CORBETTA Andrea
DE SIMONI Pietro
SESSI Michela
VERRIER Alexandra

Report of Assignment 3 :

Search Engine Optimization

- 1 -

Table of Contents

Introduction

The aim of this assignment is to understand practically how search engines work. In particular the exercise involves search engine optimization.

In order to test our abilities and fulfil the requirements we have created some web pages and a small website using Wordpress.com. Our purpose was to publish a small guide of Stockholm for exchange/Erasmus students, the website includes pages for food, activities in Stockholm, travelling around Stockholm, nightlife and swedish behaviour and habits.
The link is the following :
https://mystockholmdotblog.wordpress.com/
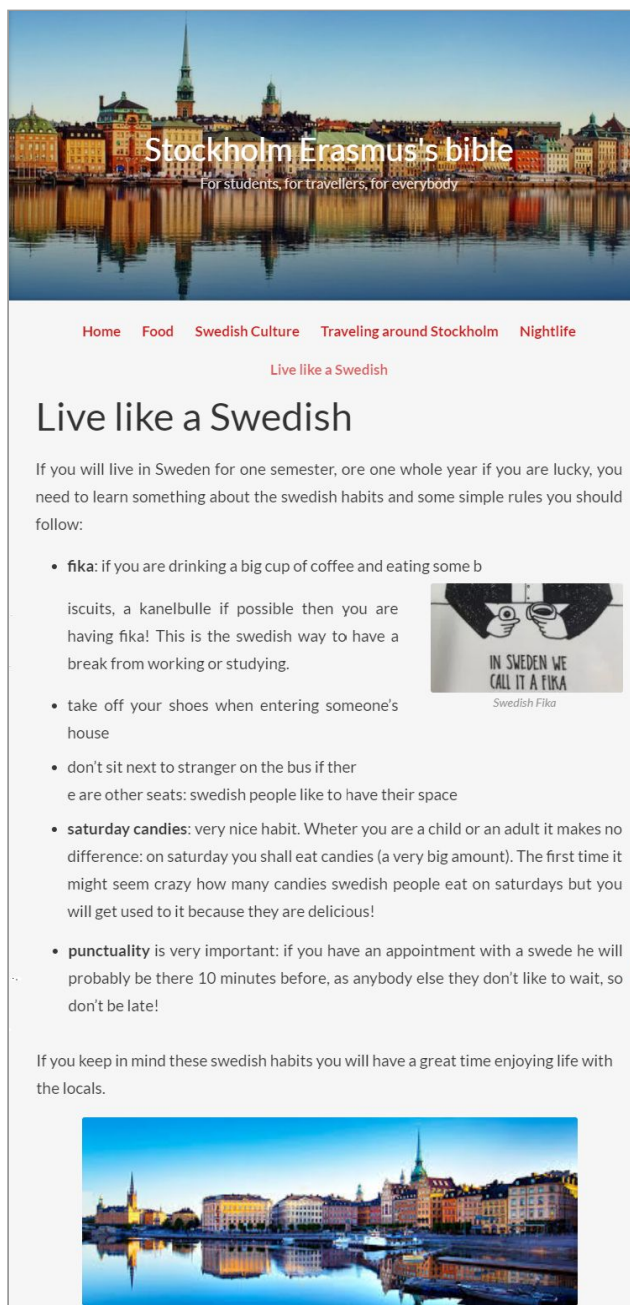
Optimization target

From our mini website, we choose one page that we want to optimize. Particularly we select the page regarding facts about swedish people: every country has peculiar ways of doing, different customs and singularities. Therefore, the section wants to report all the peculiarities in being swedish.

The page link is the following:
https://mystockholmdotblog.wordpress.com/live-like-a-swedish/

We want to remember that the page is updated with improvements, so the initial page information, no more available at the link, are reported as screenshots and comments in this and next sections.

Figure 1 shows the initial webpage.

In order to start our work we selected our query regarding this specific page:
q: (swedes OR swedish) habits

Figure 1 *: Initial target webpage*

Initial Indexing

We used SiteSeeker in order to index our pages. In this case, the site is not a real one, we created it with the only purpose of didactic support. For that reason, as the number of pages on our mini-site is too small, we index also a portion of some competing public website. We selected the 10 following websites:

- https://www.visitstockholm.com/
- https://www.thelocal.se/20150422/my-swedish-habits-that-foreigners-just-dont-get/
- https://hejsweden.com/en/how-to-be-swedish-in-10-easy-steps/
- https://www.irishtimes.com/life-and-style/abroad/ten-things-i-ve-learned-from-two-decades-living-in-sweden-1.3068733
- https://theculturetrip.com/europe/sweden/articles/11-habits-you-pick-up-when-living-in-sweden/
- https://matadornetwork.com/read/10-swedish-habits-adopt/
- http://www.oenoforos.se/swedish-food-habits.aspx
- https://blogs.transparent.com/swedish/swedish-checkout-habits/
- http://lostinstockholm.com/2009/08/27/10-things-to-know-about-swedes/
- https://brightside.me/wonder-places/30-habits-of-the-swedes-that-make-their-country-one-of-the-happiest-in-the-world-608910/

By indexing the previous servers we obtain an index composed by 143 webpages (*Figure 2*).

Produktversion	6.10.0954								
Type and timestamp	requested / scheduled / notified, started **2018-12-03 14:09** and lasted 38 minutes								
Full crawling enabled	https://blogs.transparent.com, https://brightside.me, https://hejsweden.com, https://matadornetwork.com, https://mystockholmdotblog.wordpress.com, https://theculturetrip.com, https://thelocal.se, https://www.irishtimes.com, https://www.visitstockholm.com, lostinstockholm.com, www.oenoforos.se								
Minimal crawling enabled	-								
No crawling	-								
Indexed objects	143 web pages								
Language	140 English, 2 Swedish, 1 Spanish. 0 other								

Servers

Description »

Document count	Server	Status	Mapped to (DNS: etc/hosts)	Type	Limit	Link	NOX	Dup	Dup2	Dead
16	https://blogs.transparent.com	OK		standard	32%	2/1	1	0	0	0
25	https://brightside.me	⚠		standard	100%	2/1	0	0	0	0
1	https://hejsweden.com	⚠		standard	100%	1/1	0	0	0	1
25	https://matadornetwork.com	⚠		standard	100%	2/1	0	0	0	0
10	https://mystockholmdotblog.wordpress.com	OK		standard	20%	1/2	0	0	0	0
24	https://theculturetrip.com	⚠		standard	100%	2/1	0	1	0	0
0	https://thelocal.se	OK		standard	-	0/1	0	0	0	0
0	https://www.irishtimes.com	OK		standard	-	0/1	0	0	0	0
1	https://www.visitstockholm.com	⚠		standard	100%	1/5	0	23	0	0
24	lostinstockholm.com	⚠		standard	100%	1/1	0	0	0	0
17	www.oenoforos.se	OK		standard	34%	2/1	0	2	0	0

Figure 2 : Indexing results

A relevant observation has to be done: SiteSeeker does not work with synonyms. That is a considerable point in our exercise, in fact keywords are searched just in absolute terms, there are no searches for related words.

For example if we search just for the first word in our query **q,** "Swedes", the tool can't understand the related relevancy of the word "Swedish". In order to solve that problem we add both the keywords connected with the operator OR.

We compute the query **q** on those webpages and obtain the following ranking of the 15 retrieved documents:

Figure 3 : *Top position of the ranking*

Figure 4 : *Target page position*

Figure 3 and *Figure 4* show the position (8th over 15) of our optimization target page and the top of the relevant links. Of course, our page is not in the top position, therefore there is space for index improvements.

Text optimization

We begin the improvement process by executing an on-site optimization. Indeed we want a perfect match between our webpage and the query.

We estimated that the URL of our target page could include more keywords, especially "habits". Hence, without changing the domain name we adapted the path name to our query.
- old path:

https://mystockholmdotblog.wordpress. com/live-like-a-swedish/
- new path:

https://mystockholmdotblog.wordpress.com/discover-swedish-habits/

Besides, we also changed the main title. As the title of a webpage is important in the search process, we changed its formulation for a better match with the query **q**: from "*Live like a Swedish*" it became "*Discover the Swedish habits*". This leads to a more specific topic according to our search.

To improve the keyword placement in our page we added some more in relevant position inside the body text, moreover bold and italic text are performed. Avoiding keywords' spamming, this allow the search engine to identify easily the main topic of this specific webpage.

In addition, we improved tag meta-description of the images. Indeed, we added some relevant keywords in the description of the picture as is shown in *Figure 5* and *Figure 6*. This step is important for the search engine but also for the user because it represents the text that will be displayed if the image could not be properly loaded, for instance if there are some network problems.
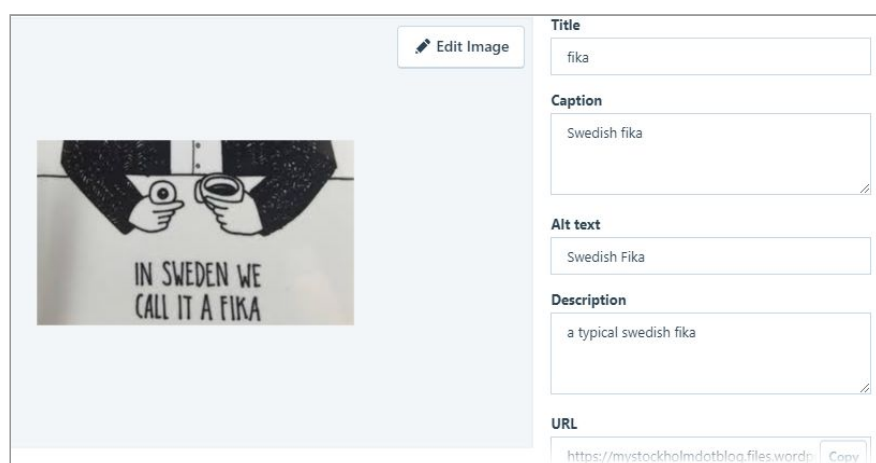

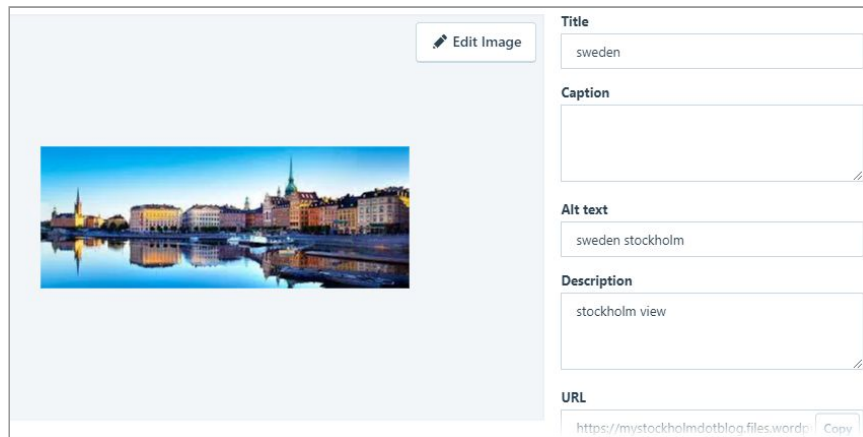*Figure 5 : Tag meta-description of the first picture*

Figure 6 : Tag meta-description of the second picture

Link optimization

However we know that it's the off-site ranking factors that will give your web pages the biggest ranking boost. In this way we did a link optimization to add incoming links to our target page.

We noticed some similarities between the different sections of our mini-website. So we decided to add a link from the web page about food to our target page, in fact there is a related topic: the concept of *fika*. We also improved our home page to link to the different section we have. Moreover, we used some keywords from the selected query in the different links to optimize the link reputation.

Nevertheless we have chosen to not use techniques such as "black-hat" search engine optimization (SEO) because this is not used for real-life websites. We think it's a "links' spamming" technique. Indeed "black-hat" SEO is a link optimization method in which the website's owner add hiding links under images or invisible links.

Connection to and from the Outside World

To improve our visibility we need some other websites to link to our pages. Obviously this is not an easy step. There are two ways to increase the number of incoming links: buying links or making good contents and wait for other webpages to notice you.
The first method is quicker but the pages that will point at your page will be very low-ranked, in fact they will be just full of artificial links. It takes much more time to perform the second method, but on the other hand it will give greater results. Having incoming links from a few high ranked pages will improve your rank much more than having many incoming links from link farms.

So in order to improve our page ranking our website could benefit of incoming links from sources like:

- source1: https://erasmusu.com/fr/erasmus-stockholm/experiences-erasmus
- source2: https://www.visitstockholm.com/
- source3: other blogs with similar content such as the following
  - https://www.nomadicmatt.com/travel-blogs/the-saturday-city-stockholm/
  - http://www.blondeatlas.com/blog-1//stockholm-travel-guide

We selected some webpages dealing with the same topics and which could in the real life link our page if we would improve the content.

This step in the improvement process is important and could lead to a real improvement of our ranking in bigger search engine results such as running our query on "google.com".

Final Indexing

After the optimization of the target page we re-indexed the same servers.
We ran the same query **q** and retrieved 21 documents. Our webpage is now on the third position. We improved its rank of 5 places, moving from the 8th position to the 3rd position.

Figure 7 : *New target page position*

We notice that there are still two pages above and conclude that this could be avoided by using "black-hat SEO" or by receiving some incoming links from the outside world.

Discussion and Conclusion

Various observations

Through the process of optimization we observed the improvement of our page thanks to the techniques explained above.

Moreover we could see another aspect of the linking environment. In fact our page received a few comments classified as spam *(Figure 8)*. Some users may want to improve their website's views by adding comments to our page, containing links to other pages (probably their pages). If our blog was considered reliable and it had a high ranking, the 'spam user' would benefit of a possible ingong link from our page in order to improve his/her page rank.
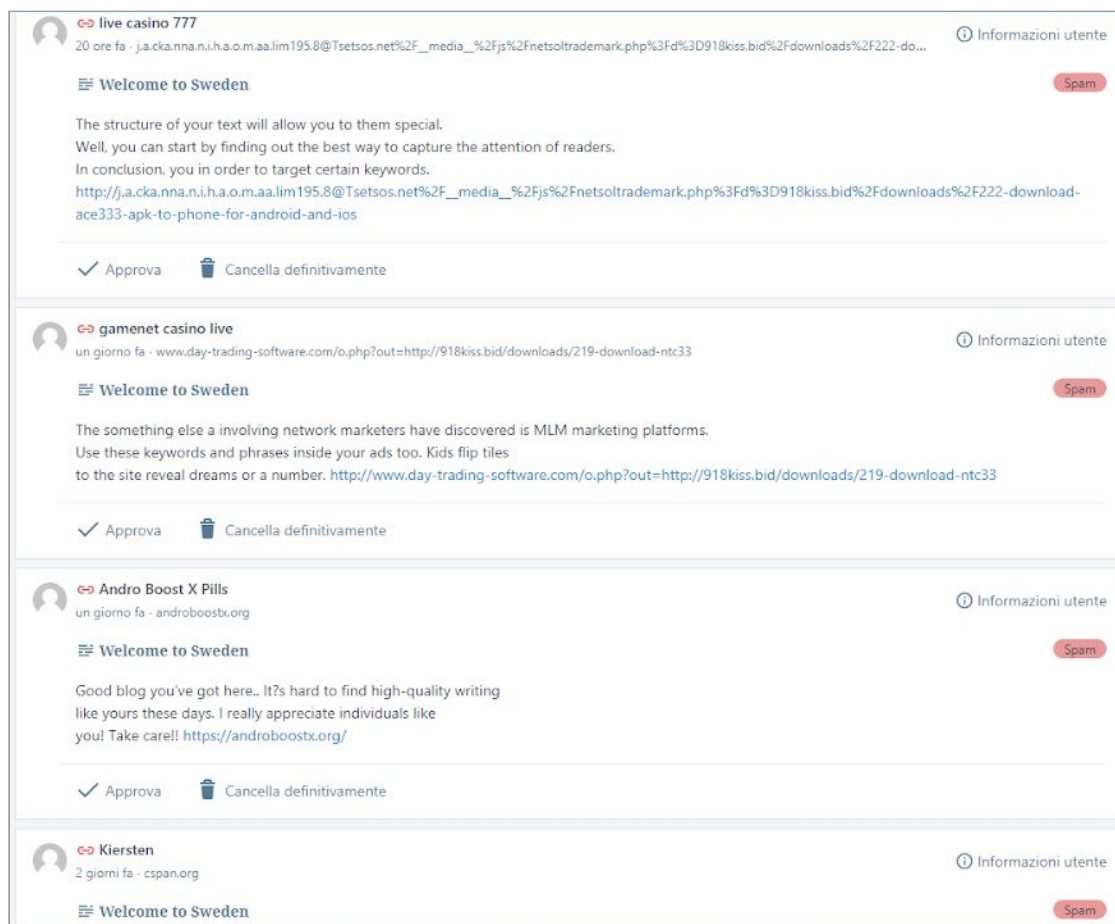

Figure 8 : *Spam comments received*

Deeper analysis of the final indexing result

The final indexing shows a particular ranking (*Figure 7*): even if our page has improved its position (3rd), the second position is a different page from our mini website: the first welcome article. Before the changes in text and links, this article was not even reported in the ranking from SiteSeeker. The improvements were clearly done with the purpose to obtain the best ranking for the target page, but some side effect appeared. We added all the links to our different sections of the site in the welcome article in order to have also one more link to our section "*Discover Swedish Habits*" (*Figure 9*). The consequence is the higher presence of the keywords according to our query **q** and the resulted ranking**.**
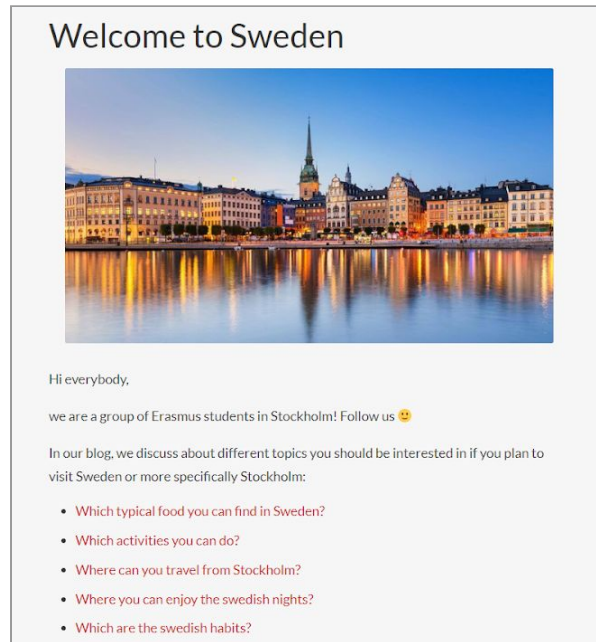

Figure 9 : First welcome article

To improve the ranking of our target page regarding this page we could use NOFOLLOW for each link from the article which not reach the target. We didn't choose to do it because we think that in a real-life site we want to optimize this page without decreasing the visibility of other pages for other queries.

Conclusion

Nowadays websites can reach millions of people. Many companies, and also bloggers, make their business on this. Information sites, like newspapers and blogs make money proportionally to the views, because having many visitors allows them to sell advertising.
The more people visit your website, the higher price you can sell your advertising spaces. Another important sector which struggles to get more visibility is e-commerce. When people look for a certain product they will be likely to buy it on one from the first sites that they find. For an e-commerce company appearing just one or two positions higher in web searches can improve the retails of millions of dollars.
Hence, we learnt in this experiment that even if search engine optimization takes time and money, it's more than important for the visibility of our websites, the numbers of users' visits and the gain of the business.