graph, in which the head are the few dominant cases and the long tail is the enormous world outside of the hits, with smaller numbers on the y-axis but larger numbers on the x-axis. (Wikipedia, 2016b)

2.3.2 User Patterns

One interesting way of handling the long tail is by looking at Twitter (Agarwal et al., 2012). When an event has occurred, many people will tweet about this certain thing which in turn makes people curious. If a website related to this topic acts quickly, it might be able to optimize for these events and lead users to their site. Since this can lead to your site having rather unique key phrases, it might be an effective method of optimizing for long tail queries. Something interesting that is found in another study is that 95 percent of clicks are on organic links (Jerath et al., 2014). However, for the more specific and unusual keywords the sponsored ones gain a bit higher rate of clicks, since these are users who are looking for something specific rather than just browsing around. This hints towards an advantage of using sponsored links for long tail keywords. The advantage of the long tail is further strengthened when looking at different channels (Brynjolfsson et al., 2011). The catalog channels (physical sales) are compared to those of the Internet channel (web sales) and concludes that niche products do indeed increase sales on the Internet more than they would do in catalog channels, which further points towards a promising future of long tail optimization. A final interesting theme for long tail studies is that of travel destinations, which has been previously studied (Pan and Li, 2011).

When people are trying to put an image of a destination into words it follows a power-law distribution. There are a few places/phrases that are well known and often searched for, while simultaneously a large amount of niched phrases are used to search for less well known places, these searches are smaller in individual numbers but together account for a large chunk of the search volume.

It is worth noting that not all platforms are long tail markets. One study looks at Google Play (Zhong and Michahelles, 2013), the market for browsing applications for Android units, which concludes that the hits are highly dominant, where the long tail part of the market gets overshadowed. This proves how important it is to know your users when optimizing, the long tail is not always as relevant.

3 Method

For the practical part of this thesis, the goal is to find keywords related to a subject and inserting them on the website owned by iProspect. Furthermore, data will be gathered to analyze how well the Moz tools works for optimization regarding the relation between keywords and website ranking. In other words: this section will cover how to select good keywords and how to ideally use them to increase the Google rank.

3.1 Keyword Analysis

The first step to take was keyword analysis. This was performed by inserting a word from a particular subject into "keywordtool.io" and in this case the subject chosen was search engine optimization since it is what iProspect works with (meaning they have resources and a website fitting to the subject), and it also fits the overall theme of the thesis. The words inserted into the website was "sökmotoroptimering" (search engine optimization in swedish) and its abbreviation "SEO". When the most relevant words had been extracted, they were saved and then inserted into another tool, the Google Keyword Planner, to compare keywords regarding their competition with advertisers (in which a low competition is desired). The words were saved into an Excel-sheet, as well as its data regarding competition.

A file from iProspect that includes more words related to SEO was used to repeat the previous process of gaining suggestions and data. This was mainly done to gather a larger quantity to experiment with regarding the impact of good keyword placement (where the keywords should be ideally placed to gain a high grade in Moz), especially long tail-words with smaller search volume. A search distribution graph was created to gain an overview of the search volume and the keywords, to more clearly see the long tail among these words.

The most relevant keywords were hand-picked by selecting the ones with a lower competition in the Google Keyword Planner so that they could more easily be used on the iProspect website without having to compete with too many advertisers. These words were further narrowed down by using the Keyword Explorer from Moz, where the keywords with low difficulty and high opportunity was chosen to avoid further competition. By looking at these metrics, one could find the most promising of the keywords to use on the website. At this point two different documents were kept: one for the iProspect

website, with a few chosen keywords, and one for the larger amount of data for further quantitative research. The total number of keywords at this point was 114.

3.2 On-page Analysis

To gain information about the impact of the selected keywords, the method for the next step was an on-page analysis which aims to analyze websites that gain a high rank for these keywords. The tool used for this task was the On-Page Grader provided by Moz. The few chosen keywords that were gathered in the narrowing-phase was inserted into the grader along with the site URL, to see their current grade. The next step was gathering the top-2 ranked websites for each of the 114 keywords, so that a comparison could be performed regarding their grades and how important they seem to be for the specific keywords. In other words, the total number of A-grades for the highest ranked websites was compared with the number of A-grades for the second-highest ranked websites. By extracting the data provided by Moz regarding the keyword placement in the title, URL, meta description, H1, body and img alt, the A-grades could be compared to the lower grades to see where the keywords appeared to have the highest impact for this grade. Additional pages with much lower Google rank was tested to see if a more significant difference could be found.

For all of the 114 keywords and top-ranked websites (rank 1), the keyword scores were saved for each grade. This was done to see how keywords were usually placed for A-grade websites, B-grades, and so on. By saving these scores, the average keyword scores for each of these grades could be calculated. It was performed with the AVERAGE-command in Excel, example:

=AVERAGE(H1 for A-grade Site 1, H1 for A-grade Site 2, H1 for A-grade site 3,).

=AVERAGE(title for B-grade Site 1, title for B-grade Site 2, title for B-grade site 3,).

These average scores can in turn be compared in between the grades. Example:

Average A-grade H1 – Average B-grade H1 = Difference.

Average A-grade title – Average B-grade title = Difference.

By doing this, one can find out in which keyword position the score vary the most and where it appears to be most significant. Since there are a lot of 0's in the scores, a smoothing of 0.1 was added to get scores that are better for comparison. 0.1 was in other words added to all of the keyword scores, since 0's are not desirable to use for calculations. With smoothing (0.1) added, the 0's can bear a small score on their own since they become 0.1 instead, and 1 will become 1.1 and so on, which in turn will add more decimals which is

useful when calculating with a large amount of scores containing many 0's, to find the small differences. It does, in other words, allow for better comparisons with more numbers.

Some sites and grades were discarded due to the Moz-grading not working on pdf-files (which in a few cases were the highest ranked ones). For the keyword "konverteringsoptimering" some additional tests were performed on lower ranked websites to see if more differences could be found on lower ranks. A further attempt at testing how Moz grades is by testing a nonsense keyword on some of the websites with high grades, to prove that the Moz grade does indeed rely on proper keyword placement.

The iProspect website was sent to webposition.com as well as the chosen keywords, to gain a report containing the current ranks for the keywords on the website. The site was then modified with these things in mind, to see if any significant changes could be seen in its page rank over time. Important headlines were moved further up on the site, and the content was rewritten to include the chosen keywords, since we had statistics regarding which keywords were good to include. This means that some content on the site was removed, and completely new content was written related to the keywords, while some text was mostly moved. After one week, the same process was performed. The keywords along with the website was sent after the content had been edited.

4 Results

4.1 Keyword Data

When choosing keywords for iProspect, those with an average search volume per month below a threshold of 400 were specifically picked, since they can be categorized as long tail (less popular than the most searched for keywords in the subject), compared to words like "SEO" that has a search volume of 4400 per month (Such words are naturally included on the site anyway). Also, the most suitable ones were hand-picked as described in section 3.1. Figure 4.1 displays the keywords that were chosen for iProspect since they fitted their website and had a reasonable competition below a threshold of 0.50 (some additional keywords were also added simply because they were relevant for the website, although they had a bigger competition).

Keyword	Avg. Monthly Searches (exact match only)	Competition
seo byrå	320	0.31
seo konsult	90	0.44
sökmotoroptimering konsult	10	0.41
seo jobb	50	0.34
seo specialist	260	0.15
seo sökmotoroptimering	20	0.56
sökmotoroptimering	5400	0.53
vad är seo	90	0.24
sökmotoroptimering utbildning	30	0.58
seo	4400	0.43
seo vad är det	10	0.19
seo optimering	170	0.60
vad är sökmotoroptimering	30	0.25
seo kurs	110	0.41
jobba med sökmotoroptimering	10	0.00
sökmotorer	4400	0.13
sökmotor	1300	0.13
nulägesanalys	720	0.00
digital marknadsföring	590	0.61
sökoptimering	480	0.62
sökordsoptimering	480	0.50
sökord	390	0.08
nyckelord	390	0.00

Figure 4.1: The chosen keywords for iProspect

Additionally, many more related keywords were included (total list of keywords were 114) for the follow-up experiment, but didn't quite fit the particular web page. The reason for this is that these keywords, while relevant to the subject, does not have any room on the chosen iProspect website page and might instead find their place in blog posts or other pages on the website (focus lies only on on of their pages and not all of them). A sample of such keywords:

seo pr
seo marknadsföring
sökannonsering
seo bloggar
jobba med sök
seo tips 2016
seo javascript
bäst på sökoptimering
seo tekniker

When demonstrating search volumes and keywords, graphs like the one below is often used. On the y-axis the amount of average searches per month is displayed, and the keywords are placed on the x-axis. Note that all of the 114 keywords are used in the graph, although not all can fit in the image. In this case the long tail is assumed to start from the word "webbanalys", since the more unusual queries are represented after it, as displayed in figure 4.2:

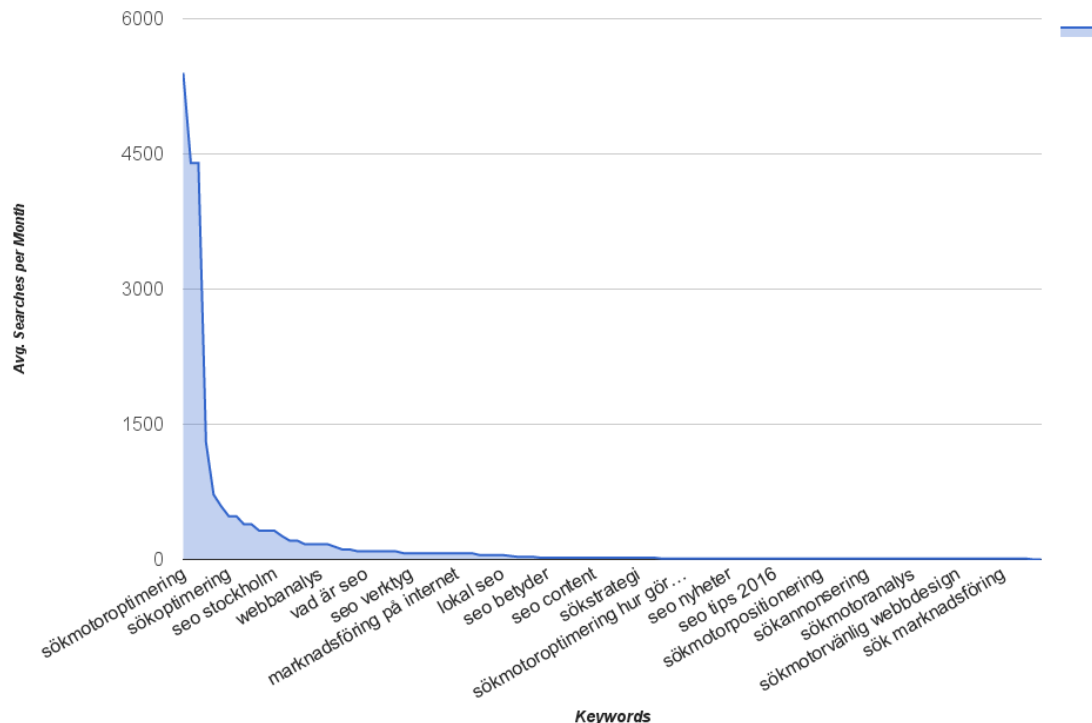


Figure 4.2: Search distribution graph of extracted keywords

4.2 Website Data

Among the top 1-ranked sites for all of the 114 keywords, the following Moz grade distribution could be found:

A-grade: 27
B-grade: 27
C-grade: 33
D-grade: 25
E-grade: 0
F-grade: 0

The top 2-ranked site Moz grade distribution:

A-grade: 31
B-grade: 25
C-grade: 36
D-grade: 17
E-grade: 0
F-grade: 1

Out of 114 websites on rank 1 for their respective keywords, only 27 had an A-grade. In other words, one does not always have to achieve an A-grade to be top ranked. This likely differs from keyword to keyword, since some will have a higher competition in between websites, if the highest ranked site has a B-grade, however, it might be desirable to strive for an A-grade to rise above it. It is apparent that other factors can have a significant impact as well since the second-highest ranked websites apparently has a few more A-grades than the ones on rank 1. Still, by saving the keyword scores for the highest ranked sites, one can compare them in between grades to find patterns to follow to increase the grades in regards of keyword placement.

By keeping 4 different Excel sheets, the average keyword placements for grades A-D can be calculated. In other words, in a sheet calculating average A-grades keyword scores, the average title, URL, meta description, H1, body and img alt is calculated so that it can be compared with the same calculations for B-grade keyword placement, which in turn can be compared with average C-grade and also D-grade. Figure 4.3 displays a sample of the A-grade sheet, and figure 4.4 displays the calculations for differences in average keyword score between grade A-B, B-C and C-D.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1														
2		Title	Title	Title	Title	Title	Title	Title	Title	Title	Title	Title	Title	Title
3		1.1	1.1	1.1	2.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1
4		URL	URL	URL	URL	URL	URL	URL	URL	URL	URL	URL	URL	URL
5		1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	0.1	1.1	1.1	1.1	1.1
6		Meta Description	Meta Description	Meta Description	Meta Description	Meta Description	Meta Description	Meta Description	Meta Description	Meta Description	Meta Description	Meta Description	Meta Description	Meta Description
7		0.1	0.1	1.1	2.1	0.1	2.1	2.1	0.1	0.1	1.1	0.1	2.1	0.1
8		H1	H1	H1	H1	H1	H1	H1	H1	H1	H1	H1	H1	H1
9		1.1	1.1	1.1	1.1	0.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1	1.1
10		Body	Body	Body	Body	Body	Body	Body	Body	Body	Body	Body	Body	Body
11		5.1	11.1	8.1	8.1	13.1	5.1	6.1	10.1	0.1	7.1	5.1	4.1	2.1
12		IMG ALT	IMG ALT	IMG ALT	IMG ALT	IMG ALT	IMG ALT	IMG ALT	IMG ALT	IMG ALT	IMG ALT	IMG ALT	IMG ALT	IMG ALT
13		0.1	0.1	1.1	0.1	2.1	1.1	0.1	0.1	0.1	0.1	1.1	1.1	0.1
14		A	A	A	A	A	A	A	A	A	A	A	A	A
15		Title	Title	Title	Average Title									
16		1.1	1.1	1.1	1.137037037									
17		URL	URL	URL	Average URL									
18		0.1	1.1	1.1	0.9148148148									
19		Meta Description	Meta Description	Meta Description	Average Meta									
20		1.1	0.1	1.1	0.7666666667									
21		H1	H1	H1	Average H1									
22		1.1	1.1	1.1	0.9518518519									
23		Body	Body	Body	Average Body									
24		3.1	5.1	4.1	5.544444444									
25		IMG ALT	IMG ALT	IMG ALT	Average IMG									
26		1.1	0.1	0.1	0.4333333333									
27		A	A	A	Average A									

Figure 4.3: A sample from a sheet with Moz-scores for A-grade sites and its average scores

Average A	Average B	Average C	Average D
Average Title	Average Title	Average Title	Average Title
1.137037037	1.025925926	0.3142857143	0.1
Average URL	Average URL	Average URL	Average URL
0.9148148148	0.7296296296	0.3857142857	0.1
Average Meta	Average Meta	Average Meta	Average Meta
0.7666666667	0.6925925926	0.2428571429	0.1
Average H1	Average H1	Average H1	Average H1
0.9518518519	0.8407407407	0.2071428571	0.1
Average Body	Average Body	Average Body	Average Body
5.544444444	18.72962963	1.885714286	1.1
Average IMG	Average IMG	Average IMG	Average IMG
0.4333333333	0.4703703704	0.1357142857	0.1416666667

Figure 4.4: The differences of average Moz scores in between grades

The most relevant differences are the ones between average A-grades and average B-grades, since those with a lower grade might have other optimization issues to solve before focusing on the keywords. The URL, title and H1 are more significant than the body, image alt and meta keywords in this case. In other words, these are the positions in which keyword inclusion appears most important, while the body text had a negative score, meaning it might even be negative to include a keyword too many times in the body text

(although this might not be true and likely varies from case to case). At the very least, the body seems to be the least important part to include keywords in. The URL had the biggest impact, so it seems important to include keywords in the URL when one wants to achieve a higher grade. These tendencies seem to differ greatly between the different grade comparisons, but assuming one wants to achieve an A-grade from having a B-grade, editing the URL seems to be a good start. After this has been corrected, the title and H1 appears to be the next step to take. Something typical for D-grades is the lack of keywords in any of the positions, so those appear to simply be websites that have optimized correctly in other ways and not for the particular keywords. Img alts seem to be consistently unimportant to the Moz grades. In conclusion, most of the average scores are decreasing from A- to D-grades so the numbers for the average A-grades should be the ones to follow.

For the 5 keywords below, the top ranked sites recieved an A-grade, so they were tested to see how they ranked without the keywords. In other words, the highest ranked websites for these keywords were tested for completely nonsense keywords, to see how high the Moz grade can be in absence of correct keyword placement (the keyword used was "hedgehog").

- seo Stockholm: C
- sökordoptimering: C
- webbstrategi: C
- konverteringsoptimering: C
- onlinemarknadsföring: C

In other words, the Moz On-Page grade does indeed give reliable information regarding the keyword placement, for the highest grades one must recieve good keyword scores. An even closer look at the keyword "konverteringsoptimering" proves that some keywords, like this one, are harder to compete with only based on the keyword placement:

From rank 1-12, 11 of them recieved an A-grade.
From rank 30-39, 5 of them recieved an A-grade.
From rank 50-59, 8 recieved an A-grade.
From rank 100-109, 8 recieved an A-grade.

Figure 4.5 includes data for the iProspect site rank for the keywords at that particular date (the page to be edited is the one called "<http://www.iprospect.com/sv/vara-tjanster/seo/>").

Keyword	Avg. Monthly Searches (exact match only)	Competition	Ranking 20160512	Sida som rankar
seo byrå		320 0.31	6	http://www.iprospect.com/sv/vara-tjanster/seo/
seo konsult		90 0.44	6	http://www.iprospect.com/sv/vara-tjanster/seo/
sökmotoroptimering konsult		10 0.41	6	http://www.iprospect.com/sv/vara-tjanster/seo/
seo jobb		50 0.34	10	http://www.iprospect.com/sv/se/careers/job-list/s
seo specialist		260 0.15	17	http://www.iprospect.com/sv/se/careers/job-list/s
seo sökmotoroptimering		20 0.56	26	http://www.iprospect.com/sv/vara-tjanster/seo/
sökmotoroptimering		5400 0.53	42	http://www.iprospect.com/sv/vara-tjanster/seo/
vad är seo		90 0.24	51	http://www.iprospect.com/sv/vara-tjanster/seo/
sökmotoroptimering utbildning		30 0.58	53	http://www.iprospect.com/sv/vara-tjanster/seo/
seo		4400 0.43	71	http://www.iprospect.com/sv/vara-tjanster/seo/
seo vad är det		10 0.19	82	http://www.iprospect.com/sv/vara-tjanster/seo/
seo optimering		170 0.60	88	http://www.iprospect.com/da/dk/
vad är sökmotoroptimering		30 0.25	90	http://www.iprospect.com/sv/vara-tjanster/seo/
seo kurs		110 0.41	91	http://www.iprospect.com/no/no/events/
jobba med sökmotoroptimering		10 0.00	92	http://www.iprospect.com/sv/vara-tjanster/seo/
sökmotorer		4400 0.13	#N/A	#N/A
sökmotor		1300 0.13	#N/A	#N/A
nulägesanalys		720 0.00	#N/A	#N/A
digital marknadsföring		590 0.61	#N/A	#N/A

Figure 4.5: Some of the keywords and their ranks for the iProspect website

It shows that out of the 114 keywords, only 15 of them had an actual rank at this point, the highest being rank 6. The other 99 keywords did not receive any rank at all, so there is plenty of room for improvement by following the guidelines that the previous results created. Note that not all of these keywords are supposed to be included on the website, only a few chosen ones (see Figure 4.1). For most of those keywords, the iProspect website aims to increase its rank. After one week had passed since the content was edited, the results from webposition had slightly changed (see figure 4.6).

SEO konsult	http://www.iprospect.com/sv/vara-tjanster/seo/	Primary	6
seo byrå	http://www.iprospect.com/sv/vara-tjanster/seo/	Primary	6
seo konsulter	http://www.iprospect.com/sv/vara-tjanster/seo/	Primary	6
organisk sökmotoroptimering	http://www.iprospect.com/sv/vara-tjanster/seo/	Primary	7
sökmotoroptimering konsult	http://www.iprospect.com/sv/vara-tjanster/seo/	Primary	7
seo jobb	http://www.iprospect.com/sv/se/careers/job-list/seo-konsult	Primary	9
seo sökmotoroptimering	http://www.iprospect.com/sv/vara-tjanster/seo/	Primary	24
sökmotoroptimering seo	http://www.iprospect.com/sv/vara-tjanster/seo/	Primary	27
Sök marknadsföring	http://www.iprospect.com/sv/se/vara-tjanster/sokmarknadsf	Primary	28
sömarknadsföring	http://www.iprospect.com/sv/se/vara-tjanster/sokmarknadsf	Primary	29
sömotoroptimering	http://www.iprospect.com/sv/vara-tjanster/seo/	Primary	43
varför sömotoroptimering	http://www.iprospect.com/sv/vara-tjanster/seo/	Primary	50
Seo optimering	http://www.iprospect.com/da/dk/	Primary	82
seo tjänster	http://www.iprospect.com/sv/se/careers/job-list/seo-konsult	Primary	95
Keyword	URL	URL Type	Position

Figure 4.6: The new ranks of the keywords on the iProspect website, after one week

5 Discussion

5.1 Keyword Phase

How one might find good keywords is tested in this thesis. In this particular case there were some limits on the keyword research. First of all, the subject of search engine optimization is quite narrow as compared to, for example, "shoes" or "cars". Reliable data is harder to get and the overall search volume is a bit small in comparison to bigger subjects. The fact that the keywords are Swedish is likely to have an impact on the work as well, since they will not have as big of a search volume as it would for a bigger subject in a bigger language. Many important keywords are likely left out as well, since there is such an enormous amount of relevant ones being searched for, but one simply has to stick to a certain limit. There are so many web searches everyday, in which a significant fraction are completely unique and harder to find in tools from Google and others. A software that can fetch keywords in real time from actual web searches or Twitter, for example, could possibly give more keywords that are relevant in the exact moment. There is also a possibility of the human factor affecting the gathering of keywords, some might be relevant even though they are overlooked. This is due to the fact that a large number of irrelevant keywords that simply look similar to the keyword from which the related ones are gathered also come up among the results in keywordtool.io and Google, so a lot of them do in fact have to be ignored, and some relevant ones might accidentally be ignored or are never found in the first place. Overall, long tail keywords are good to keep in mind since people do their searches in different manners and formulate them differently, especially considering that voice searches are becoming more popular and are likely to be unusually formulated as compared to text searches, which means that they might also belong to the long tail searches.

5.2 Website Phase

The same issue discussed in the previous section likely impacts the Moz ranks as well. The numbers might prove more reliable in a different subject, since SEO is a narrow subject as compared to bigger ones. Between the sites ranked 1 and 2 the difference was small, the sites on rank 2 actually had more A-grades than the rank 1-sites. In the case of keywords such as "konverteringsoptimering", there were consistent A-grades as far down the list as rank 100. This is an example of a keyword that is very hard to gain the highest Google rank for even with proper keyword optimization, and one might take advantage of the long tail method and use more unusual keywords instead, or as an addition.

But as always in SEO, there is a large number of factors that affect the results, which is also taken into consideration by the Moz-rank. It appears that having an overall good optimization of the website will give a high grade no matter the keyword, so proper keyword inclusion is only certain to give an edge in between two otherwise equally optimized webpages, otherwise a website with a less refined keyword optimization can still get a higher rank due to having ideal keyword placement in the links, metadata, and other factors mentioned in the background of the thesis. The usage of swedish keywords might have another problem with the Moz-score, since the URL's does not include the letters "ää" while the keywords often do, which makes it harder for certain keywords to display an A-grade in Moz. A final note regarding this aspect is that many of the tested sites are written by people that actually work with SEO, and as such they might lose a bit of relevance in regards of the reality of internet search and the optimization of top websites. So yes, the Moz grades do say quite a lot about how well you will potentially rank. This differs between keywords mainly because the competition in between websites is harder for some keywords (where there might be A-grades throughout many sites) than for others with more varying ranks, and in which only one or a few in the top receive the A-grade.

After one week, the rank had already increased for some of the keywords on the iProspect website, so the optimization was successful, but since ranked results take a significant amount of time to take effect many more changes are likely to appear in the future.

5.3 Future Research

This thesis gives results that are easy to build upon. By including a larger amount of keywords and websites, more reliable data could be gained. Additionally, as mentioned earlier, testing for different languages such as english might be preferable. Also, it might prove interesting to do the same tests for another subject to gain more results for keywords with larger search frequency. If a whole website is edited instead of a particular page, it might also give a lot more results since the edited page in this experiment could not include every keyword.

6 Conclusion

There is evidently some great tools around the Internet to use when you want to optimize your website. They make the process significantly easier and are proven to often be reliable. Both Google and Moz offers great solutions for doing this. They appear to be worth using as long as you understand all the factors involved; it is hard to simply separate the keyword aspect of SEO from the rest. While there are many different aspects that can help a site gain rank, once an overall well-optimized site would like to draw in more visitors, good keyword research and placement appears to pay off, particularly for long tail keywords.

By analyzing the potential of keywords and the competition from other websites that include these keywords, there are many hints towards which ones to use. Also, by using on-page graders there is a lot of available information regarding where the keywords should be placed to reach an ideal optimization and higher rank. The keywords should not only be focused on the body text but in positions such as headlines, titles and even the website URL's when possible, meaning that many different pages on the website focusing on many subjects might be the way to bring in the highest amount of users, since there is not room for every relevant keyword on one page alone. A way of including the highest amount of keywords possible might be through blog posts and articles, in which the keywords can easily be placed in URL's, instead of trying to stuff them all on one page. Search engine optimization requires long-term thinking and patience, since any rank statistics might take some time to change, and ideally the pages should be properly maintained and actively cared for, to always reflect the user needs based on their search behaviour. There appears to be an important relation between Moz scores and website ranking, however a high Moz score is not a standalone guarantee for a high website ranking, therefore it should mainly be used as a structural guideline when writing the content of the page. By following the method in this thesis some improvements and changes could be applied to a real website in an attempt to increase its rank. Since this experiment did yield some successful results in a short time, it does prove that editing the content to include relevant keywords in crucial positions does have the potential to give an improved website ranking for those particular keywords.

In summary, this thesis has investigated how much one can rely on some popular SEO tools when working with keyword optimization, and how to reach an increased website ranking by using them. Positive results were achieved, although some aspects might require further investigation. Overall the thesis does provide evidence for some important relations between keywords

and website ranking, particularly how optimizing for long tail keywords can prove to give a significant edge when the website competes with others of the same subject.

Bibliography

- Puneet Agarwal, Rajgopal Vaithiyathan, Saurabh Sharma, and Gautam Shroff. Catching the long-tail: Extracting local news events from twitter. In *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*, pages 379– 382, 2012.
- Chris Anderson. *The Long Tail: Why the Future of Business Is Selling Less of More*. Hyperion, 2006. ISBN 1401302378.
- Erik Brynjolfsson, Yu Hu, and Duncan Simester. Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management Science* 57.8 (2011): 1373-1386, 2011.
- Jayson DeMers. What is google rankbrain and why does it matter?, 2015. URL <http://www.forbes.com/sites/jaysondemers/2015/11/12/what-is-google-rankbrain-and-why-does-it-matter/#7e1cb5c301ae>.
- Rand Fishkin and Moz. The beginners guide to seo, 2015. URL <http://static.googleusercontent.com/media/www.google.com/en/webmasters/docs/search-engine-optimization-starter-guide.pdf>.
- Google. Search engine optimization starter guide, 2010. URL <http://static.googleusercontent.com/media/www.google.com/en/webmasters/docs/search-engine-optimization-starter-guide.pdf>.
- Google. Keyword planner: About search volume statistics, 2016. URL <https://support.google.com/adwords/answer/3022575?hl=en>.
- internetlivestats.com. Google search statistics, 2016. URL <http://www.internetlivestats.com/google-search-statistics/>.
- Kinshuk Jerath, Liye Ma, and Young-Hoon Park. Consumer click behavior at a search engine: The role of keyword popularity. Technical report, Marketing Science Institute Working Paper Series 2013, Report No. 13-106, 2014.
- Vincenza La Starza. Keyword research - inspiration for content, 2015.
- Moz. Google algorithm change history, 2016. URL <https://moz.com/google-algorithm-change>.
- Bing Pan and Xiang Li. The long tail of destination image and online marketing. Technical report, Annals of Tourism Research, 2011.
- Wikipedia. Timeline of google search, 2016a. URL https://en.wikipedia.org/wiki/Timeline_of_Google_Search.

Wikipedia. Long tail, 2016b. URL https://en.wikipedia.org/wiki/Long_tail.

Nan Zhong and Florian Michahelles. Google play is not a long tail market: An empirical analysis of app adoption on the google play app market. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, pages 499–504, 2013.