

**Department of Computer Science
and Mathematics
Munich University of Applied Sciences (MUAS)**

In collaboration with

**Website Attributes Research Centre
Cape Peninsula University of Technology (CPUT)**

The research as described in this document was done by the first author as noted below. She was a student at the Munich University of Applied Sciences (MUAS) in 2014 - the topic was given as a student project. It has not been peer reviewed, but has been edited for basic grammar and accuracy. References have been standardized as far as possible according to the Harvard system. Consider this document as a working paper, to be used for basic referencing but not as seminal source for research work. It could be useful for research in the fields of Website Visibility, Information Retrieval and Search Engines.

**A ranked summary of the on-page elements
affecting the visibility of a webpage to search
engines**

Author: Manuela Blümel (MUAS)

Student Number: 02657811

Mentor: Prof M Weideman (CPUT)

Submitted on 27 October 2014 Edited, published on 01 December 2014

Field: Computer Science

Topic: Website Visibility

Harvard Reference:

Blümel, M. & Weideman, M. 2014. A ranked summary of the on-page elements affecting the visibility of a webpage to search engines. *Working Paper*, Munich University of Applied Science, Munich. Available online: <http://www.web-visibility.co.za/website-visibility-abstracts-seo.htm>

Abstract

Search engine optimisation has become a popular topic of discussion, especially in the commercial world. There are different reasons why the owner of a website would want their site should to rank well in the search results. Different approaches are available to achieve high rankings: on the one hand the ethically correct whitehat method, and on the other hand the unethical blackhat approach. Because of the potential of incurring penalties and subsequent problems it is not recommendable to use a blackhat approach. This research only focussed on whitehat methods. The objective of this research was to determine which positive and negative on-page elements affect the visibility of a webpage most. An average ranking was compiled.

Website visibility elements were separated into those which have a positive and a negative impact on the visibility of a website respectively. The factors were divided into categories. The positive groups were keywords, internal links, content, external links, meta-tags and others. The negative groups included bad content, link farming, keyword stuffing, cloaking and advertisements. The content of several ranking models was split up into these groups to make them comparable. An average value for each group was calculated. Diagrams were generated by combining the models and calculating average values for the positive and negative on-page elements. These diagrams indicate which on-page element have the highest impact on the visibility of a website. The result was that good content, keywords, internal links, external links and meta-tags generate the greatest positive impact on the visibility of a webpage. In contrast, bad content, keyword stuffing, link farming, certain technologies, cloaking and advertising provided the greatest negative impact.

The ranking of on-page elements is not a straightforward calculation due to the fact that search engine algorithms are usually not public. The models are based on estimated values, and therefore the average of several models produced more reliable results. In conclusion it can be stated that no single on-page element have a major impact on the visibility of a website. Rather, the focus should be on a combination of factors.

Keywords

website visibility, internet, positive on-page elements, negative on-page elements, search result ranking, search engine

1. Introduction

There are many ways to improve the ranking of a website in the search results of a search engine (Curran 2004). From the viewpoint of the search engine, there are whitehat and blackhat methods for optimising a webpage. On the one hand the whitehat optimisation is an ethically correct method to increase the visibility of the website in the search results of a search engine.

On the other hand, blackhat methods concentrate on the search engines themselves and not the human audience of the website or the guidelines of search engines. One method could for example be to include keywords that are not used in any context on the website, or to use invisible text (words in the webpage code which are not visible for normal visitors but only for search engines). If the blackhat method is revealed by the search engine operators, in the worst case it can happen that the offending website is removed from the search engine index, and thus is no longer displayed in the search results. This exact event occurred and was reported on extensively at the beginning of 2006 to the website of BMW (Cutts 2006). It is not recommended to use blackhat methods, and therefore in this research the focus will be on whitehat methods.

In addition to whitehat methods, there is also the possibility to improve the ranking through paid advertising. There are, for example, the pay-per-click or pay-per-view models. The first model requires a payment per click and the second model requires a payment if the advertisement has been displayed. A website normally receives a high ranking for a list of keywords within the search results and the costs for that service depend on the chosen model and on the chosen keywords (Chen *et al* 2011).

It is possible to determine whether or not the optimisation process was successful. To this end, it is recommended to use appropriate analysis tools – refer to Section 2.2 for more details. These tools document historical values and show patterns in a graphical way.

Even if a website has achieved a high ranking through optimisation, this ranking can be affected by an algorithm update from the search engines. For example, Google has done various releases which affected some rankings negatively. These include Panda, Penguin and Pigeon. Both Panda and Penguin were

punitive in nature, since they dropped rankings when certain negative elements were detected in websites. These include thin content, link farming and other blackhat techniques. Furthermore there are also fluctuations in other elements of search engines (Bar-Ilan *et al* 2006). This could be attributed to the fact that the competition increases, and due to the re-evaluation of a website by a search engine.

In addition, it should also be noted that it is easier to achieve a better ranking for a website which is dealing with special topics and/or which is not written in English. The greater the number of competing websites, the more difficult it is to prevail, hence the strong competition between English websites. In turn, one can learn from the competition. The coding of many websites is open for inspection by any human visitor or crawler, and ideas from other websites can be used by any reader.

The focus of this research is to determine which on-page elements influence the ranking of a website in the search results of a search engine the most. The positive and negative effects of the on-page elements are listed in different rankings with average values.

2. Background and Other Research

2.1 Website visibility

In addition to the on-page elements, there are also a number of other ways to improve the ranking of a website in the search results of a search engine.

2.1.1 Backlinks

The more websites link to a webpage, the higher its relevance. These backlinks are also an indicator of how popular a website is. For this high relevance to pass onto a webpage, the following quality criteria must be met:

- The referencing websites must be reputable and well-known.
- In addition, the sites must match in content.
- The anchor text should fit thematically.

2.1.2 Social Media

Social media has become a factor to consider when building the visibility of a website (Killoran 2013). Through social media, it is possible to attract the attention of large numbers of users, or a specific target group. There are also many opportunities in the social media for indicating trust in or support for a given product, service, person or other entity. By using the "Like" button on Facebook, for example, or sharing the contribution, a website can be promoted by a person to their respective contacts.

In addition to the most relevant off-page optimisation, there are a variety of other off-page optimisation approaches that are not addressed in this research (SideOPS SEO Company n.d.).

2.2 Website Analysis Tools

With the help of analysis tools, such as Google Analytics, Alexa, Websiteoutlook and RankCheck, it can be determined how successful optimisation has been. These tools offer various forms of representation of data such as graphs, tables, maps and other options to create useful reports. Furthermore, a history of activity is created, and historical data can be displayed and analysed. At the time of writing, these tools are all free and available online.

2.3 Website usability

The usability and visibility of a website are related indirectly. If the degree of user-friendliness of a website is low, the bounce rate will be high, since not many users will visit other pages on the same website. If a website is user-friendly, it is likely that it will be recommended and thus gets known by many others. In addition, the user-friendliness of a website could be improved by on-page optimisation. Since on-page optimisation includes the focus on quality content, important usability criteria have already been met. Finally, some improvements of the architecture could reduce loading times and improve usability as well.

3. Impact of on-page elements on the ranking of a website

In this section the impact of on-page optimisation on the ranking of the website in the search results of a search engine is presented.

To obtain comparable values of various statistics, comparable groups were created. These groups were designed to cover most of the factors listed in the statistics. The groups include: keywords, internal links, external links, content in general and meta-tags. Then all factors that belong to a group were added (with their percentage as weight), and it was calculated which percentage of all factors of the specific statistic they have. Those comparable values were then inserted into a chart ordered by groups (including the average for each group), and a further chart was created, displaying the averages.

These steps are graphically displayed in Figure 1. Each factor with a blue arrow belongs to the group of external links. All these percentages were summed up and divided by the overall percentage. The result is the average percentage of the group external links for the model. These will also be calculated for the other models. In Figure 4 and 6 the actual results of these calculations are shown.

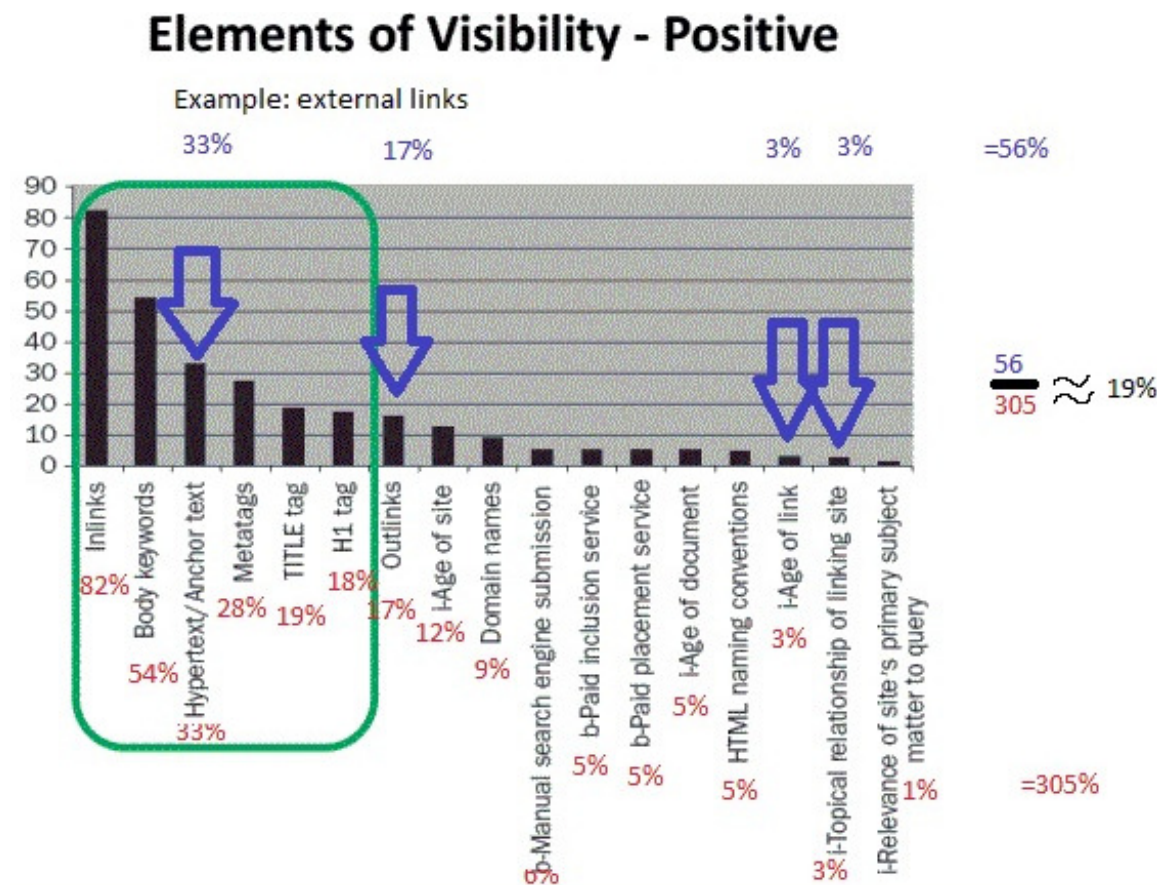


Figure 1: Example of calculating the average impact of external links on the website visibility (Weideman, 2009: 54).

The ranking models used here are listed in Appendix 2 to 5.

3.1 On-page elements with a positive impact

The Weideman and Sullivan models, and the Spearman correlation were used to calculate the average of all values with a positive impact. Figure 2 indicates the average impact of all relevant on-page elements on a site's ranking and its effect in the search results of a search engine.

Average weighting of the various positive on-page elements of a website's ranking in the search result list of a search engine

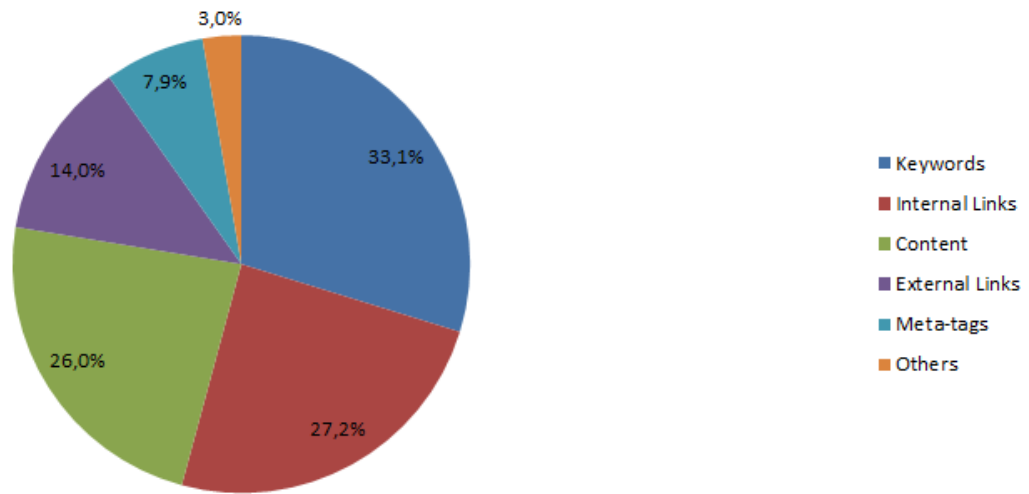


Figure 2: Average rating of positive on-page elements (Schwartz, 2014) (Sullivan, n.d.) (Weideman, 2009: 54).

Figure 2 shows that keywords (33.1%), internal links (27.2%), content (26.0%), external links (14%) and meta-tags (7.9%) could be considered as the most important factors.

3.1.1 Keywords

Keywords are important since search engines do not understand the content of a webpage, but can only scan and index it. In some cases the keywords of a search query can be directly matched to the keywords of a website. This would result in a higher ranking in the search results. The more prominent the keywords of a webpage are, the higher the probability that the webpage will be listed higher for the queries of a user in the search results. By optimising the keywords of a webpage for a specific topic, the probability is higher that a user will get this specific website as a search result when searching for something related to the topic (Velasquez 2012). Keywords are typically used in the URL, title, metatags, headings and body text of a webpage (Rimbach *et al* 2007).

3.1.2 Internal Links

Internal links are links that point to a page or area on the same website, possibly on a different webpage. The internal links are used by search engines to detect the structure of a webpage. Internal links do contribute towards higher rankings in the search results.

3.1.3 Content

Well-written content will cause users to recommend this website or link to other users. Through qualitative links a website can earn a higher ranking in the results of a search engine. It should be noted that the content and the complexity of the text should be aligned to the target group. Furthermore, the content should be well structured with headings. Good content also includes multimedia like videos or images. Meta-information for these files (title, alt title, description tag) should be provided as well. In addition, the content of the website should not contain spelling or grammar errors.

3.1.4 External Links

External links are those links on a webpage which refer to an external website. It is important to gain relevant and thematically appropriate links. If the reference provides more information, it also has a positive effect on the receiving website. Spamdexing links will be judged negatively by search engines and usually lead to a decrease in ranking of the relevant webpage. However, a webpage owner has little control over links being created on other websites, pointing to his/her webpages. Through constructive link building, these links can be built up over time.

3.1.5 Metatags

Metatags are HTML tags that contain information about the webpage where they appear. These metatags are not visible to the user, and are optional. However, they can become visible when a given webpage appears on a search engine result page. The most relevant metatags are:

- Title tag
- Description tag and
- Robot tag (index, follow).

In Figure 3 it becomes clear how the metadata title and description metatags are displayed in the search results.

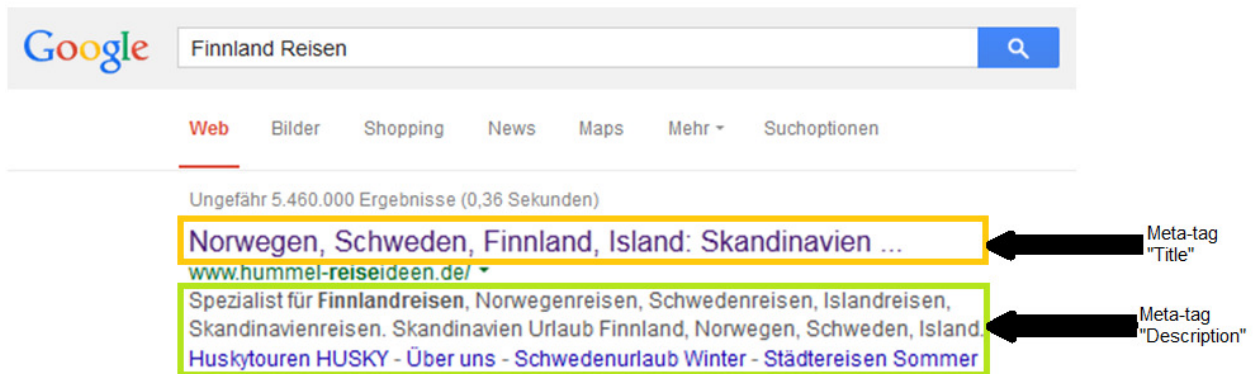


Figure 3: Example; Google - meta-information in search results (www.google.com).

If no metadata has been specified, the search engine uses any text from the website. This can cause the description appearing on the screen to be nonsensical, and it would not attract visitors (Goldsborough n.d.).

It is therefore important to choose a meaningful and concise title and description. Often, this opportunity is missed and instead the title is used as a greeting or it contains meaningless text. If the theme of the site can be summarised in a few words, a higher numbers of users can be attracted and a higher ranking can be achieved, because it increases the relevance of the website. Therefore it would be better to choose the title tag: "Lawyer John Doe - Probate Specialist Lawyer Munich" instead of "Welcome to the website of the law firm Smith". In order to increase the ranking of a website on search queries with location information, a geo-metatag can be used. It could for example contain the location of an office.

3.1.6 Other factors

There are other important on-page elements, which have a smaller impact on the ranking of the website in the search results of a search engine. Those include:

- Site download speed
- HTML naming conventions (Weideman 2009)
- URL format (Sullivan n.d.)

In addition, it should be mentioned that a low loading time of a website has a positive impact on the user experience (see Section 2.3.). Also, Google has included this factor as one of their official (approximately) 200 ranking factors.

By adhering to HTML conventions, it is easier for search engines to scan a webpage and it is more likely that the website is displayed in the same way on each browser.

It is also advisable to use a short, concise and meaningful URL. For the user it will be easier to enter a short URL when directly accessing a website. A short URL is more memorable for the visitors of a website and thus increases the likelihood that he/she will visit the page again.

Figure 4 indicates the average values of the level of impact of the on-page elements from the different models, when combined.

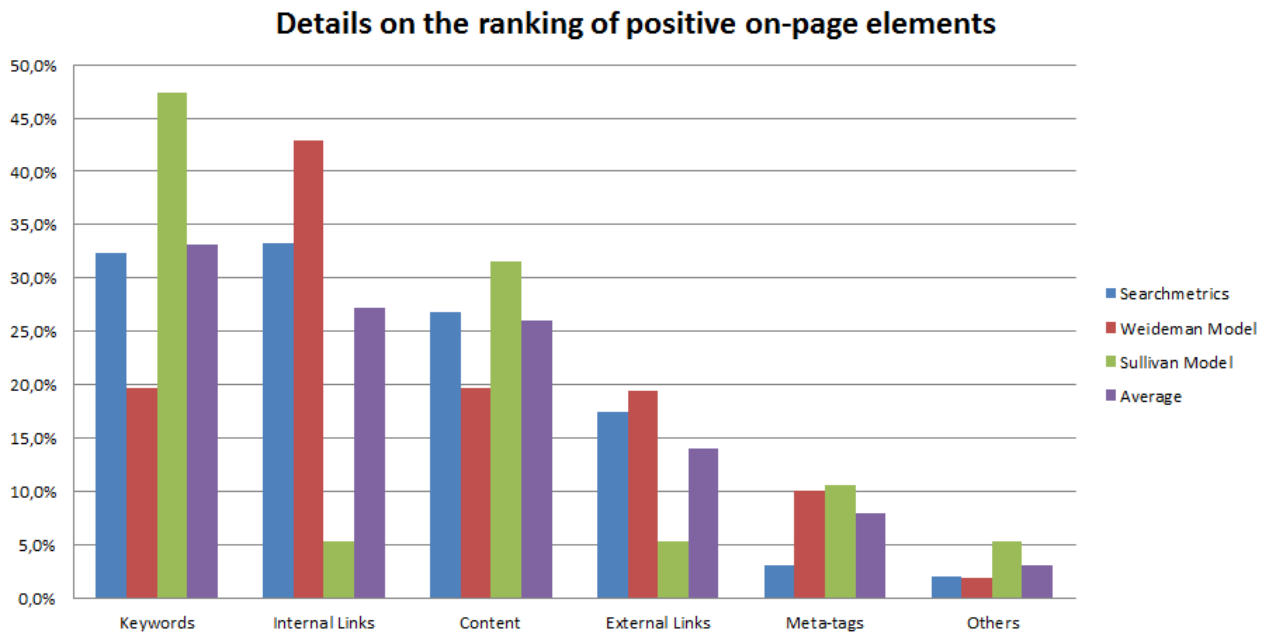


Figure 4: Comparison of the average values of the impact of positive on-page factors (Schwartz 2014) (Sullivan n.d.) (Weideman 2009).

3.2 On-page elements with a negative impact

For the survey, the average values with a negative impact on the ranking of a website (from both the Weideman and Sullivan model) were used. These on-page elements should not be used as part of website design. Figure 5 shows the average impact of the negative on-page elements on the ranking of a website.

Average weighting of the various negative on-page elements of a website's ranking in the search result list of a search engine

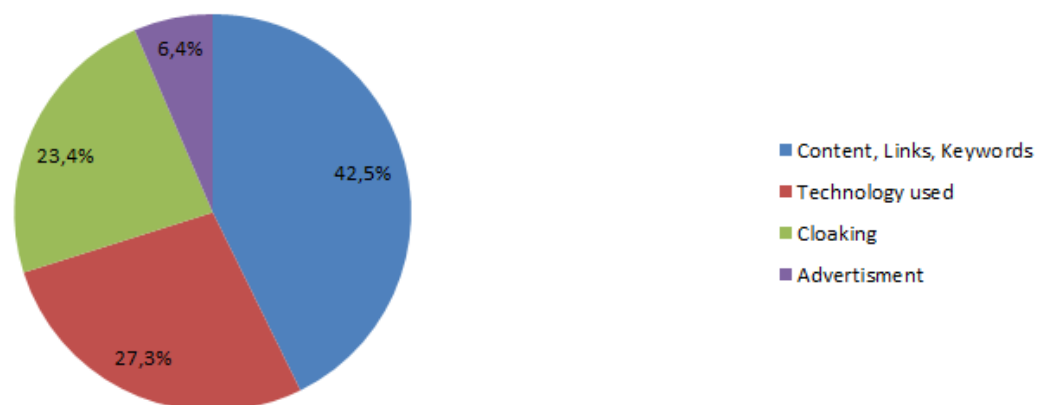


Figure 5: Average rating of negative on-page elements (Schwartz 2014) (Sullivan n.d.) (Weideman 2009).

Badly written content, link farming and keywords spamming combined earned a total of 42.5%, the highest negative impact on the ranking of a website. If the HTML tree is unclear, a user probably will not visit the website again (Garofalakis *et al* 1999). If the content of a website is thin or otherwise of a low quality, and the keyword density is high enough to become cumbersome to read, it would have a bad impact as well. The second biggest negative factor was the technology used, and frequent hardware problems of the server with an impact of 27.3%. These technologies are:

- Frequent server outages
- Flash
- Frames
- Low server response times
- Doorway pages
- Dynamic webpages
- JavaScript

(Weideman 2009)

The use of frames is a webpage feature which the user will experience as sections of the webpage remaining in place, while others move up and down. This has often been used in webpage design to, for example, fix the navigation pane of a webpage, while the users scroll up and down inside the content area.

However, search engine crawlers often see a single framed webpage as a multiple-page website. It can therefore happen that a visitor selects one of the results of a search engine result page, and only a single frameset is displayed. The use of frames can also negatively affect the way favourites in browsers are stored, and the way webpages are printed out.

A doorway page is an intermediate page that is search engine optimized and points to the actual webpage. The homepage of one website might have thousands of doorway pages with thin content, all linking to this homepage. Although this might fool a crawler into allocating a high inlink count to this homepage, users might be frustrated by the lack of useful content on the doorway pages.

A dynamic website displays content dynamically, often as the result of a user query on a (hidden) database. This dynamic page is not stored on a server like static webpages are, but exists only on the user screen until it is removed by the browser's next page request. As a result, dynamic webpages are not readable by search engine crawlers.

Well-written content is also crucial for the visibility of a website (see Section 3.1.). If parts of a website can only be reached through executing a piece of JavaScript code (for example, as used in some drop-down menus), it is often not possible for the search engines to index this content. Another reason for avoiding JavaScript is that it can be disabled in the browser, and a user in this case could not be able to access the website. With the help of a sitemap, proper indexing of websites could be done, bypassing this problem. Redirects through JavaScript will not be recognized.

At 23.4%, cloaking has a significant negative impact on the ranking of a website. Cloaking refers to different webpages being served (when requesting a given webpage), depending on the requestor. A human visitor will see a well-designed page, attractive to human users, while a search engine crawler will see and index a text-rich, often unattractive (for humans) webpage. This feature can also be used to lure the user to advertisements. A search engine can often not distinguish between positive or negative elements, since both are used in practice to present a highly visible webpage.

Finally it should be mentioned that a website that carries a high proportion of advertising is ranked lower, with an impact of 6.4%.

Figure 6 shows the composition of the average values of the individual models (Weideman plus the Sullivan Model).

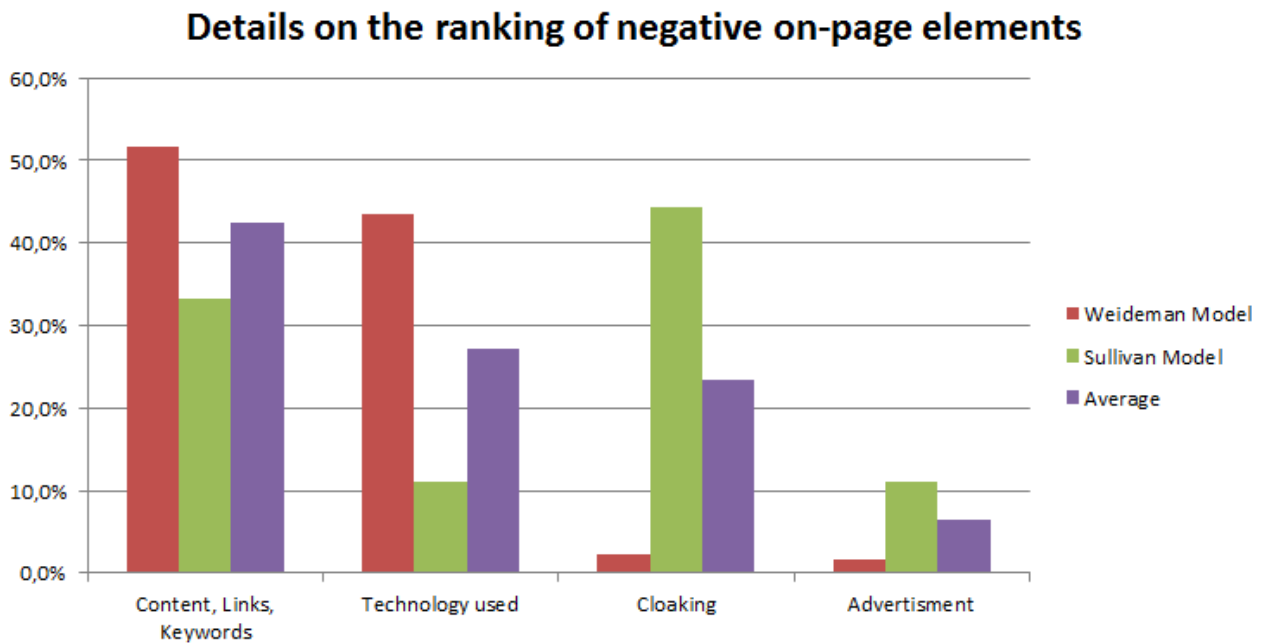


Figure 6: Composition of the average values of the impact of negative on-page factors (Schwartz 2014) (Sullivan n.d.) (Weideman 2009).

4. Interpretation and Conclusion

Creating a weighted listing of the impact of the on-page elements on a website's ranking in the search results of a search engine requires careful consideration. The reason for this is that the algorithm which is used for ranking by search engines remains a trade secret. It is not clear exactly how the on-page elements relate to each other and how each factor is evaluated by a search engine. It is therefore only possible to establish estimates which are based on experience, or perform tests with a test website. Gathering detailed numbers of percentage are not possible, due to the secrecy of the algorithm of search engines like Google (Evans 2007).

Since the models are based on estimated values, a single ranking cannot be calculated. However, if more models are included and an average value is determined, the ranking becomes more meaningful. The average values of these models provide a reasonable estimation of weighting. In this case, if the values are close to each other it is not possible to rank one aspect higher than the other.

In addition to these on-page elements there are a number of other elements, which do not affect the ranking of a webpage significantly. These ones are not included in this research. If a webpage should be optimized for search engines it is advisable to focus first on these aspects which are most relevant for the visibility of a webpage.

It should be also noted that as a result of both evaluations (Figures 2 and 5) there were no positive or negative factor which individually affect the visibility of a website strongly.

The differences between the Weideman Model, the Sullivan Model and the Spearman correlation could be attributed to the fact that all models are based on estimated values. Also, these models were created at different times.

Finally it should be noted that the first three search results attract the highest number of visitors (Rogers 2002). Using all optimisation approaches combined, and also ensuring that all content is of a high standard and keyword rich, will assist in increasing the ranking of a webpage. Higher ranking becomes more difficult when there are more or stronger competitors in the same field.

References

Bar-Ilan, J., Levene, M. & Mat-Hassan, M. 2006. Methods for evaluating dynamic changes in search engine rankings: a case study. *Journal of Documentation*, 62(6): 708-729.

Chen, Y. & He, C. 2011. Paid Placement: Advertising and Search on the Internet. *The Economic Journal*, 121(556): F309-F328.

Curran, K. 2004. Tips for achieving high positioning in the results pages of the major search engines. *Information Technology Journal*, 3(2): 202-205.

Cutts, M. 2006. *Ramping up on international webspam*. Online: <https://www.matcutts.com/blog/ramping-up-on-international-webspam/> [12 September 2014].

Evans, M.P. 2007. Analysing Google rankings through search engine optimization data. *Internet Research*, 17(1): 21-37.

Garofalakis, J., Kappos, P. & Mourtoukos, D. 1999. Web site optimization using page popularity. *Internet Computing*, IEEE, 3(4): 22-29.

Goldsborough, R. n.d. Helping Others Find You on the Web. *Personal Computing*: 74.

Killoran, J.B. 2013. How to Use Search Engine Optimization Techniques to Increase Website Visibility. *IEEE Transactions on Professional Communication*, 56(1): 50-66.

Rimbach, F., Dannenberg, M. & Bleimann, U. 2007. Page ranking and topic-sensitive page ranking: micro-changes and macro-impact. *Internet Research*, 17(1): 38-48.

Rogers, I. 2002. *The Google Pagerank algorithm and how it works*. Online: http://mira.sai.msu.ru/~megeera/docs/IR/search/pagerank/pagerank_explained.pdf [October 2014].

Schwartz, B. 2014. *Moz Local Ranking Factors 2014 Study Shows Mostly The Same Formula*. Search Engine Land. Online: <http://searchengineland.com/moz-local-ranking-factors-2014-study-shows-mostly-formula-205823> [27 October 2014].

SideOPS SEO Company. *Off -Page SEO*. Online: <http://sideops.com/off-page-seo/> [26 October 2014].

Sullivan, D. n.d. *The Periodic Table Of SEO Success Factors*. Search Engine Land. Online: <http://searchengineland.com/seotable> [27 October 2014].

Velasquez, J.D. 2012. Web site keywords: A methodology for improving gradually the web site text content. *Intelligent Data Analysis*, 16(2): 327-348.

Weideman, M. 2009. *Website visibility: the theory and practice of improving rankings*. Oxford: Chandos Publishers.

Appendix 1

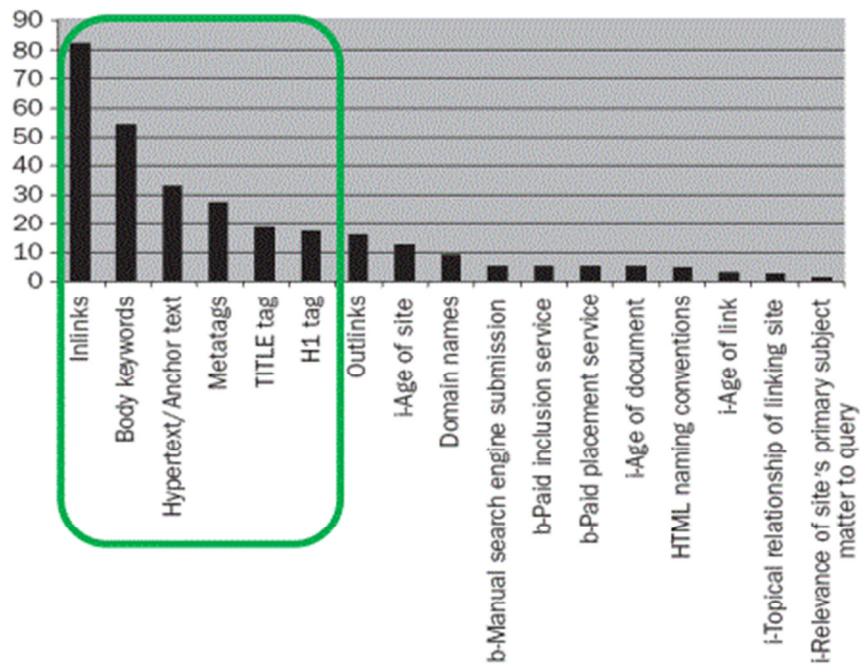
This is a summary of where tools such as Google Analytics, Alexa, Websiteoutlook and RankCheck can be found.

Name	Source
Google Analytics	http://www.google.com/analytics/
Alexa	http://www.alexa.com/tools
Websiteoutlook.com	http://www.websiteoutlook.com/
RankCheck.org	http://rankcheck.org.cutestat.com/

Appendix 2

The Weideman model with those on-page elements that have a positive impact on the ranking of a website is shown here.

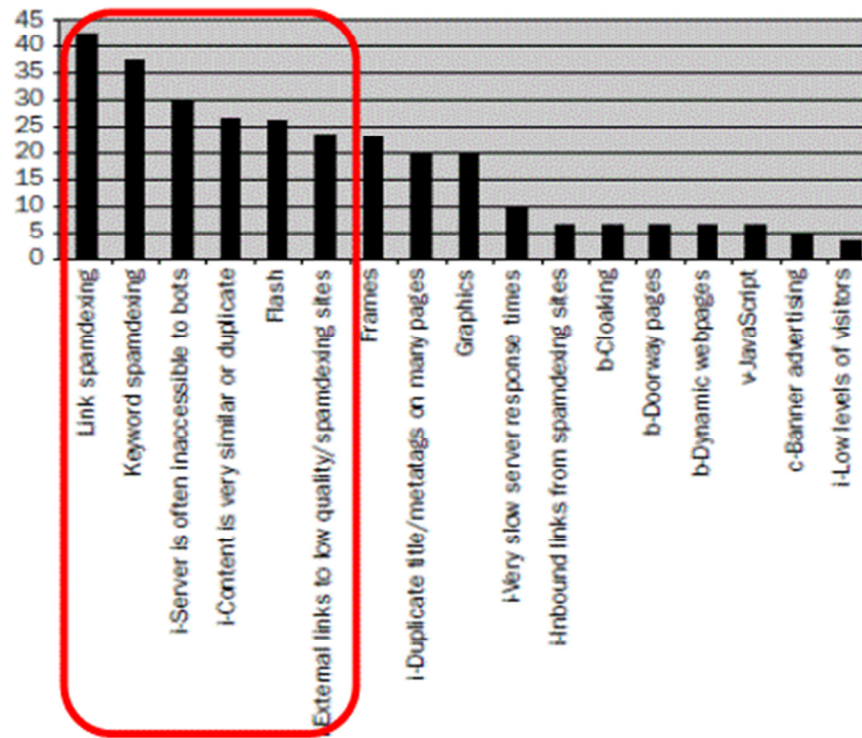
Elements of Visibility - Positive



Appendix 3

This Appendix contains the Weideman model with those on-page elements that have a negative impact on the ranking of a website.

Elements of Visibility - Negative



Appendix 4

This Appendix indicates the Sullivan model with those on-page elements that have a positive and negative influence on the ranking of a website.

ON-THE-PAGE FACTORS

These elements are in the direct control of the publisher

CONTENT

Cq	QUALITY	Are pages well written & have substantial quality content?
Cr	RESEARCH	Have you researched the keywords people may use to find your content?
Cw	WORDS	Do pages use words & phrases you hope they'll be found for?
Ce	ENGAGE	Do visitors spend time reading or "bounce" away quickly?
Cf	FRESH	Are pages fresh & about "hot" topics?
Vt	THIN	Is content "thin" or "shallow" & lacking substance?
Va	ADS	Is your content ad-heavy, especially "above the fold"?

HTML

Ht	TITLES	Do HTML title tags contain keywords relevant to page topics?
Hd	DESCRIPTION	Do meta description tags describe what pages are about?
Hh	HEADERS	Do headlines & subheads use header tags with relevant keywords?
Hs	STRUCTURE	Do pages use structured data to enhance listings?
Vs	STUFFING	Do you excessively use words you want pages to be found for?
Vh	HIDDEN	Do colors or design "hide" words you want pages to be found for?

ARCHITECTURE

THE PERIODIC TABLE OF SEO SUCCESS FACTORS

Search engine optimization — SEO — seems like alchemy to the uninitiated. But there's a science to it. Below are some important "ranking factors" and best practices that can lead to success with both search engines and searchers.

ON-THE-PAGE SEO					OFF-THE-PAGE SEO								
CONTENT	HTML	ARCHITECTURE	LINKS	TRUST	SOCIAL	PERSONAL							
Cq ⁺³ Quality	Ht ⁺³ Titles	Ac ⁺³ Crawl	Lq ⁺³ Quality	Ta ⁺³ Authority	Sr ⁺² Reputation	Pc ⁺³ Country							
Cr ⁺³ Research	Hd ⁺² Description	Ad ⁺² Duplicate	Lt ⁺² Text	Th ⁺¹ History	Ss ⁺¹ Shares	Pl ⁺³ Locality							
Cw ⁺² Words	Hh ⁺¹ Headers	As ⁺¹ Speed	Ln ⁺¹ Numbers	Ti ⁺¹ Identity	Ph ⁺³ History								
Ce ⁺² Engage	Hs ⁺¹ Structure	Au ⁺¹ URLs	Vp ⁻³ Paid	Vd ⁻¹ Piracy	Ps ⁺² Social								
Cf ⁺² Fresh	Vs ⁻¹ Stuffing	Am ⁺¹ Mobile	VI ⁻² Spam										

OFF-THE-PAGE FACTORS

Elements influenced by readers, visitors & other publishers

LINKS

Lq	QUALITY	Are links from trusted, quality or respected web sites?
Lt	TEXT	Do links pointing at pages use words you hope they'll be found for?
Ln	NUMBER	Do many links point at your web pages?
Vp	PAID	Have you purchased links in hopes of better rankings?
VI	SPAM	Have you created many links by spamming blogs, forums or other places?

TRUST

Ta	AUTHORITY	Do links, shares & other factors make site a trusted authority?
Th	HISTORY	Has site or its domain been around a long time, operating in same way?
Ti	IDENTITY	Does site use means to verify its identity & that of authors?
Vd	PIRACY	Has site been flagged for hosting pirated content?

SOCIAL

Sr	REPUTATION	Do those respected on social networks share your content?
Ss	SHARES	Do many share your content on social networks?

PERSONAL

Pc	COUNTRY	What country is someone located in?
Pl	LOCALITY	What city or local area is someone located in?

Appendix 5

This Appendix indicates a Spearman Correlation between those on-page elements that have a positive and negative influence on the ranking of a website.

