

# Search engine optimization

Bc. Roman Peček

---

Master thesis  
2011



Tomas Bata University in Zlín  
Faculty of Applied Informatics

---

Univerzita Tomáše Bati ve Zlíně

Fakulta aplikované informatiky

akademický rok: 2010/2011

## ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Roman PEČEK**

Osobní číslo: **A09485**

Studijní program: **N 3902 Inženýrská informatika**

Studijní obor: **Informační technologie**

Téma práce: **Search Engine Optimization (SEO optimalizace)**

Zásady pro vypracování:

1. Vypracujte literární rešerši na téma SEO (Search Engine Optimization).
2. Popište, jak vidí stránku vyhledávače a jak uživatel.
3. Navrhněte postupy, jak vylepšit hodnocení stránky uživatelem a jak vylepšit hodnocení stránky vyhledávačem.
4. Analyzujte možnosti poškození stránky.
5. Získané postupy aplikujte na internetový portál s flash hrami a vyhodnoťte jejich úspěšnost.

Rozsah diplomové práce:

Rozsah příloh:

Forma zpracování diplomové práce: **tištěná/elektronická**

Seznam odborné literatury:

1. ENGE, E., SPENCER, S., FISHKIN, R., STRICCHIOLA, J., The art of seo – mastering seo, 1st ed., OReilly Media, October 21st 2009, ISBN: 978-0-596-51886-8
2. SOUDERS, S., High performance web sites essential knowledge for frontend engineers, 1st ed., OReilly Media, September 11th 2007, ISBN: 978-0-596-52930-7
3. SOUDERS, S., Even Faster Web Sites: Performance Best Practices for Web Developers, 1st ed., OReilly Media, June 10th 2009, ISBN: 978-0-596-52230-8
4. Search engine optimization starter guide [online], 2010 [cit. 2010-10-20], [http://static.googleusercontent.com/external\\_content/untrusted\\_dlcp/www.google.com/cs//w/engine-optimization-starter-guide.pdf](http://static.googleusercontent.com/external_content/untrusted_dlcp/www.google.com/cs//w/engine-optimization-starter-guide.pdf)
5. SEOmoz the beginners guide to seo [online], 2010 [cit. 2010-20-10], <http://guides.seomoz.org/beginners-guide-to-search-engine-optimization>
6. KUBÍČEK, M., Velký průvodce seo, 1st ed., Computer Press, August 2008, ISBN: 978-8-0-251-21955

Vedoucí diplomové práce:

**Ing. Tomáš Sysala, Ph.D.**

Ústav automatizace a řídicí techniky

Datum zadání diplomové práce:

**24. února 2011**

Termín odevzdání diplomové práce:

**18. května 2011**

Ve Zlíně dne 24. února 2011



prof. Ing. Vladimír Vašek, CSc.  
*děkan*



doc. Mgr. Roman Jašek, Ph.D.  
*ředitel ústavu*

## **ABSTRACT**

Cílem této práce bylo popsat techniky, které se používají k optimalizování webových stránek. Teoretická část je zaměřena na techniky, které mohou zlepšit uživatelský požitek a také požitek internetového vyhledávače.

Praktická část ukazuje tyto techniky v praxi. Je zde ukázáno, jak byla webová stránka zdokonalena, aby byla atraktivnější pro vyhledávače a také pro uživatele.

Klíčová slova:

SEO, search engine marketing, internet marketing, short tail, long tail, CDN, content delivery network, search engine optimization

## **ABSTRACT**

The goal of this thesis was to describe techniques which are used for website optimization. The theoretical part is focused on techniques which could improve users experience and search engines experience as well.

The practical part shows the techniques in practice. There is shown how was the website improved to look more attractive for both – search engines and users.

Keywords:

SEO, search engine marketing, internet marketing, short tail, long tail, CDN, content delivery network, search engine optimization

Na tomto místě bych rád poděkoval Ing. Tomáši Sysalovi, Ph.D. za cenné připomínky a odborné rady, kterými přispěl k vypracování této diplomové práce.

### **Prohlašuji, že**

- beru na vědomí, že odevzdáním diplomové/bakalářské práce souhlasím se zveřejněním své práce podle zákona č. 111/1998 Sb. o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších právních předpisů, bez ohledu na výsledek obhajoby;
- beru na vědomí, že diplomová/bakalářská práce bude uložena v elektronické podobě v univerzitním informačním systému dostupná k prezenčnímu nahlédnutí, že jeden výtisk diplomové/bakalářské práce bude uložen v příruční knihovně Fakulty aplikované informatiky Univerzity Tomáše Bati ve Zlíně a jeden výtisk bude uložen u vedoucího práce;
- byl/a jsem seznámen/a s tím, že na moji diplomovou/bakalářskou práci se plně vztahuje zákon č. 121/2000 Sb. o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) ve znění pozdějších právních předpisů, zejm. § 35 odst. 3;
- beru na vědomí, že podle § 60 odst. 1 autorského zákona má UTB ve Zlíně právo na uzavření licenční smlouvy o užití školního díla v rozsahu § 12 odst. 4 autorského zákona;
- beru na vědomí, že podle § 60 odst. 2 a 3 autorského zákona mohu užít své dílo – diplomovou/bakalářskou práci nebo poskytnout licenci k jejímu využití jen s předchozím písemným souhlasem Univerzity Tomáše Bati ve Zlíně, která je oprávněna v takovém případě ode mne požadovat přiměřený příspěvek na úhradu nákladů, které byly Univerzitou Tomáše Bati ve Zlíně na vytvoření díla vynaloženy (až do jejich skutečné výše);
- beru na vědomí, že pokud bylo k vypracování diplomové/bakalářské práce využito softwaru poskytnutého Univerzitou Tomáše Bati ve Zlíně nebo jinými subjekty pouze ke studijním a výzkumným účelům (tedy pouze k nekomerčnímu využití), nelze výsledky diplomové/bakalářské práce využít ke komerčním účelům;
- beru na vědomí, že pokud je výstupem diplomové/bakalářské práce jakýkoliv softwarový produkt, považují se za součást práce rovněž i zdrojové kódy, popř. soubory, ze kterých se projekt skládá. Neodevzdání této součásti může být důvodem k neobhájení práce.

### **Prohlašuji,**

- že jsem na diplomové práci pracoval samostatně a použitou literaturu jsem citoval. V případě publikace výsledků budu uveden jako spoluautor.
- že odevzdaná verze diplomové práce a verze elektronická nahraná do IS/STAG jsou totožné.

Ve Zlíně

.....  
podpis diplomanta

# CONTENTS

<b>INTRODUCTION .....</b>	<b>9</b>
<b>1 THEORY.....</b>	<b>10</b>
<b>1 HISTORY.....</b>	<b>11</b>
<b>2 INTRODUCTION TO SEARCH ENGINE OPTIMIZATION.....</b>	<b>17</b>
2.1 USER EXPERIENCE .....	18
2.2 SEARCH ENGINES' EXPERIENCE .....	20
<b>3 BASIC STEPS.....</b>	<b>22</b>
3.1 STAY ONLINE.....	23
3.1.1 Server timeouts.....	23
3.1.2 Blocked IP addresses.....	24
3.1.3 Bandwidth limits .....	24
3.2 KEYWORD RESEARCH.....	24
3.2.1 Google's AdWords Keyword Tool .....	27
3.2.2 Google's AdWords Traffic Estimator tool.....	28
3.2.3 Yahoo! Search Marketing .....	28
3.2.4 Seznam.cz Keyword Tool .....	28
3.3 DOMAIN NAME .....	29
3.3.1 Domain name hunting .....	29
3.3.2 Domain length .....	29
3.3.3 Domain tools .....	30
3.3.4 Domain name recommendations .....	30
3.4 URLs .....	31
3.4.1 URL length.....	31
3.4.2 URL types.....	32
3.4.3 URL characters.....	32
3.4.4 URL recommendations .....	33
3.5 FASTER LOADING.....	34
3.5.1 Content delivery network .....	34
3.5.2 CSS sprites .....	35
3.5.3 Compression.....	36
3.5.3.1 Compression during delivering.....	36
3.5.3.2 Compression during development .....	37
3.5.4 Content caching.....	37
3.5.4.1 Expires header.....	39
3.5.4.2 Max-age directive .....	39
3.5.5 Styles, scripts position.....	39
3.6 ON-SITE FACTORS .....	41
3.6.1 HTML title tag .....	41
3.6.1.1 Keywords position .....	42
3.6.1.2 Title length.....	42
3.6.1.3 Title tag content .....	42

3.6.2	Meta tags .....	43
3.6.2.1	Description tag length .....	43
3.6.3	Heading tags .....	43
3.6.3.1	H1 content .....	44
3.6.4	Images .....	44
3.6.5	Website structure .....	46
3.6.5.1	Flat structure .....	46
3.6.5.2	Deep structure .....	46
3.7	OFF-SITE FACTORS .....	46
3.7.1	Backlinks .....	47
3.7.1.1	Backlinks hunting .....	47
3.7.1.2	Backlinks' anchor text .....	48
3.7.2	Social media marketing .....	48
<b>4</b>	<b>BLACK HAT OPTIMIZATION .....</b>	<b>50</b>
4.1	NOT RECOMMENDED TECHNIQUES .....	50
4.1.1	Duplicate content .....	51
4.1.2	Useless content .....	51
4.1.3	Backlinks business .....	51
4.2	FORBIDDEN TECHNIQUES .....	52
4.2.1	Hidden text .....	52
4.2.2	Cloaking .....	53
4.2.3	Fake links .....	53
<b>II</b>	<b>PRACTICE .....</b>	<b>54</b>
<b>5</b>	<b>WEBSITE OPTIMIZATION.....</b>	<b>55</b>
5.1	WEBHOSTING SELECTION .....	55
5.2	KEYWORDS RESEARCH .....	56
5.3	DOMAIN NAME .....	58
5.4	URLS .....	59
5.5	FASTER LOADING.....	61
5.6	ON-SITE OPTIMIZATION.....	63
	<b>CONCLUSION .....</b>	<b>67</b>
	<b>BIBLIOGRAPHY .....</b>	<b>68</b>
	<b>LIST OF ABBREVIATIONS .....</b>	<b>69</b>
	<b>LIST OF FIGURES .....</b>	<b>70</b>
	<b>LIST OF TABLES .....</b>	<b>71</b>
	<b>APPENDICES .....</b>	<b>72</b>
	<b>APPENDIX P I: APPENDIX TITLE.....</b>	<b>73</b>

## INTRODUCTION

Since the Internet was “born” our lives have started changing. People spend more time glued to the computer screens. The way of living, as had been known before, disappeared. The Internet walked a long way before it matured in to the medium which we know today. The Internet slowly became a global highway. Anyone who wants to succeed has to be a part of this highway, but this is not enough, he has to be seen by others. It is very easy to be a part of anything, but being just a part never brought success. It is necessary to be an active “part”. In the online world it is very hard to succeed.

If you run an online shop, it is not only about people, who you know, it is about people who do not know you and they do not know anything about your efforts, it is necessary to tell them that you run an online shop. Everyone who wants to be seen in the online world has to catch users and bring them to his website.

In the real world it is not enough when the website is copied to the webhosting. There has to be someone who looks after the website. It is necessary to react on market changes, on modern trends and changes which are made by competitors.

One way how tell the users that the website is up and is ready to welcome them, is through the search engines. These search engines have become very popular in the last decade. They are so popular that the success of the website depends on how the site is seen by search engines, so that it is necessary to know how to make the website more attractive for these programs.

This thesis is divided in to four parts. Each of these parts is focused on aspects which are related to the search engines optimization.

In the first part there is a brief history of search engines. It gives an idea how the situation was when the Internet was a small network consisted of only a few computers.

The next chapter is focused on two different visitors of the website – search engines and human beings. It describes techniques how to make the website look attractive for users and for search engines as well. It is usually not so easy, as it could look.

The last part of this thesis shows the approaches in a production environment. This chapter shows the whole process of optimization from the beginning to the end.

## **I. THEORY**

## 1 HISTORY

Probably the first search engine which was developed in 1989 by Peter Deutsch, Alan Emtage and Bill Wheelan was Archie. This computer program was developed at McGill University. The name Archie is a contraction of a word “archive”. Archie was not a true search engine it scoured the Internet for content on anonymous FTP servers all over the world. It collected all the names of the files and then it was possible to find the file if you knew the name of that file. In fact it was a searchable list of files. Before Archie it was necessary to know two details – the name of the file and the address of the server where the file was located, the information about the file and the server could be obtained only by word of mouth or by email. If the name of the file was known, it was easy to ask Archie where the file is and then download it. By year 1992 the Archie had indexed more than 200 public FTP servers. It could sound a bit silly today but in that time it was almost the whole Internet. When the Archie was at its peak 30 Archie engines crawled the Internet. Archie is sometimes called the grandfather of search engines.

If the Archie is the grandfather of search engines, Veronica could be called theirs’ grandmother. Veronica was developed at University of Nevada by Fred Barrie and Steven Foster. The name Veronica was chosen because it is an acronym for “Very Easy Rodent-Oriented Netwide Index to Computerized Archives”. This program was developed to search for a Gopher files. A Gopher server stored only text-plain documents – FTP stores all files, not only text files. At Veronica’s peak this program searched through a database of 5500 Gopher servers and had indexed more than 10million of Gopher files.

By year 1993 the Internet started changing, before this year the Internet consisted mainly of FTP servers, Gopher servers and email servers, but this year World Wide Web pages started to be more popular. So that the first URLs list which was developed by Matthew Gray, was introduced. Its name was World Wide Web Wanderer. In fact it was a series of computer programs called robots, these robots went through the all links on the websites and put them into the database called Wandex. At first the WWW Wanderer’s goal was to measure how the Internet grows, but then it had the first database of WWW URLs. In June, 1993 Wanderer had 130 records in its database, but in January, 1996 it had more than 100 000 records in the database. Thanks to these robots it was possible to see how the Internet was growing.

The proper search engine was created in 1993, its name was Aliweb. It was developed by Martijn Koster. The Aliweb offered webmasters to submit a special file with information about the website. This file was written by human being, the information about the website were – site description and a list of keywords. By the time this thesis is being written Aliweb's website is still up, its address is [www.aliweb.com](http://www.aliweb.com).

The next generation of search engines was a bit smarter they used spiders to crawl the Internet. These spiders visited the website and got the information about keywords on the website from its titles, headers tags and from URLs. These engines did not rank any website they just only put the keywords into the database. So that when the search query came, the search engine returned a list of pages where the keyword had been found. These results were not quality.

Probably the first search engine which did some kind of raking was Excite, this search engine's previous name was Architext. This search engine was developed by 6 students at Stanford University, these students were Graham Spencer, Joe Kraus, Mark Van Haren, Ryan McIntyre, Ben Lutch and Martin Reinfried, it was in year 1994. This search engine still works and it is possible to find it at [www.excite.com](http://www.excite.com).

Another big player on this field was Yahoo!. This project started in 1994. It was a website owned by David Filo and Jerry Yang, Ph.D. students at Stanford University. At first it was a list of websites' URLs which these two students were interested in. When the number of links started to grow they decided to divide these links into categories. In fact it was not a proper search engine it was rather an online searchable directory which allowed people to search through the links. These links were divided into the categories by human beings – Yahoo! did not use any robots to do the work. Probably this made Yahoo! so popular, because people could distinguish whether the link is quality or not. In 1994 Yahoo! was so popular that it had 1 million visitors per day and 100 000 unique visitors per day. Filo and Yang said that their directory was called Yahoo! because they were “yahoos”, that is why the Yahoo! was called Yahoo!.

The first full-text search engine was WebCrawler. This search engine began as a seminar project of Brian Pinkerton at Washington University. It was written for his personal needs, but after a short time his schoolmates wanted to use such a tool, so that he made a web interface for it. Search engines which had been developed before WebCrawler could only

index pages by their titles and URLs, but WebCrawler came with different approach, it crawled the whole content of the page, all texts, so that it indexed every word on the page. After few months WebCrawler had 15 000 visitors per day. In 1997 WebCrawler was bought by Excite.

Another search engine which was created in 1994 was Lycos. It was using “spider concept”. This search engine was born at Pittsburgh’s Carnegie Mellons University by Dr. Michael Mauldin. Lycos was named after the spider Lycosiade lycosa. When the Lycos went public it had 54 000 documents in its database. Lycos came with a new idea, it was the first engines which was doing some kind of raking of relevance. By August 1994 Lycos had 394 000 documents in its database. By January 1995 it had 1,5 million documents and by November 1996 it had more than 60 million documents. So it is possible to say, that this search engine was quite popular due to its ability to rank the searched documents.

In January 1994 Steven Kirsch created InfoSeek, at first this service was paid, but in August 1994, he decided to offer it for free. In February 1995 this engine was renamed to InfoSeek Search, after that it was a proper search engine as are known today. It had a user-friendly interface, it also had a news feed. But what made InfoSeek Search so popular was deal between InfoSeek Search and Netscape Navigator – InfoSeek Search became the default search engine for Netscape’s web browser, Yahoo! was the default search engine before InfoSeek Search. This deal made it much more popular than ever before.

By the end of 1995, Digital Equipment Corporation introduced a new search engine called AltaVista. This search engine was innovative because it let users to ask questions instead of simply inserting keywords – for instance user entered a question such as “Where is London” and the search engine returned an intelligent answer, instead of returning millions of pages which contain words “where”, “is” and “London”. It also gave users the ability to use Boolean operators such as And, Or, Not, But, so that it was possible to create more sophisticated search queries. This search engine ran on the most powerful processors available at that time and it also ran a special edition of Unix. Due to these facts, these servers were able to handle high traffic – more than million of hits a day.

Between years 1995-1998 there were some search engines which were trying to become popular, for instance – The Northern Light – created in 1995, HotBot – created in 1996, this search engine was famous because it was using cluster and was able to index up to 10

million of pages every day, so that it was easy for him to return to the previously indexed page and check whether there are any changes or not, LookSmart – created in July 1996, GoTo – created in 1997, first concept of pay-per-click advertising, companies agreed to pay a fee every time the user clicked on their ad in search engine, Ask Jeeves – created in 1997, this search engine used NLP – natural language processing, this technology let the users to enter the search queries in everyday language.

Probably the most famous search engine which is the most using on the Earth started in 1998. Its first name was BackRub. This search engine was developed by two students Larry Page and Sergey Brin at Stanford University. The originality of this search engine was that it was analyzing back links which pointed to a given website. As the project grew it needed more space so that Page and Sergey decided to move it to the dorm room. When this search engine reached one terabyte level of data, it was renamed to Google, a play on googol – it is a number followed by 100 zeroes. At first they wanted to sell it to Yahoo!, but the owner of Yahoo! told them to come back when they commercialize it. After this “fail” Google got a cheque for \$100 000 from Sun Microsystems. So this was the first capital of Google Inc.. Until September 21, 1999, Google still had “beta” label on its website. Google gradually grew up, at first it run in a dorm room, then in a garage and after few years Google moved in to its own complex called Googleplex. By October 2003 it had indexed more than 3,3 billions of pages, millions of images, countless newsgroups and much more. Google became a phenomenon. People started using terms like “googling” – which means searching something on the Internet. Google is the most using search engine nowadays its market share is more than 84%. It handles more than 400 million queries per day. Google became the most using search engine in many countries all over the world.

Almost in the same time when the Google started appearing, Microsoft came up with its own search engine called MSN Search, it was a part of Microsoft’s MSN portal. At first it was a search tool for portal customers. When the MSN Search was publicly available it was the most using search engine on the Earth, because it was a default search engine in Microsoft’s Internet Explorer web browser. At first MSN Search was using results from Inktomi, then for a short time results from AltaVista and after that Microsoft developed its own search engine. This search engine was updated weekly and sometimes daily. As an image search MSN Search used third party engine – Picsearch. On September 11, 2006 MSN Search was replaced by Microsoft Live Search – this search engine stopped using

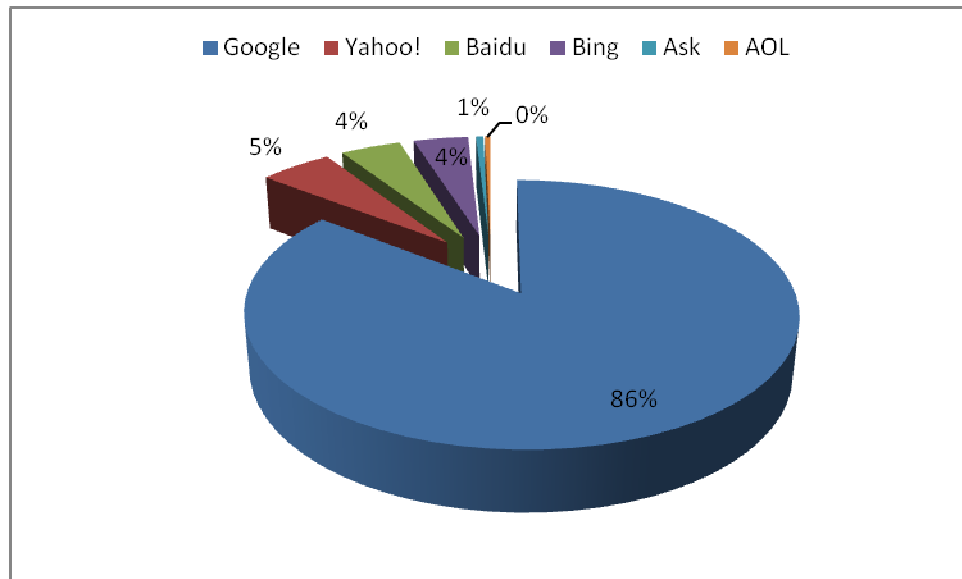
Picsearch, because it had its own search algorithm for searching for images. On June 3, 2009, Microsoft Live Search was officially replaced by Bing. On June 29, 2009, Microsoft and Yahoo! announced that Yahoo! will be using Bing as its search engine in next 10 years.

In year 1996 Ivo Lukačovič introduced his online catalogue. The functionality was quiet similar to Yahoo!. This catalogue was divided into categories and each category had its own subcategories.



Picture 1 The first Seznam's webdesign from the year 1996

On today's Internet there are six search engines, obviously there are more but these search engines have significant amount of percentage. The most using search engine is Google, it has 84.64%, the second biggest search engine is Yahoo! and its share is 5.15%, the third place belongs to Baidu – this search engine is the biggest one in China, it has 4.3%, the fourth is Bing, its share is 3.91%, Ask has 0.53% and the sixth is AOL with 0.38%. There is no doubt that Google is the most using search engine on the Earth, but despite this fact it is possible to find countries where Google is not number one, for example in the Czech Republic the number one is search engine called Seznam.



Picture 2 Search engines market share

## 2 INTRODUCTION TO SEARCH ENGINE OPTIMIZATION

In today's world it is necessary to be seen when you run a website. It is not only about that the website is uploaded on a webhosting, but that is not all. It is necessary to look after the website and think about every change which is made on the website, anything what is done in a right way could bring visitors to the website or if it is done in a wrong way it could send the visitors straight to the competitor's website. So that it is quite important to think carefully about every step which is made to be sure that this single step will not ruin the whole work. Because the wrong step could destroy everything what has been done so far.

Search engine optimization is a method used by website developers, owners or even all members of company which helps the website to be ranked better by search engine. This could mean that the website could be placed on a better position in search engine's results. The position in search engine's result list is really important because people are not willing to go through tens of result pages due to this fact the best possible way is to be placed in the first ten results on the first page, sometimes it is easy to get there but sometimes it is almost impossible if there is no a large budget, everything depends on which keywords were chosen and how good the SEO strategy is.

Search engine optimization does not consist only of single steps which are done only once in a life time of the website. It is an ongoing process which makes people who are in charge with SEO to be still updated and still trying to guess the algorithms of search engines. They have to know about changes which were made on competitor's website and they have to react on them. This process is quiet tricky because a person or team who is trying to change the website to look more interesting for search engines never knows the exact algorithm of the concrete engine, so that it is always about guessing things which are important for the search engine and trying to use them to make the website look more attractive for this engine.

Search engines which are the most using nowadays are clever enough that they could recognize not only from which country the user is but they could recognize from which city the user is and give him relevant results. Let's imagine that the user is looking for a restaurant in his city, on search engine's result page he will see the results which belongs to his area.

There are two groups of factors which could be affected by developer and could help to improve website position in search engine's results. One group is called On-site factors and the second one is called Off-site factors. Both of these groups should be involved in the search engine optimization strategy, because, as it was mentioned previously, search engines do not look only on a website as it is, they look "around", it means they look at other pages and match the information which are found on these websites with the website which is being ranked, these factors which are not on the website are called Off-site factors, for instance, hyperlinks from thematically similar websites, hyperlinks from websites catalogues, notes on social networks etc. The other group is called On-site factors, these factors are literally ON the website, it means for example a structure of the website, title texts, meta tags, content of the website, filenames, speed of the website, request time, alternative texts for images which are not shown etc. It is easier to affect factors which are on the website, it is a bit harder with factors which are off the site, because it is not always possible to affect thousands of hyperlinks' text which led to the website or to persuade people to link more to the page, it needs more effort to improve the off-site search engine strategy.

## 2.1 User experience

Despite the fact that this thesis is about optimization for search engines, it is necessary to say that the main reason why the websites are optimized is that they want to gain visitors from search engines, so that it is obvious that the users' experience is very important. It should have never been forgotten that the websites are for users not for search engines. It is necessary to care about visitors because the better the experience is the more visitors will return to the website.

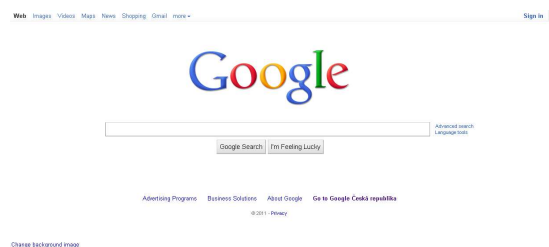
The other reason why the user experience is so important is, that this way it is possible to get back links to the website the nature way. It means for example that a visitor who is satisfied with the website will place a link, which leads to the website, on his own website. This is the best way how to obtain back links, because the website gets these links gradually, from time to time.

It is also always better to have satisfied visitors who are willing to visit the website again. There are several ways how to improve users' experience, some of them are easy to implement but some of them are not so easy to use and need more effort.

Probably the most obvious way is to have a fancy website design, because this is the first part of the web application which is seen by visitor once he comes to the website. This web design should be easy to navigate and everything on the webpage should be easy to understand, it is good to avoid using difficult webpage layout, because the user could get “lost” on the page. It is good to know how people skim the webpage, when they come to the website they do not read all the text or look properly at every image they just only skim and scan the layout. So that it should be clear at first sight what the website is about. There should not be any disturbing elements which would confuse the visitor of the web page. Let’s have a look at some examples – the best example could be Google’s website, the main goal of this website is to help users to find something on the Internet, so that the dominant element of this page is a big text box, which is for queries which are inserted by users. Another good example is Facebook’s website, the aim of this site is to register as much users as possible to give them the ability to present themselves online, so that there is only one registration form which tells users “ok, here you are, so log in or register”, there are no other elements, beside login form, which could make user confused, everything is clear at first sight.



Picture 3 Facebook’s home page – easy to navigate



Picture 4 Google’s home page – no disturbing elements

Beside bad designed website, there is another problem – users are not willing to wait for a long time until the page loads or responds to their requests. That is why it is good to find a solution how to decrease the website’s loading time and response time as well. The users

will be satisfied that they do not have to wait for a long time until the page loads. According to Jakob Nielsen a delay of 0.1second could be considered as a normal response time – for instance when the user clicks on the table column and this column highlights itself in 0.1second, 1second user feels that the website is trying to finish his request – for example sorting a table, when the delay is bigger, the user starts feeling that the application is not responding. When the delay is around 10seconds the users are starting losing concentration to the task, they feel like the application is not responding. There are several ways how to decrease website's loading time, some of them will be discussed later in chapter "3.2 Faster loading".

## 2.2 Search engines' experience

Despite the fact that the websites are mainly done for human beings, it is necessary to bear in mind that search engines look at the websites as well. So that it is necessary to think how to make the website good looking for users and for search engines as well. It is not an easy task. Search engines do not look at the page like human beings do, they do not have eyes. Search engines do not care about fancy design, they work on code level. Search engines rather care about the structure of the website, how it is written, which elements are used, how they are used etc. The only possible way how to get these information from the website is to download its source code and parse it, this is the way how the search engines work. Because of this reason it is necessary to have valid source code, because when the code is not valid, the search engine's parser could get lost in it, it means for instance that texts which should be in `<h1>` HTML tag are not in it so that the parser could not distinguish what the most important keywords are. When the parsed page is valid it will get more points than a page which is not valid. There are other ranked factors, which could help to get more points, for example search engines are quite similar to human beings in some cases, they do not like waiting, they like websites which are small, it means that they do not need to download a large amount of data to parse the page. Another possible way how to look more attractive for search engine is to have an authoritative domain suffix, such as .gov, .edu or .mil – these domain suffixes could help you to be ranked better, but it is not possible to get some of these suffixes, because for example .gov is only for United States' government websites. Search engines like original content, so that it is very important to have original

texts on the website, when there are no original texts, it could be ranked by negative points and these points could decrease the total rank of the website.

There are some problems with some kinds of elements on the website, for instance, some search engines are not able to index content which is inside flash elements, so that it is not recommend using flash menus, because some search engines could not read the text inside these elements. Another problem is caused by images, some websites use images with text as a menu items, it is not so bad like flash elements, because it could have an alt tag which describes what the image is about, but for menu items it is not recommended.

Search engines are interested in how often the website is offline, they are like ordinary users, when they try to visit the website and it is not available, they mark it as unreachable and give it some penalty points. So that it is necessary to have website on trustful webhosting servers, because once the website lost the position in search engines results, it is sometimes very hard to get it back.

The worst scenario is when the website developer tries to hide something before search engine, for example when he shows one version of website to human beings and another version to search engines, this could lead to deletion from search engine's results at all. Developers should avoid these situations. Website should have the same look, despite the fact, whether the visitor is, if it is computer program or human being.

### 3 BASIC STEPS

On today's search engine market, there are several types of search engines, some of them look for images some of them look for videos or even for books. These search engines could bring some users to the optimized website, so that it is recommended to optimize all components on the website – images, multimedia etc.. There are some factors which are more important for search engines and of course there are some which are less important, but when everything is optimized in a right way, it always helps. It would be very naive to think that optimizing only some parts of the website will bring some success, the optimization should be done in a whole otherwise it would be wasting of time and money.

Another good thing is to ensure that the content is indexable – it means that there are no difficulties in indexing. To ensure that the content is visible for search engine it is good to have all the texts on the page in HTML form, because search engines are not able to recognize whether the image contains a text or not, some advanced engines are able to find out whether there are texts inside the image or not, but the website is always optimized for more than one search engine, so it is always a good solution to have texts in a HTML form.

If the website consists of more than one page it is necessary to tell the search engine's crawler that the other pages exist as well, otherwise they would not be crawled. These crawlers use links to hop from one page to another one, this is the only way how the crawler could find a new page which might be interesting for him. So that it is necessary to make up a good link structure which would ensure that all the pages are accessible through these links. Here are some cases which could make the page inaccessible for the web crawler – page which is accessible only through the form, it means the only way how to visit the page is to fill in the form and submit it crawlers are not able to do this. This is typical for log in forms. Another case is a link in a java script. For instance - if the java script's click event is used for linking, there is no way how to tell the crawler to click on a button to call the event. Probably the most surprising reason why the page is not reachable for some crawlers is that there are many links on a page which links to the "unreachable" page. For example Google's crawler only visits from 100-150 links on a given page then it stops spidering. It is recommended to use maximal 100 links per page otherwise there is a possibility that the rest of the links will not be visited by the crawler. One solution how to tell the crawler about all the pages is a XML sitemap.

Search engines like good-quality websites, the way how they distinguish whether the website is quality or not is that they look for high-quality links which led to the website. So that it is important to have a well-organized website structure and navigation because it is the only way how to obtain links from other websites naturally. Otherwise the website could face suspicion that it uses bad SEO approaches and could be penalized. Well organized website's navigation is a must, because user should still know where he is and what to do next. If the website's structure is organized the wrong way, user could get to the place where he did not want to be and will do not know how to get back. Such confused user navigates away from that site, because he did not find any useful information, despite the fact that the information could be "somewhere" on the website. User should need as few clicks to get the wanted information as possible.

Another solution which could make it easier for both users and search engines is to have keywords-targeted URLs. It means instead of having URLs similar to this one "www.mydomain.cz/index.php?x=klhj34oik23cbe99ui" it is better to change them to more rememberable ones. If the URL contains keywords which are related to the website the users could remember it easier because the URL makes sense to them and the search engine knows what the website is about from the URLs as well. So the previous URL could be changed to something similar to this "www.mydomain.cz/videos/45". It is much easier to remember. One possible way how to find out how the website's navigation should be organized is to use technique called "card sorting".

### **3.1 Stay online**

Probably the most important step is to ensure that the website is accessible to search engines. It should be checked from more than one place, because some problems could appear - for instance, the website is accessible from one place but it is not accessible from another one, the problem could be in the connection between the server and the client or whatever.

#### **3.1.1 Server timeouts**

Every search engine has its own timeout - it is a value, which defines for how long the search engine waits until it gave it up and marks the website as inaccessible. Another reason why the website could be marked as inaccessible is a timeout generated by a server.

### 3.1.2 Blocked IP addresses

If the website is hosted on a webhosting, it is highly likely that this hosting is a shared one, it means that more than one website is hosted on one server. This could cause a big problem, because all domains which are hosted on the same server share the same IP address and could be affected. Therefore when one domain causes problems, for instance it is a spammy one, the search engine could block either the domain or IP address, or even both. Sometimes search engines block the whole IP range. It is clear that the website could get into trouble without knowing it and without its fault. There is one solution how it is possible to find out whether the IP address is blocked or not – there is a way how to use Bing to search for all websites which share the same IP address – “IP:address” query, if the IP address is not blocked the Bing returns a list of all websites which have the same IP address, but when the IP address is blocked Bing returns that it has not found anything. It is recommended checking from time to time if everything is alright.

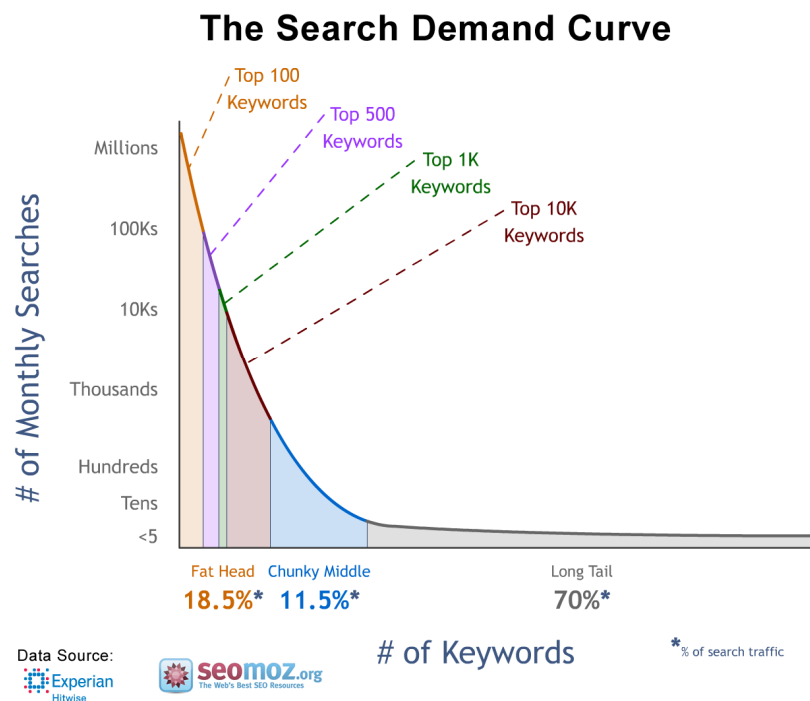
### 3.1.3 Bandwidth limits

Almost every webhosting has limits – number of emails, hard drive space, number of databases, size of database, number of ftp accounts, bandwidth etc. From a SEO perspective the most important is the bandwidth limitation. Let’s say that the webhosting sets the bandwidth limit to 10 GB/month and imagine what happens when the website becomes more popular than it has ever been before. More visitors will come, so that more requests for the website will come and more data will be transferred between visitor’s computer and the webhosting server. It is necessary to ensure that the website stays available when the bandwidth limit is exceeded.

## 3.2 Keyword research

Keywords are the words which are used by human beings to find some information, product or anything on the Internet. These terms are filled into search engine’s search bar. According to filled terms the search engine returns a list of results, each item in the list represents one webpage or website. These terms are obtained by search engines from web pages, then the search engine matches the keywords with the address of the web page, then when the user types the specific keyword into the search engine, the search engine checks the database of keywords and addresses and returns the results.

There are two types of keywords – Long Tail and Short Tail keywords. Short Tail keywords are the keywords which have 5000 searches per day, or even 500 searches per day could be considered as Short Tail keywords. It could look like these keywords are the most important to the website, but it is not true, these keywords send only less than 30% of the visitors to the website plus it is quiet hard to optimize for these keywords because many other people optimize for these specific keywords as well, the rest 70% are send by so called Long Tail keywords. The Long Tail keywords are the words which have only few searches per day or even only one search ever, but these keywords bring the majority of visitors. These keywords are usually three or more words in length. There are many Long Tail keywords, so that it is much easier to optimize the website for these keywords because it is possible to choose them from a large amount of words. For instance keyword “note-book” could be considered as a Short Tail keyword, it could bring thousands of visitors but only few of these visitors are willing to make a purchase, on the other hand keyword “DELL Vostro 3500” could be considered as a Long Tail keyword and could bring more valuable visitors who will be willing to make a purchase or will return to the website later. These visitors are better because they make repetitive traffic.



Picture 5 Long tail terms [7]

Keyword research is probably the most importing part of your SEO strategy, because if the site or page is optimized for wrong keywords it could cause that this site would never be

found by anyone. It is necessary to think carefully which keywords will be chosen. Thanks to keyword research you can predict shifts in demand, make a website which will be highly visited by customers. One way how to find out the right keywords for the website is to look at competitors' websites and try to guess for which phrases these websites are optimized. But this way will not bring new ideas, you will only get keywords which are used by someone else who thinks that these keywords are searched by users, it could be true or not. Another more creative way is to close your team into the room and tell them to think about the words which the best describe the website, it could bring new words as well, but the disadvantage is that these people know a lot about the website, they know what they would be looking for, which words are related to the website. This method could be confusing because members of the team are insiders they know the website, so that they could not pretend that they are ordinary users searching for products which are offered by the website. The best possible way how to find out keywords which describe the website would be hiring many people and let them look at the site and ask them to fill in a form about the site. This method would bring the most valuable data and the best results, but it is not often possible to do it this way, because of the budget. It would be incredibly expensive if thousands of users would be used. There is a way which is slightly similar to the last mentioned one, it is possible to use search engines records, to find out which phrases are searched by users.

Another slightly different way is to use people's mistakes. This could sound like something bad but it is not, it is just only a way how to use typos which are quite often when people do not know how to write words properly. It could be caused by different languages, positions of keys on a keyboard or simply they do not know how to spell the word the right way. It is often that people spell words the wrong way, mainly when the word is not from their mother language. Let's have a look at some examples – for example the word “university” could have many typos: univerzity – when the “s” letter is exchanged with “z” letter or it is also possible to write unyverzity etc. There are many ways how to write this word, people are inventive so it is possible to find some unusual ways how to write words.

Search engines record every searched phrase it means that every searched phrase which is entered into a search engine's search bar is recorded. So that, search engines know how often the concrete keyword was searched. Due to these records, it is possible to use tools which are provided by search engines to find out which keywords are the most searched

and how competitive they are. Despite the fact that it is possible to get information how often the keyword is searched it is not possible to say how valuable this concrete term will be for your website and how many visitors it will send to the website, because it is necessary to bear in mind that you are not the only one who optimized for this keyword, there could be thousands of other pages which are optimized for this concrete keyword as well, they could have higher budget than you, so that they could rank on better positions. Another possible way is that your competitors could buy these keywords, it means that it does not matter how good you are in optimizing, because every time user searches for these bought keywords, he will get the results with your competitors websites at first places and you will be somewhere after them.

### **3.2.1 Google's AdWords Keyword Tool**

Despite the fact that this tool is focused mainly to help to find words for paid advert campaigns it could also be used as a tool for finding words for the website for organic search results. Google AdWords keyword tool provides such information like related terms – it means terms similar to the entered one, search volume – it means how many times the concrete keyword was searched, price of this keyword – how much you will pay when you choose this keyword. There are two ways how to enter keywords, one is that the keywords are written into a Keyword Variations box and Google AdWords Keyword Tool simply mixes these words together and returns combinations of these words. It is also possible to enter keywords which should not be used during mixing. Another option is that it is possible to choose whether you want to get broad, exact or phrase match. When the matching option Broad match is set the result list will consist of words which are suitable for long tail optimization.

When the Site-Related keywords search option is set, Google AdWords keyword tool will need the URL of the website which will be used for looking for keywords. Then in result list there will be information about each keyword, for instance keyword search volume – it means how many times this keyword was looked for etc.

The screenshot shows the Google AdWords Keyword Tool interface. The search criteria are set to 'Web'. The tool suggests several keywords, and a table provides search volume and competition data for each.

Klíčové slovo	Konkurence	Celosvětový objem vyhledávání za měsíc	Místní objem vyhledávání za měsíc
onlinovky		22 200	22 200
flash		68 000 000	68 000 000
hry		3 350 000	3 350 000
minihry		90 500	90 500
zdarma		1 500 000	1 500 000
hry online		450 000	450 000
online hry		450 000	450 000
hry zdarma		201 000	201 000
logické hry		22 200	22 200
flash hry		9 900	9 900
hry sk		49 500	49 500
hry pro dívky		49 500	49 500

Picture 6 Google's AdWords Keyword Tool

### 3.2.2 Google's AdWords Traffic Estimator tool

The URL of this tool is <https://adwords.google.com/select/TrafficEstimatorSandbox>, it is quite interesting tool as well. It could help to guess how many visitors the chosen keyword will bring to the website. It is mainly focused to help to set up a paid advert campaign, so that is why it is possible to guess how many competitors use this word. Because the more expensive the word is, the more competitive the word is.

### 3.2.3 Yahoo! Search Marketing

Yahoo! also provides tool for looking for keywords. It is necessary to register before trying this tool.

### 3.2.4 Seznam.cz Keyword Tool

It is quite tricky to find this tool but it exists. It is not so powerful like for example Google's AdWord Keyword Tool, but it could still give you a brief look at how often the chosen keyword was searched. It does not give you a list of potential keywords, but it could help to find out the most searched combinations of the keywords, for instance it is necessary to get the information about the word "onlinovky", this word is simply entered into the Seznam's search bar and then the results are returned – as the result it is returned the search

volume of this concrete keyword – it means minimal search volume per day, maximal search volume per day and average search volume per day – but this is only one part of the result, another one is that it is returned the most searched phrases containing the specific keyword, in this case it means that it is returned ten other possible phrases, such as “cz onlinovky”, “onlinovky www” and so on. It is always very useful to get some other ideas and know their search volume.

### **3.3 Domain name**

One of the most important parts of the website is its domain name. It is a key factor which distinguishes whether the website will be successful or not. It does not matter if it is a personal website, company website or just a blog, it is always necessary to think carefully about the domain name, because these domain names are used by people to get to the website. There is another way how to get to the website – via IP address, but the domain name way is much easier.

#### **3.3.1 Domain name hunting**

It is always good when the domain name consists of the keywords which describe the topic of the website the best. The approach how to find a proper name is a bit similar to keyword research it is possible to use the same tools like in keyword hunting. One way how to find proper names is to write down a list of words, which are related to the topic of the website. Once the list is done, it is possible to pair the words or add prefixes or suffixes or both to create domain names. When the appropriate name is found, it is necessary to check whether the name is available for registration, or someone has already had the same idea and registered that name.

#### **3.3.2 Domain length**

Good domains are short domains it means if there are two available domain names the shorter one should be chosen. There are many reasons why to choose the shorter one – one of these reasons is it is easier to copy and paste – for example when one user wants to send the URL to his friend, he needs to copy it and then paste it somewhere, if the domain name is too long, there is a significant possibility that he does not copy the whole domain. Another reason why the domain should be short is that the domain could be written on

business cards and shorter words are easier to remember. But the main reason why the domain name should be as short as possible is there is lower possibility to make a mistake when the user is typing it in to the browser's address bar.

### 3.3.3 Domain tools

There are many tools which could help to find out whether the domain is available for registration or not, almost every domain name registrar has its own tool how to check it. Good tools are these which use the AJAX because it is more interactive and there is no need to refresh the website after each search, just only type in the wanted domain name and this type of tool returns the availability of the domain. It is possible to find such tool for example here – [www.web4u.cz](http://www.web4u.cz), [www.domjax.com](http://www.domjax.com) or [ajaxwhois.com](http://ajaxwhois.com).

### 3.3.4 Domain name recommendations

The website's domain name should be unique otherwise it could be a disaster for the whole business, because people could spell it wrong and the competitor or domain trader who owns the other name would profit from it and the user do not get the wanted website. Nice example of this was Flickr.com. Flickr only owned the domain flicker.com, but they had forgotten to register the name flickr.com, it was disaster for this server, because people used to spell it wrong and significant traffic went to the flickr.com instead of flicker.com, only one letter difference but this one letter made a huge hole in Flickr's plan, Flickr had to buy this domain, otherwise it would lost many potential users. So never use domain name which could be similar to registered domain, or buy all possible combinations to avoid situation similar to Flickr's one.

If the website is meant to be a global one, it is good to have a .com suffix. It is for reason that some people still think that .com is the only suffix which is out there, so that they automatically add this suffix to each domain. It is similar to every country concerned website, users from this country are used to use domains with country's suffix – for example people from the Czech Republic are used to .cz domain suffix, so if the website has the ambitions to be well known in the specific country, it should have also registered the nation domain name. Otherwise it would be losing type-in traffic.

Some domain names are so hard to spell that people need more than three tries to get to the wanted website. So it is always good avoid using letters which are “hard” to pronounce.

These letters are q,z,x,c,p but everything depends on language, because each language has its own letters which could be misheard.

Domain names should be not only as short as possible but they should be as easy as possible as well. By easy is meant, that the name should not consist of numbers or hyphens or even Roman numbers. This could be confusing for users for many reasons – it could be hard to remember when the domain is like this one “my-web-site.com” – there is too many words and hyphens in there, another good example of a bad domain name could be “3webtutorials4dummies.com” – this domain name could represent “free web tutorials for dummies” but it is very hard to remember that there are two numbers which represent words not numbers – 3 = free and 4 = for. It is possible to use this type of writing in texts on mobile phones, but it should not be used in domain names, because this is only meaningful to native speakers or people who know the modern trends in writing texts, other people could not get the “message” which is hidden in this domain name. The worst case scenario is, when both of these bad techniques are used together for example “3-web-tutorials-4-dummies.com” this address would be hardly rememberable and much traffic would go to some other websites with similar domains.

### 3.4 URLs

Search engines like keywords in URLs, they place some weight on them. But sometimes when the URL structure is badly designed it could be a signal that it is a spammy website. URLs should be self-explaining it means when user looks at the URL he could easily identify where he is and what the content on the website is about. Nice example of this is when one user sends a link to his friend then the friend looks at the browser’s address bar and knows what the website is about.

#### 3.4.1 URL length

The shorter the URL is the easier it is to remember or paste. This is similar to domain names – URL should be as short as possible because there are many situations which need short URL, for example shorter URL always looks good on business cards than the long one.

### 3.4.2 URL types

There are two types of URLs – static and dynamic, each of them is treated differently by search engines. Dynamic URLs consist of two parts – one part specifies the file and the second part specifies variables and their values, here is a sample of a dynamic URL “www.mywebsite.com/index.php?variable1=value1&variable2=value2”. The part which specifies the file is www.mywebsite.com/index.php, the part which specifies variables and their values is “variable1=value1&variable2=value2”, question mark is a delimiter which separates the file part and the variables and the ampersand is a delimiter which separates variables.

On the other hand static URLs consist only of the file part it means that there are no variables and their values, for example “www.mywebsite.com/index.html” is a static URL.

Search engines could distinguish whether the URL is static or dynamic, they like static URLs more, because there are no symbols like a question mark, an ampersand or an equal mark. Some engines even ignore the part which is after the question mark.

There are tricks how to make dynamic URLs look like static ones. One technique which could translate dynamic URLs into “static” ones uses a .htaccess file and an Apache module mod\_rewrite. But this approach is available only on Apache servers.

### 3.4.3 URL characters

Most of today’s servers are Linux/Unix-based servers, so it could make problems sometimes, because Unix-like systems are case sensitive, it means that folders “Product” and “product” are two different folders despite the fact they are different only in one letter. This could be very tricky, because when the proper URL is “www.mywebsite.com/Home/Product/index.html”, it is not possible to access the index.html file by any other way than get in to the “Home” directory and so on, so when the user types in to the browser’s address bar URL like this “www.mywebsite.com/home/product/index.html” he gets in to a totally different directory – in to “home” instead of “Home”. The recommended approach is to use only lowercase characters it is very easy to write “home” instead of “Home”. Users are used to use only lowercase characters when they are typing in the address. If the website already has case-

sensitive URLs, the URLs with uppercase characters should be redirected to ones which are in lowercase characters.

#### 3.4.4 URL recommendations

URLs should use as less folders as possible, it means that content should not be hidden somewhere deep in the website structure – like this “/folder1/folder2/folder/3...../folder56/index.html” – this is not the right way how to do it.

When the name of the folder in the URL consists of more than one word, these words should be merged together by plus sign or hyphens, anything else could confuse search engines, the most using way how to join words together is by hyphens. Search engines could recognize that it is more than one word and find out the other words from the name – for instance “my-home-directory” is a nice example how it should look like.

Some URLs have many additional data, which are not necessary and when these data are removed it still points to the same webpage as the URL with these data, disadvantage of this is users are clever enough to find out this trick and will link to the page by the shortened version of the link, but some of them will link by the proper version of that link, what does it mean for the website? It means nothing good for the website, because these two URLs will point to the same content, literally to the same page. This could cause a huge problem because this content could be considered as a duplicate content by search engines and this is what the most sites are trying to avoid, it does not matter that it is only one page, there are two different URLs with the same content – this is how the search engine sees it and it could penalize the whole website, because of it.

If the website will be divided into several categories, it is necessary to make up how these categories will be represented in URL, one way is to assign a number to each category so the URL would look similar to this one “www.mywebsite.com/1/2/index.html“, of course it is possible to do it this way but there are two main disadvantages - the visitors would not know what the number one means in the URL, search engines would not know either and instead of the number there could be a keyword which would bring more benefit to the whole website. So the solution is to put there a keyword which would be more descriptive than the number, let’s have a look how this example could be changed to look more familiar to users and to search engines and to make more sense as well -

“www.mywebsite.com/shoes/sandals/index.html” – this is far more descriptive than numbers.

### 3.5 Faster loading

The time which is needed to download the website is one of the key factors which affect the ranking result. The faster the site is the higher amount of points which are necessary to rank higher are obtained. Search engines bear in mind this time. It is also good to have a faster website because it is much more pleasant to visitors. They do not have to wait for a long time until the website loads they can rather enjoy the site instead of waiting time. The key factor how to improve the performance of the site is to make lower requests to the server. These requests are made to get the content of the specific page – this content is for instance – images, scripts, styles. Another way how to decrease the loading time is not to download the content which has already been downloaded, this approach is called caching this is useful for returning visitors, because they will use scripts, images or style sheets from the browser’s memory. The last way which will be described in this chapter is compression this way is used when there is no way how to avoid downloading the content, in fact compression will not decrease number of requests necessary to get the content, it will decrease the time of downloading of the content, in most of the cases significantly.[3]

#### 3.5.1 Content delivery network

CDN (Content delivery network) is a network which delivers content to the users. These networks are here because users can access your website from the whole world and they are not located at the same place, if they are not at the same place it means different times for packets to arrive from the server, because of the different distances. Lets imagine that you are accessing website which’s servers are located in Washington in the USA from Prague in the Czech Republic, a distance between these two cities is approximately 6900km, all packets have to travel this way because all the content is located in Washington, this traveling takes time which makes the user to wait until the whole page loads. To decrease this time, you should use CDN – this network consists of servers, which are situated in many cities over the globe. So when the user wants to visit your site, she will download the content from the closest server. Let’s use the example which was described before – the visitor is in Prague, the server is in Washington, but now CDN is involved in this scenario. Let’s

say that this CDN has one of its servers in London in United Kingdom. Once the visitor accesses the site she will get the HTML document from the original server but the content such as images, scripts, style sheets, flash files etc. will be downloaded from the CDN's server situated in London, the distance between London and Prague is about 1000km which is almost seven times less than the distance between Prague and Washington. The time which is needed to download these files from London is lower than the time necessary to download this content from Washington. CDNs are usually good choice when you have visitors from different locations, but there are situations which make solution with CDN slower than without it – the main problem is delay which is made by the way between the selected CDN server and the visitor, there could appear problems with routing, so that the data will travel longer distance in spite of the fact the geographical distance is shorter. [1][2]

Let's have a look at the most visited websites on the globe, which CDNs they use.

Website	CDN
<a href="http://www.amazon.com">http://www.amazon.com</a>	Akamai
<a href="http://www.aol.com">http://www.aol.com</a>	Akamai
<a href="http://www.cnn.com">http://www.cnn.com</a>	
<a href="http://www.ebay.com">http://www.ebay.com</a>	Akamai, Mirror Image
<a href="http://www.google.com">http://www.google.com</a>	
<a href="http://www.msn.com">http://www.msn.com</a>	SAVVIS
<a href="http://www.myspace.com">http://www.myspace.com</a>	Akamai, Limelight
<a href="http://www.wikipedia.org">http://www.wikipedia.org</a>	
<a href="http://www.yahoo.com">http://www.yahoo.com</a>	Akamai
<a href="http://www.youtube.com">http://www.youtube.com</a>	

Table 1 The biggest players' CDNs [1]

### 3.5.2 CSS sprites

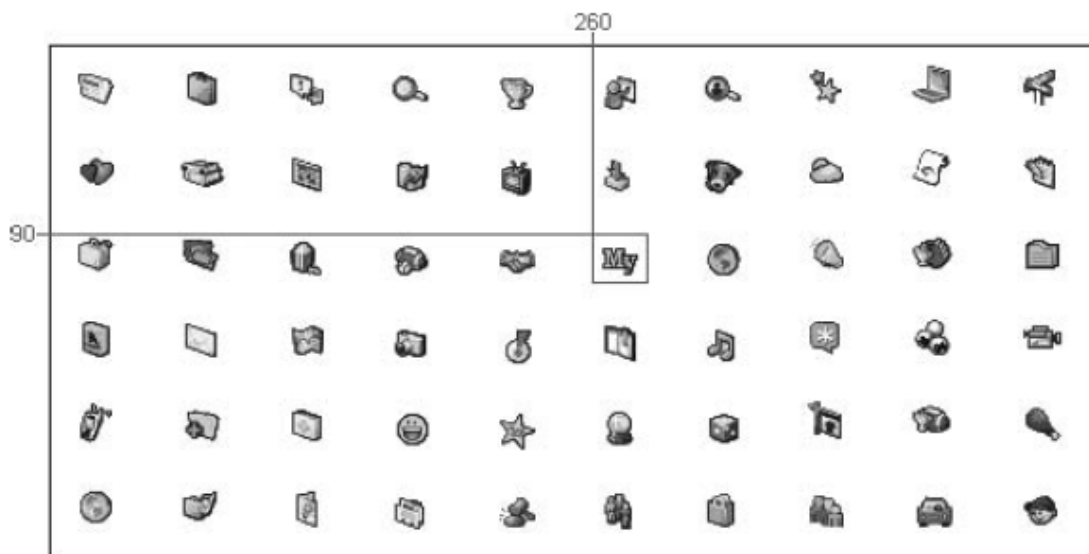
CSS sprites are used to decrease the number of requests necessary to download the images. This approach uses CSS tags which tell the browser from where to where the background image for the specific element is.

Let's imagine that you have fifteen DIVs on the page, each of these DIVs has image as a background, this background image is specified in style sheet file, each of these DIVs has its own ID and this ID is used in CSS file to set the background-image property. Let's say that DIV number one with ID="one" has image "one.png" set as a background, DIV num-

ber two with ID="two" has image "two.png" set as a background, DIV number three with ID="three" has image "three.png" set as a background and so on until DIV number fifteen. It means there are fifteen DIVs and fifteen images. So that the user's browser will need to make fifteen requests to download all these images.

When you combine all your images into a table, you will get only one image file. Then you will use the CSS tags "background-position", "width", "height" to specify which part of the matrix you want to use.

So let's use the same scenario as before, but instead of fifteen separated image files, we will have only one, which consists of the fifteen images aligned into a matrix. Then the browser will download only this image which costs only one request and then CSS will do the rest of the work. This approach will significantly reduce the time necessary to load the page.[1]



Picture 7 CSS sprite [1]

### 3.5.3 Compression

When we have no way how to decrease the number of requests necessary to download the page's content, it is right time to involve compression.

#### 3.5.3.1 Compression during delivering

This type of compression reduces the size of the HTTP response which is returned to the user's browser from the server. To use this type of compression you need to change set-

tings of the web server. Nowadays the most popular compression method is Gzip. In fact it is not literally compression during delivering, the data are taken, compressed and then send to the client in compressed form, on the client's site data are uncompressed and used by browser. This compression method has its pros and cons. Pros are that there is less data to transfer from the server to the user, this means the user is not made to wait for the content. Cons are that the process of compression and decompression takes time, and there are issues with some browsers and proxy servers.

Gzip compression is done by module which is configured to be used by an Apache web server. The name of this module has been `mod_deflate` since an Apache 2.x and `mod_gzip` in an Apache 1.3. When this module is used, the user will see the page much faster. [1]

Here is a table which shows the savings achieved by a `mod_deflate`.

Example	Components (HTML, CSS, JS)	Total size	Size savings	Response time	Time savings
Nothing Gzipped	48.6K, 59.9K, 68.0K	177.6K	-	1562 ms	-
HTML Gzipped	13.9K, 59.9K, 68.0K	141.9K	34.7K (19.7%)	1411 ms	151 ms (9.7%)
Everything Gzipped	13.9K, 14.4K, 18.0K	46.4K	130.2K (73.8%)	731 ms	831 ms (53.2%)

Table 2 Page weight savings for different levels of compression [1]

### 3.5.3.2 Compression during development

During development time it is possible to reduce the amount of data which the user will be made to download to see the page. There are no doubts that the largest items which are downloaded by a client when she visits our site are images. Designers of the site could not miss this fact. There are two main types of images on the page. These two types are photos and graphics. Both of these types should be compressed, there are several ways how to do that. Graphic files have much “unnecessary” information inside, for example it is possible to remove information such as the author's name, it is good that the user knows who the author of the image is, but he does not need to know this to see the image. Another quite often used method is to decrease the number of colors (color depth).

### 3.5.4 Content caching

Today's web pages include many components so that it is necessary to reduce time needed for their downloading, or find other way how to get the requested content without down-

loading it at all. This solution only helps to decrease loading time for returning visitors first-time visitors do not get any profit from this solution. If the developer cares about returning visitors he should use caching, there are several ways how to do it. There are two possible types of visits – a visit with a primed cache and a visit with an empty cache. A visit with an empty cache means that the user has never visited the website and a visit with primed cache means that the user has already visited the website. The impact of this improvement depends on how often visitors come to the website with a primed cache. According to Tenni Theurer's research it was shown that 40-60% of unique users have a primed cache, these users could recognize the differences in loading times. The same study showed that the number of page views with a primed cache was 75-85%. Caching is mostly used with images, scripts, style sheets and Flash, but it should be simply used with all components on the website. In other words any component which changes infrequently should be cached. For example a HTML document should not be cached because it contains a dynamic content. One way how to distinguish whether to cache the component or not is to check "Last-Modified" dates, these dates could help to decide. [5] [1]

One problem with caching is that the browser's cache remembers the filenames, so that even when the content of the file is changed and the browser has already have cached version in its memory, it uses the cached version from local pc instead of the new version from the server, because it just only checks the filenames not the content of these files. One solution how to deal with this problem is to use filenames with versions. Let's have a look at an example – the CSS file has a name "style.css", so when the user comes to the website the browser downloads this file and saves it on the local hard drive, next when the users comes the browser only checks whether to use the stored file, if so, it uses the file from local memory instead of downloading it from the server. But when the changes are made in the CSS file stored on the server and the user comes after these changes, his browser will not know anything about these changes and will use the file stored on the local drive instead of downloading the new changed one. So if there is a need to make browsers always download the newest CSS file, the filename should consist of the name + version – for instance instead of a simple "style.css", the filename could be "style\_1\_0.css" and when the changes are made, the file could be renamed to "style\_1\_1.css". This makes browser to download the newest file from the server.[1] [2]

#### 3.5.4.1 *Expires header*

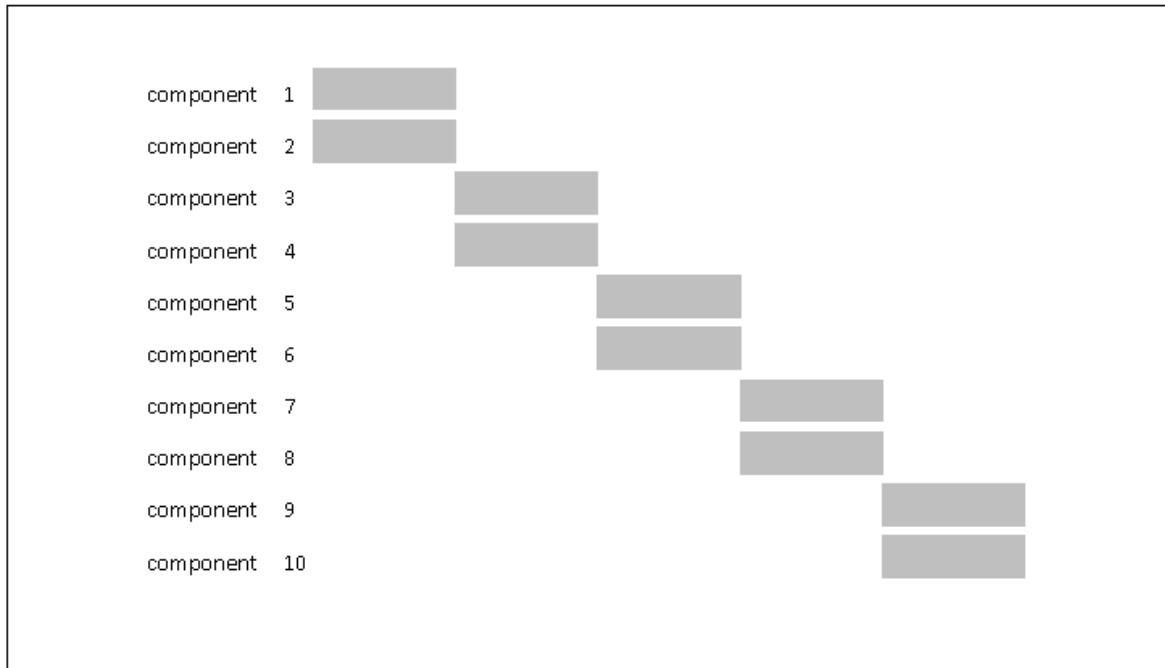
Browsers and proxies save the content to reduce the loading time of the page. The web server could tell the browser until which date the saved content is valid, if the date is in the future, the browser uses the cached copy of the content, but if the date is in the past, it downloads the content from the server. This approach reduces the number of requests and the time needed to load the page. The reduction is gained due to the far future expires header – it tells the browser that for instance the image will be valid until April 15, 2020, until that date the browser will not download the image again, it will use the cached one. This approach is good but it has two disadvantages – it uses the time taken from the client's pc, so that the server's time could be different to client's one – for instance, the real date is September 10, 2011, expire header is set by server on September 13 2011, but client's computer has the time set to the September 10, 2012, due to this fact it downloads the content every time he visits the website, because the client's pc thinks that the content is outdated. Another big advantage is, what happens when the expiration date comes, it is necessary to manually change the server's configuration to another expiration date, this change has to be done, otherwise every single file, previously downloaded from the server, would be considered as outdated, so that it would be necessary to download each file once again, every visit.[1]

#### 3.5.4.2 *Max-age directive*

One solution how to overcome disadvantages of expires header approach is to use a max-age directive. This directive tells the browser for how long the cached items are stored since first download. The directive has only one parameter – time, which is in seconds. This parameter defines for how long the content is valid. If this period of time is exceeded the component is downloaded again with a new time stamp. It is possible to combine both approaches – expires header and a max-age directive, because browsers which do not support HTTP/1.1 ignore the max-age directive. If both solutions are involved, the max-age overrides expires header. [1]

#### 3.5.5 **Styles, scripts position**

Some components from the HTML document are downloaded in parallel it means more than one component at a given time. A position of component could affect this behavior.



Picture 8 Components loading

Style sheets are components which are downloaded in parallel. A position of style sheets is a key factor which tells the user that the website is still loading. Because if style sheets are on the wrong place, the user could start feeling that nothing is happening and the page is not loading. There are two main options where to place the style sheets – at the bottom of the HTML document or at the top of the document, the position of the style sheets in the page does not affect the download time, but it affects rendering. The problem with style sheets is when they are placed at the bottom of the page, because it causes that the page stays completely blank until the style sheets are loaded, once they are loaded, the whole page appears at once. This could confuse the user. The better place for style sheets is at the top of the HTML document, in the “head” using the “link” tag, because the style sheets are loaded at first and then the browser knows how each element of the page looks like, so there is no need to wait until the whole page is loaded, the progressive rendering could be used. The advantage of this approach is the user has as feeling that the page is responding and something is happening. [1]

Scripts are components which are not downloaded in parallel, it means that only one script file is being downloaded at given time. A position of scripts could also make problems, because progressive rendering is blocked for all the content below the script. For instance, if the script file is in the middle of the HTML document all the content below this script

need to wait until the script is loaded, the rendering is stopped, once the file is loaded the rendering could go on. The thought is to put the scripts as low as possible (if it is possible), because the more content is rendered before the scripts are being downloaded. The best position where to put scripts is at the bottom of the HTML document, after all components, because loading of these scripts do not affect anything else. [1]

### **3.6 On-site factors**

On-site factors are the factors which are literally on the website. Search engines are “only” computer programs which use information which are publicly available, so that when they visit the website, they expect to find elements which they could use to rank the whole website. What these elements are will be described in this section. The whole point of an on-site SEO is to make the website look good for search engines they want to find keywords, meta tags, titles, alternative texts and many more elements to rank the site properly. First engines used to rank websites only according to what they had found on them, in these times, there were no off-site factors, the whole point was to develop a website which was search engine friendly, that meant insert keywords into every HTML tag. Today’s search engines are much cleverer they could recognize whether the website has any added value for the user or not, they count how many times the keyword appears on the page, if it is too often, the whole site could get lower points than the website which has only few of them, but on the right place.

#### **3.6.1 HTML title tag**

This tag is inside the head section of the HTML page. According to Seomoz’s survey, 72 SEO specialists think that this tag is the most important factor when Google ranks the webpage. It should contain keywords and a short description of the website – but really short. If the website does not have this tag, search engines could rank it, but it gains only few points, so this tag should be always used. Each page of the website should have its own title text it means there should not be two exactly the same titles. The title tag could also help users, because it is displayed in the browser’s bar, or some browsers display it on a tab, in the tab bar, so the user still knows what on the tab is, even when he is on the other tab, just one look to find out where the page is “hidden”. [3] [4] [6]

Here is an example of the title tag written in HTML.

```
<title> Here is the content of the title </title>
```

### **3.6.1.1 Keywords position**

The best position where the keywords should appear is at the beginning of the title tag. Sometimes it is possible to see that some websites have at the first position of the title its name, it is also possible to do it this way, but it is rather for branding reasons than for SEO reason. If the brand is not very famous it does not matter where it is, so the best position in that case is to place it at the end of the title tag – it is still in the title but does not have such a big impact. When the brand name is inside the title tag, it should be separated from the rest of the title – for example some special characters could be used for this – “>”, “|” or “-” anything which clearly separates the brand name from the title is possible. This character does not affect ranking, it is rather for esthetic reasons – it is more readable for users, they could distinguish what the title is and what the brand name is, because it is sometimes very hard to find out this information. [8]

### **3.6.1.2 Title length**

In fact the length of the title tag is not so important for the website itself, it is important for users, because the title text is sometimes used as a description text on the search engine’s result page. So it is recommended think carefully what the text is. Due to the fact that the text is displayed on the search engine’s results page, the length of this text has to be limited. The recommended length is 65 characters including spaces, the rest characters are ignored by the search engine. Some search engines allow a longer title text, for instance Google allows up to 70 characters title text, in some cases, but SEO is not about optimizing only for one specific search engine.[4] [6]

### **3.6.1.3 Title tag content**

Each title text should include keywords, but it is not all about keywords. It should be descriptive, so when the website is meant to be an ecommerce website, it should be clear for the first sight – sample of a title text for an online shop selling cars – “Buy a car – the best available cars online” – the user could see that this site sells cars online. It is possible to make experiments with the texts and then use the text which fits the best and brought the

most visits to the website. The title text should be similar to the one which would be used for paid ads – descriptive and short. [9] [3]

### 3.6.2 Meta tags

Other quiet important tags are meta tags, mainly the description tag, which describes the website. It is used to describe the website, more concrete than title tag, more accurately. It could be a short advertisement for the website. Sometimes it is very hard to write about the website in a way which attracts users. The description tag should consist only of true information because search engines are clever enough to find it out.

In early times a quite import tag was the “keyword” tag, but webmasters used to use this tag, they filled it by keywords which were not related to the website, so that search engines stopped using this tag. But it does not matter that it should not be included in the source code. [3]

Another tag, which could sometimes help search engines distinguish in which language the website is written, is a “Content-language” tag. This tag is used only by some search engines, for example Google does not use it, it finds out in which language the website is from the content. But it does not make it worse when the tag is there.

There are more meta tags, but they have not any impact on the ranking. These tags are “generator”, “author”, “cache-control”, it is also possible to define a new tag, for instance “test” it could be used by programs which read the page’s source.

#### 3.6.2.1 Description tag length

It is quite similar to the title tag – the description has to be maximal 160 characters length, due to the fact that each search engine shows a different number of characters – Google shows 160 characters, Bing more than 200 and Yahoo! 165 characters, so the lowest number is 160, because of the Google, the description should be maximal 160 characters long, otherwise the website risks that the description will not be shown all. [6]

### 3.6.3 Heading tags

Heading tags are the tags which begin with H in HTML document. These tags should be hierarchically ordered, it means the most important headline should be in the H1 tag, the

less important content should be in the H2 headline tag. The hierarchy is similar to a structure of the newspapers – the most important text is the name of the newspapers – so it could be a H1, the next important texts are the headlines of sections – it could be a H2, next important texts are headlines of the articles – so this could be a H3 and so on.



Despite the fact each H tag has its own predefined font size, font weight etc. it is possible to redefine it to look as the website developer wishes. So the H1 tag is normally the biggest one – the biggest font size, font weight and so on, everything is huge, so that there is sometimes a need to customize it to fit in to the website. It is possible to make it look as the website developer wishes because search engines do not look how the tag looks like they just only check what inside this tag is. CSS styles are used for customization, only user sees the difference between customized tag and uncustomized one. [3]

### 3.6.3.1 H1 content

Sometimes it is possible to use the same text as is used for the title with keywords. But it is not possible when the title text is longer, because the H1 tag is in a position where only short phrases fit. In these cases it is recommended to use a shorter heading tag with the most important keywords from the title tag.

### 3.6.4 Images

Images are elements which make the website look more attractive to users, they are parts of design. These elements are not readable by search engines, they has to be described some-

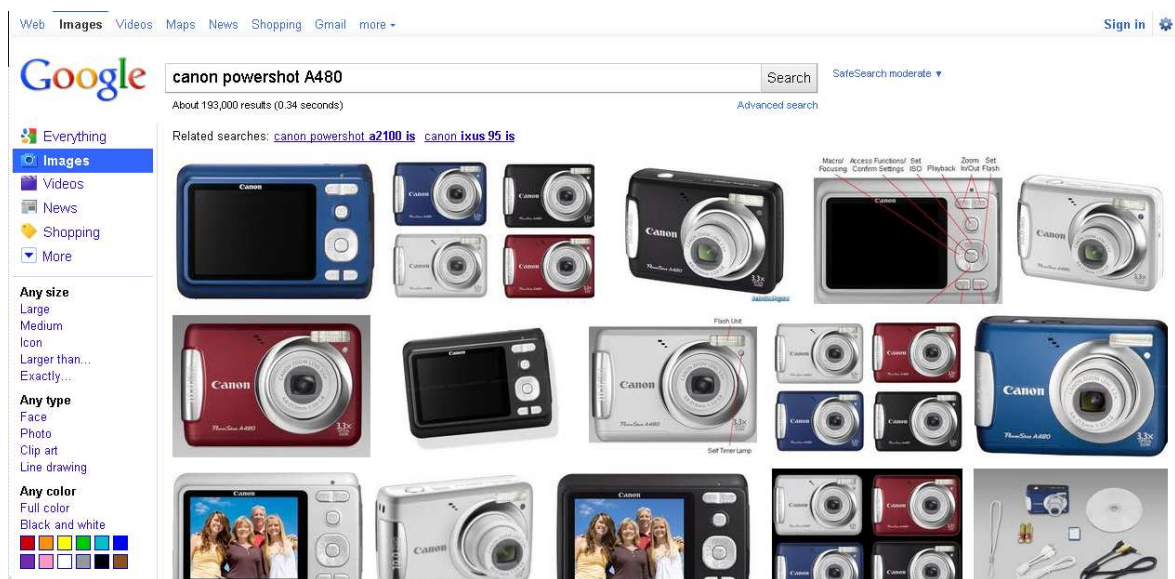
how else. There are two properties of the images which help describing them. One of this attribute is an alternative text – this text is used when the browser is not able to show the image, showing of images is disabled or for blind people which use special software for browsing the Internet. This attribute should be descriptive, it should describe as much as possible, let's say there is an image of a camera “Canon Powershot A480”, so the alternative text could be “camera Canon Powershot A480” it is descriptive enough, so that users could easily identify what the image is about. Search engines could identify it as well they could not look at the image they use the alternative text for that reason. [3]

Another attribute which is used by search engines to find out what on the image is, is its file name. This name and a path to the file is in the attribute “src” of a HTML tag “img”, it is an opportunity to use keywords in the file name because the search engines notice it.

```

```

These attributes are used by special search engines – these engines look for images and then users are able to find them. So that these attributes could bring some benefits, because there is a space for keywords, which could bring users to the website. For example one of these types of search engines is here “images.google.com”.



Picture 9 Google's images search

### 3.6.5 Website structure

Each website should be as easy to navigate as possible, because users do not want to get lost at the website. They want to find information that is all. For websites which have less than 10 000 pages, all the content should be accessible through maximal three clicks from the home page. For websites which have one million pages, all content should be accessible through maximal five to six clicks otherwise the user could get lost or could lost interest of the website. One solution how to find out how users navigate on the website is called “card sorting” – it is a technique used to find out which item belongs where it is done from users’ perspective. All items are written onto a set of cards, the group of users then order the cards in the logical flow, into groups that seem to fit together. There are two types of websites’ structure – deep and flat. Website’s structure should be as flat as possible. [8]

#### 3.6.5.1 Flat structure

When the website structure is flat, it means that there are no more than three clicks from the home page to each page on the website. This structure reduces the number of page visited by the user until he gets to the wanted information. When this structure is used, it is necessary to have less than 100 links at each page, because search engines are not willing to follow more than 100 links.

#### 3.6.5.2 Deep structure

Deep structure is like a flat house, it has levels which represent related items. The entrance to this house is a home page then the user needs to go through many links to get to the wanted information. Each level opens a new set of links. Typical for this structure is that the user needs to click on more than five to six links, sometimes even more.

## 3.7 Off-site factors

Off-site SEO approaches are techniques which are not used directly on the website, but somewhere else – by somewhere is meant on other servers. One of the most important elements which are located on other servers, are links. The first search engines had only

ranked the website according to on-site factors, then Google came up with its idea to rank the website according to its popularity, by popularity is meant the number of links – the more popular website the more links lead to this site because its content is more valuable for the users which link to this site. In these days it was quite easy to cheat on the search engines, because it was only about the amount of links, now it is a bit difficult. Everything counts. So the final position in the search engine's results is a result of many factors – some of them are on the site, others are off the site, the final result is combination of both. [3] [8]

### **3.7.1 Backlinks**

Backlinks are the links which lead to the website. It is one of the most ranked factors. But not all the links have the same value, some of them are more valuable than others, it depends on which site they are placed on. The key factors which distinguish whether the link is valuable or not is the topic of the site or if the website is authoritative enough to bring more points.

#### ***3.7.1.1 Backlinks hunting***

There are many ways how to get links which point to the website. The golden rule of this method is to get as many links from the websites which are related to the optimized website as possible. One way is to have interesting content, so that users link on the website naturally themselves, but is it sometimes hard. Sometimes there has to be a little “help” which helps to get links. One of these approaches is to register the website into the online catalogues, these links will not be high quality but it could boost the website's start, then users will find out that the website exists and start visiting it and linking to it. Another way is to exchange the link with your competitor – site owner number one links to the site owner number two and the owner number two links back to the owner number one. This approach could be used but not very often, because some search engines could consider it as a barter trade which is sometimes penalized. Another way how to get links is to have a blog and add a link into articles which are related to the website. [3]

### 3.7.1.2 Backlinks' anchor text

An anchor text is the first sign which tells search engines what the web page is about. When the search engine finds a link, it adds it to its database. The advantage of the anchor text is, for example when the web page is written in Flash. Search engines are not able to crawl websites written in Flash, so the only way how to guess what on the page is, is the link's anchor text. It is a text in a HTML tag "a".

```
<a href src="mywebpage.html">ANCHOR TEXT</a>
```

The more backlinks the website has, the more points it gets in search engine's algorithm. The anchor text should consist of keywords, which are related to the website.

An importance of backlinks and the anchor text was shown in October 2003, when a group of people decided to use this power to express what they think about American president candidate George W. Bush. It was during an election period when the first effort appeared on the Internet. It invited people who own a blog, website or just only one page to join this kind of protest. They decided to link to the White House biography of George W. Bush, it would not be anything unusual if the normal anchor text would be used, but they decided to use "miserable failure" as an anchor text, during next six weeks people placed such many links on their websites that when someone tried to look for "miserable failure" on Google, the first result which was returned was the webpage of George Bush biography. This approach was named "Google bomb", few years after Google changed its algorithm to avoid this type of manipulation, but it is still possible to use other search engines such as Yahoo for this trick. For example when the "miserable failure" is entered in to Yahoo, biography of G. W. Bush is on the third place. Another "bomb" was launched in April 2004, it was against Ken Jacobson, the anchor text was "waffles" – it was as successful as the previous one, the "miserable failure". [3]

### 3.7.2 Social media marketing

The social media marketing is a form of marketing which promotes the website on social networks like Facebook, MySpace, Twitter, Digg, YouTube and so on. The biggest advantage is that on these networks it is possible to get in touch with many people. These networks provide the ability to share comments between users, so that when one user writes something about the website in his comment, others will see this comment and could react

on it. Beside the comments these sites allow users to upload videos, so there is a possibility to upload a video which promotes the website and the users will find the promoted website.

[8] [3]

## 4 BLACK HAT OPTIMIZATION

The black hat optimization is an approach which is a bit controversial. These techniques are used to move the website in search engine's result a bit up. It could be compared to a forbidden competition – if one trader uses techniques which are against good manners, he is sometimes penalized by authorities. But it does not work the same way with search engines, because they are run by private companies, so they are not able to punish the sites which use the techniques which are against good manners, because these techniques are not forbidden by authorities. There have to be some rules, what is allowed and what is not allowed. These rules are sometimes publicly available, for instance Seznam.cz has its own list of these techniques which tells the user what not to do to stay in the Seznam's results. The list is divided in to two parts – forbidden techniques and not recommended techniques. The not recommended techniques could cause that the website could be penalized – it means that it adds some negative points during ranking, so that the website loses its position in search engines' results and is moved somewhere far away from the first page of results, but is still present in the search engine's results. The worse case is when some forbidden techniques are involved, when the search engine finds it out, the website is completely removed from the results. After this removal, it is very hard to get back in to the results, so the recommendation is - use only techniques which are proven, otherwise it could ruin the whole work. The websites which are dependant on visitors from search engines should be very careful which techniques are using in other words they should not use any techniques which might remove the website from the results.

### 4.1 Not recommended techniques

These techniques do not cause that the website is removed from the search engine's results, but they cause that the site is moved far away from the first page of results. It is sometimes similar to complete removal, because when the majority of visitors come from the search engine, the website loses visitors and the website without visitors is totally useless, because as was mentioned in the first part of this thesis, websites are done to gain visitors.

Some of these techniques could be tricky, because they could be used for a good SEO which helps the website in ranking, but if they are used wrong, they could be considered as bad SEO techniques, the border is sometimes very narrow.[3]

#### 4.1.1 Duplicate content

A duplicate content is a content which is placed on more than one site. It is same like in the academic world. If someone tries to persuade people that the work which has been done by someone else is his own work, he is usually punished. This is same in online world, when someone tries to use texts which are from someone else's page, the search engines recognize this behavior and penalize the whole website, because it is not useful for users to find the same text on more pages, it is nothing new for them.

The duplicate content is a content which is accessible by more than one URL. This could happen when the URL structure is badly designed, for example when there are parts in URL which are unnecessary. Users find out this fact and start linking on the page by two a bit similar URLs, they could look similar only to users, search engines usually have different opinion – two similar URLs are still two URLs, does not matter how much similar they are. So that the content is accessible from two URLs and might be marked as duplicate one. The solution for this situation is to redirect one URL to the other one this tells the search engines that these two URLs are in fact only one URL. [10]

#### 4.1.2 Useless content

The content which is marked as a useless content is usually content which is auto-generated by programs or is abnormally filled with keywords. This content is a text which has a high keywords density, it means that the keyword is repeated every few words. Another content which could be considered as a useless content is a text which does not thematically match to the rest of the website, it is clear that this text is there only for reason to improve the search engine's ranking of the website. This text has no value for users because it is there mainly for search engines, but when they find out that this is the only reason why the text is there, they penalize the whole website, because it is cheating, no one wants to read auto-generated text without meaning.[3] [10]

#### 4.1.3 Backlinks business

Backlinks trade is a method used for getting more links which lead to the website. The key factor is that there are companies out there which provide special bars, which are filled with links. This bar is simply included in the HTML page and thanks to this bar links which lead to other sites, which also use the service which is offered by the company, are

shown. The link which leads to the website is also shown but on other websites. The links should not appear this way they should appear by the natural way, one by one. It is usually very easy to identify these bars on the page, because links which are inside are not usually thematically similar to the webpage. [3]

## **4.2 Forbidden techniques**

Forbidden techniques are techniques which in most of the cases lead to complete removal from the search engine's results. These techniques should never be used otherwise the website loses its visitors from the search engines. These methods are usually these which cheat on the visitors. If the website uses these techniques it does not deserve to be in the search engines results, because in most of the cases these websites are not valuable for users. [10]

### **4.2.1 Hidden text**

A hidden text is a text which is somehow invisible for users. It is usually a text which is intended to be only for search engines' robots. The invisibility could be achieved by many methods. First of them is probably the easiest to achieve – the font color is exactly the same like background color, so that the user could not see anything beside the background, but on the background there is a text. Robots see this text, because they look in to the source code of the HTML document. The font color and background color could be set by CSS style. Another technique also using CSS is that the text is completely hidden. Each element of the HTML page has a CSS property – “visibility” which could be set to “none” and the text is hidden when the page is rendered by browser, but search engines do not care about CSS style, so that they could see everything what on the page is. Another approach which is using in newspapers or even on contracts is a small text. The small text is a text which is hardly to read for ordinary human being. Its font size is set to very low value, so that the user is not able to read the text.

In fact only few of these methods literally hide the text, but all of them have something in common, they try to manipulate search engines' robots by giving them something what is “hidden” before ordinary user. [10] [8]

### 4.2.2 Cloaking

Cloaking is a method which is used to manipulate the robots. By cloaking is meant that there are two versions of one webpage, one of these versions is for search engines' robots and the other is for human beings. The version which is intended to be for a robot could be totally or partly different from the one which is for visitors. Despite the fact that this method needs the most effort it is still one of the most using. It uses the information which is provided by the robot which goes to the webpage the information is in the request for the web page, once the algorithm, which is in charge with page switching, finds out that the visitor is not a human being, it sends the version which is intended to be for robots. [10]

### 4.2.3 Fake links

Fake links are links which send the user somewhere else than they claim. Normal links use HTML property "href" which defines the target of that link.

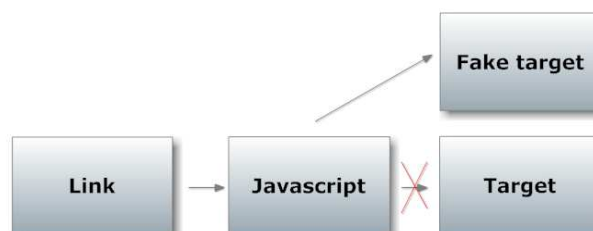
```
<a href="google.com">Google.com</a>
```

Once the user/robot clicks on such a link, he is redirected to the URL which is specified in the source code.

Fake links use the ability of javascript to manipulate the website elements. Once the user clicks on the link he should be send to the URL defined in the "href" property, but thanks to javascript which catches the click event, once this event is caught it is not so hard to redirect the website wherever the developer wants. [10]



Picture 10 Links normal behavior



Picture 11 Fake links behavior

## **II. PRACTICE**

## 5 WEBSITE OPTIMIZATION

This part of the thesis is focused on optimization of a website, to make it more valuable and friendly for users and for search engines as well. It is shown in this part how the previously mentioned techniques could be used to improve user experience. At first the website was launched without these changes and then after the changes so that it was possible to compare these two versions of one website.

### 5.1 Webhosting selection

The whole application is written in a PHP language and it uses MySQL as a database server. This combination is very famous, so that it is not so hard to find thousands of webhosting providers. The website is meant to be mainly for Czech and Slovak visitors, but the majority of them will be from the Czech Republic, because of that the first criterion was, that the server has to be located in the Czech Republic, the second requirement was that the server has to support PHP and MySQL.

The first webhosting provider was chosen Kapusta.cz, because of the positive references which were found on the Internet. This webhosting was quite new and everything looked like it is a good choice for the beginning – servers were located in a datacenter which is in Prague. This webhosting offered 600MB disc space for data and unlimited traffic.

During the first weeks when the webhosting had being tested everything looked good, but during February, 2011 the webhosting was destroyed – according to its owner servers were hacked, backups were lost and the websites were unavailable for more than half a month. Because of this incident another webhosting selection had to be done.

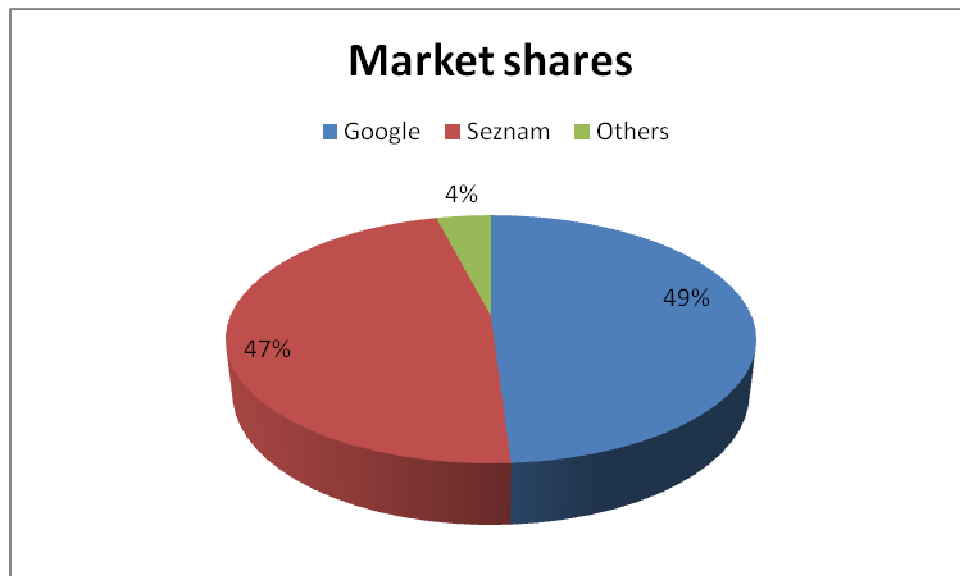
The next webhosting which was chosen was from a company – Onebit.cz, this company has been on the market for more than four years and the availability of their servers is around 99.999% - their servers have being constantly measured since 7<sup>th</sup> February, 2007. This company has two separated datacenters so in case of failure of one of them then other one is still available. This webhosting offers 350MB disc space but the traffic is limited. The traffic limit is set to ten gigabytes per month it should be sufficient for the first months. When the limit is exceeded the company sends an email to inform the owner about this fact, but the website will be still running. The last information which was obtained was that their servers are not blocked by any of the search engines.

Because of the references, history and the used technology as the final webhosting was chosen Onebit.cz.

## 5.2 Keywords research

The main topic of the website, are online flash games which are played inside the web browser. The website will be visited mainly by younger people, let's say in the age from eight to thirty-five, they play these games more than any other group. The website is focused on people whose native language is Czech. Because of that reason only Czech words or the words which are used by people in that group are chosen.

Tools which were used to find the proper words are Google Adwords Keyword Tool, Google Adwords Traffic Estimator Tool and Seznam.cz Keyword Tool. Because the visitors are mainly from the Czech Republic mostly search engines which are used by Czech people were used for keyword research. The Czech search engines market is divided like this:



Picture 12 Czech search engines market shares [11]

The Google is being the most used search engine in the Czech Republic, because many search engines use it as their native search algorithm – for instance Centrum Holding, Atlas or Jyxo. Seznam has its own algorithm. Due to this fact there are only two search engines. So the keywords and concurrency rate was mainly obtained from these two engines.

Keywords	Searches per day	
	Seznam	Google
online hry	14 004	6 700
hry zdarma	3 931	3 667
animace	133	903
onlinovky	4 901	603
super mario	180	493
3D hry	466	220
karetní hry	129	220
strategické hry	130	180
simulátory	170	180
hry pro holky	177	97
střílečky	123	97
sportovní hry	173	80
skákačky	256	63
akční hry	192	63
brutální hry	192	16
oddechové hry	66	16

Table 3 Number of searches per day

<input type="checkbox"/>	Klíčové slovo	Konkurence	Celosvětový objem vyhledávání za měsíc ?	Místní objem vyhledávání za měsíc ?
<input type="checkbox"/>	☆ oddechové hry	<div style="width: 10%;"></div>	590	480
<input type="checkbox"/>	☆ brutální hry	<div style="width: 10%;"></div>	590	480
<input type="checkbox"/>	☆ akční hry	<div style="width: 15%;"></div>	1 900	1 900
<input type="checkbox"/>	☆ skákačky	<div style="width: 10%;"></div>	2 900	1 900
<input type="checkbox"/>	☆ sportovní hry	<div style="width: 10%;"></div>	2 400	2 400
<input type="checkbox"/>	☆ střílečky	<div style="width: 10%;"></div>	3 600	2 900
<input type="checkbox"/>	☆ hry pro holky	<div style="width: 10%;"></div>	4 400	2 900
<input type="checkbox"/>	☆ simulátory	<div style="width: 10%;"></div>	6 600	5 400
<input type="checkbox"/>	☆ strategické hry	<div style="width: 10%;"></div>	8 100	5 400
<input type="checkbox"/>	☆ karetní hry	<div style="width: 10%;"></div>	6 600	6 600
<input type="checkbox"/>	☆ 3d hry	<div style="width: 10%;"></div>	14 800	6 600
<input type="checkbox"/>	☆ super mario	<div style="width: 10%;"></div>	13 600 000	14 800
<input type="checkbox"/>	☆ onlinovky	<div style="width: 10%;"></div>	22 200	18 100
<input type="checkbox"/>	☆ animace	<div style="width: 10%;"></div>	33 100	27 100
<input type="checkbox"/>	☆ hry zdarma	<div style="width: 10%;"></div>	201 000	110 000

Picture 13 Google AdWords Keyword Tool – recommended keywords



Picture 14 Seznam's keyword tool

According to conducted research the chosen keywords are: **online hry, hry zdarma, animace, onlinovky, super mario, 3D hry, karetní hry, strategické hry, simulátory, hry pro holky, střílečky, sportovní hry, skákačky, akční hry, brutální hry, oddechové hry.**

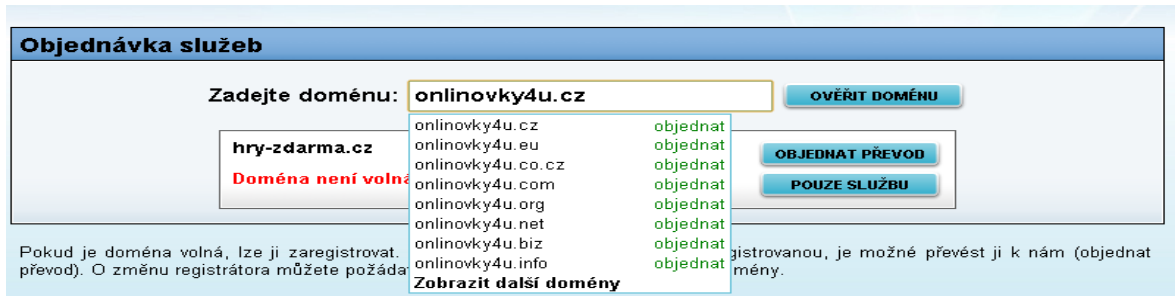
### 5.3 Domain name

The domain name had to meet a few requirements – the suffix had to be .cz because it was meant to be for Czech people, it had to contain at least one keyword and had to be easy to remember. Another requirement was that it had to be available for registration and the length of the name had to be one or two words maximal. As a domain registrar was chosen Web4U.cz because of their user interface, it uses AJAX, which helped to find the domain name quicker.

All the previously mentioned keywords were checked but none of them was available, so it was necessary to add something what would be easy to remember and understandable by young people.

Picture 15 Web4u.cz domain name tool – domain is not available

The final domain name is onlinovky4u.cz – one keyword plus acronym 4U and the .cz suffix. An inspiration was found in the registrar’s domain name – 4U – acronym for “for you” – in Czech “pro Vás”.



Picture 16 Web4u.cz domain name tool – domain is available

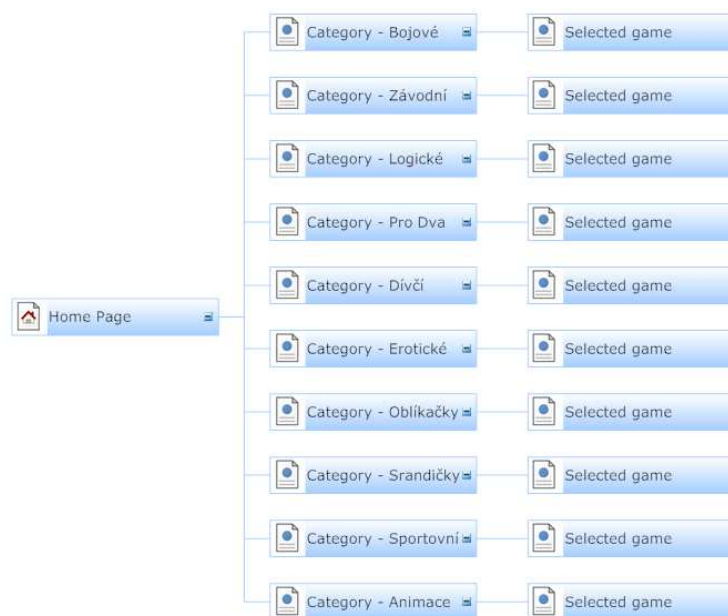
## 5.4 URLs

The structure of the website is as flat as possible. The maximal number of clicks which are needed to get to the furthest page is two. The whole application is written in PHP Zend Framework, which uses mod\_rewrite, so that the URLs are user and search engines friendly, it means that there are no numbers at all in the URL.

Here is a content of the .htaccess file, which provides the mod\_rewrite functionality, it rewrites the URLs:

```
RewriteEngine on
RewriteBase /
RewriteRule !\.(js|ico|txt|gif|jpg|png|css|swf)$ index.php
```

Here is a site structure:



Picture 17 Website structure

There are three ways how to get to the game. The first one is from the home page – there are links which are represented by thumbnails of the most playing games, the newest games, the best games, the most discussed games, recommended games, random games and a sample from each category. The second way is – at first it is necessary to go to the specific category and from there to the specific game. The last one is from the game page this page contains clickable thumbnails of other games, so that it is possible to change the game immediately.

The URLs are self-explaining, they are full of keywords. Each category name is present in the URL and each game name is there as well. The address contains only letters, numbers and hyphens that is all.

Homepage URL:

**www.onlinovky4u.cz**

There are nine categories. The example of a category URL:

**www.onlinovky4u./onlinovky/kategorie/hry/bojove**

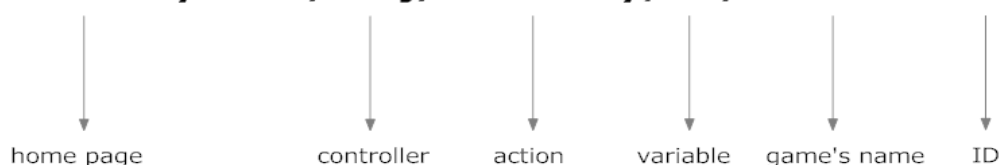


Picture 18 Example of a category URL

Category names are – bojove, zavodni, logicke, prodva, divci, eroticke, oblikacky, sran-dicky, sportovni and animace.

The example of a game URL:

**www.onlinovky4u.cz/hrej/onlinovky/4u/Love-Run-99**



Picture 19 Example of a game URL

Once the user visits the webpage with the game, the action “onlinovky”, which is in the controller “hrej”, checks the variable “4u” and gets the name of the game + ID of that game

– e.g. Love-Run-99 – the name of the game is “Love Run” and the ID is “99”. When the name is made by more than one word, the words are separated by hyphens. The ID and the game’s name are separated by hyphen as well.

Here is a PHP algorithm, which separates the ID and the name of the game, so that it is possible to find the game in the database:

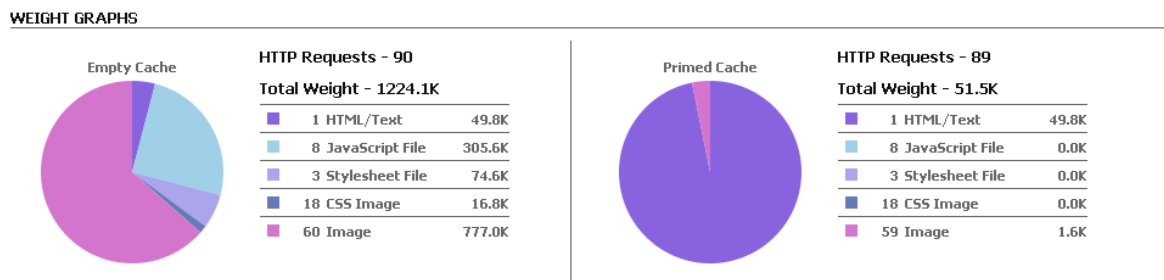
```
$selected=$this->_request->getParam("4u");
$selected=explode("-", $selected);
$selected=$selected[sizeof($selected)-1];
```

### 5.5 Faster loading

The progress in loading speed was measured by YSlow tool – this tool comes from Yahoo! and is intended to be a tool which helps developers to decrease the time needed for loading the page and the size of the data.



Picture 20 YSlow’s performance score before the changes were made



Picture 21 YSlow’s statistics before the changes were made

Several changes were made to decrease the size of the data which were transferred from the server to the user’s browser.

The biggest impact had the change of the expire headers. The first version of the website did not use any expire headers, so that all the content was checked whether it is updated or not, every single visit. Number of requests before the change was 89, but after the change it decreased to 9. Expire headers were set in .htaccess file by this directive:

*ExpiresActive On*  
*ExpiresDefault "access plus 2 years"*

Another significant change was in the data size, when all the CSS files were combined into one file and then compressed by this tool - <http://www.minifycss.com/css-compressor/>. The compression rate was 31.3%. Quite similar tool was used to compress the javascript files - <http://www.minifyjavascript.com/>. The compression rate was 19%.

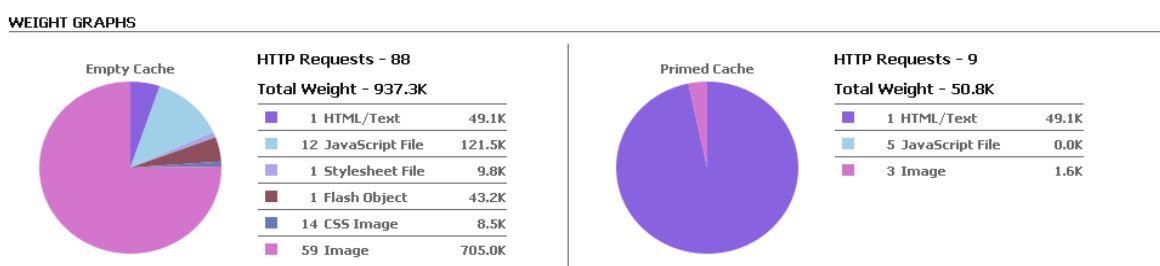
The next step which was made was compression during delivering. The size of data which were transferred from to server to the user decreased from 1224.1KB to 937.3KB. The compression rate was 23%.

Another change which was made was that all the script files were moved to the bottom of the page and all the CSS files were moved to the top of the page.

Probably the most important change for the users was that before any of these changes were made, the time which had been needed to download all the data was 7.02s but after the change, it was decreased to 3.041s. The difference between the old time and the new one is almost 4 seconds.

**Grade B** Overall performance score 86 Ruleset applied: YSlow(V2) URL: <http://www.onlinovky4u.cz/>

Picture 22 YSlow’s performance score after the changes were made



Picture 23 YSlow’s statistics after the changes were made

The overall improvement is 15 points in the YSlow tool.

YSlow parametres	Before	After
Make fewer HTTP requests	F	F
Use a Content Delivery Network (CDN)	F	A
Avoid empty src or href	A	A
Add Expires headers	F	F
Compress components with gzip	F	B
Put CSS at a top	A	A
Put JavaScript at bottom	C	A
Avoid CSS expressions	A	A
Make JavaScript and CSS external	n/a	n/a
Reduce DNS lookups	A	A
Monify JavaScript and CSS	C	B
Avoid URL redirects	A	A
Remove duplicate JavaScript and CSS	A	A
Configure entity tags (Etags)	F	A
Make AJAX cacheable	A	A
Use GET for AJAX requests	A	A
Reduce the number of DOM elements	C	C
Avoid HTTP 404 (Not Found) error	A	A
Reduce cookie size	A	A
Use cookie-free domains	A	A
Avoid AlphasImageLoader filter	A	A
Do not scale images in HTML	A	A
Make favicon small and cacheable	A	A

Table 4 YSlow ranked parameters before and after the changes

## 5.6 On-site optimization

As a tool which could measure the on-site SEO improvement was chosen [www.seo-servis.cz/power-sila-stranky/](http://www.seo-servis.cz/power-sila-stranky/). The first version of the website got 43% as an overall mark. It was necessary to make many changes, because the website had been missing many important elements, such as title tag, description tag, robots.txt etc. Website was not valid either, there was one error.

At first, elements which are not dependent on keywords were added – such as robots.txt file. Validation errors were corrected so that the website was valid and does not have any broken links, which would call HTTP error 404 – page not found.

This document was successfully checked as HTML 4.01 Transitional!	
<b>Result:</b>	Passed
<b>Address:</b>	<input type="text" value="http://onlinovky4u.cz/"/>
<b>Encoding:</b>	utf-8 <input type="text" value="(detect automatically)"/>
<b>Doctype:</b>	HTML 4.01 Transitional <input type="text" value="(detect automatically)"/>
<b>Root Element:</b>	HTML

Picture 24 w3.org syntax validator result – validator.w3.org

The next in the row were texts, which were added – mainly title text. Each web page has a unique title, so that there are not any identical titles.

The chosen keywords are:

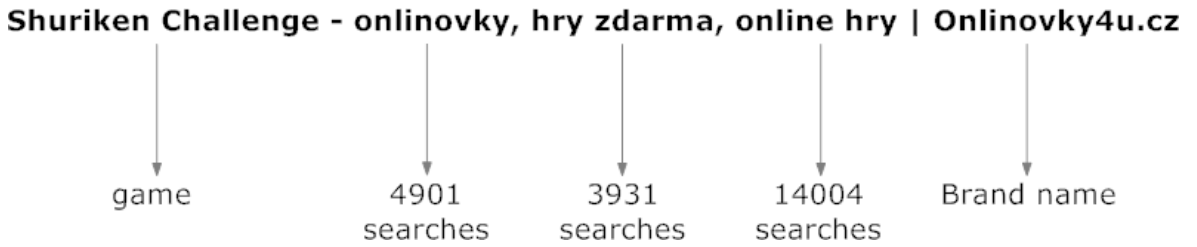
Keywords
<a href="#">hry zdarma</a>
<a href="#">animace</a>
<a href="#">onlinovky</a>
<a href="#">super mario</a>
<a href="#">3D hry</a>
<a href="#">karetní hry</a>
<a href="#">strategické hry</a>
<a href="#">simulátory</a>
<a href="#">hry pro holky</a>
<a href="#">střílečky</a>
<a href="#">sportovní hry</a>
<a href="#">skákačky</a>
<a href="#">akční hry</a>
<a href="#">brutální hry</a>
<a href="#">oddechové hry</a>

Table 5 Chosen keywords

These words were inserted in to the most of the page's elements. As the most important keyword, was chosen the word “onlinovky”, because of the domain name and its memorability. Due to this fact, the title tag's content begins with the word “onlinovky”, its length is 61 characters. It also contains other keywords like “online hry”, “skákačky” and so on. Each category has its own title tag, which contains keywords specific for that category, so for example the title text of the category “bojové” is:

*“Bojové hry, onlinovky, brutální hry, akční hry | Onlinovky4u.cz”*

The title text of a page where the game is placed, begins with the name of the game and then continues with keywords like “onlinovky”, “hry zdarma”, “online hry” and “Onlinovky4u.cz” – the keyword “Onlinovky4u.cz” is there because of the branding.



Picture 25 Game’s page title

Page	Title	Length
<a href="#">Home</a>	Onlinovky, online hry, skákačky, hry zdarma   Onlinovky4u.cz	60
<a href="#">Category - Bojové</a>	Bojové hry, onlinovky, brutální hry, akční hry   Onlinovky4u.cz	63
<a href="#">Category - Závodní</a>	Závodní hry, simulátory, sportovní hry zdarma   Onlinovky4u.cz	62
<a href="#">Category - Logické</a>	Logické hry, hry na přemýšlení, onlinovky   Onlinovky4u.cz	58
<a href="#">Category - Pro dva</a>	Hry pro dva, onlinovky zdarma, karetní hry   Onlinovky4u.cz	59
<a href="#">Category - Dívčí</a>	Dívčí hry, hry pro holky, dívčí onlinovky   Onlinovky4u.cz	58
<a href="#">Category - Erotické</a>	Erotické hry, hry pro dospělé, erotické hry zdarma   Onlinovky4u.cz	67
<a href="#">Category - Oblíkačky</a>	Oblíkačky, dívčí oblíkačky, módní onlinovky   Onlinovky4u.cz	60
<a href="#">Category - Srandičky</a>	Srandičky, oddechové hry zdarma, zábavné hry   Onlinovky4u.cz	61
<a href="#">Category - Sportovní</a>	Sportovní hry, závodní hry, sportovní onlinovky   Onlinovky4u.cz	64
<a href="#">Category - Animace</a>	Animace, zábavná videa, animace zdarma, klipy   Onlinovky4u.cz	62

Table 6 Titles

The next tags which were improved were Meta tags – meta description, meta author, meta robots and meta keywords.

Meta	Content
<a href="#">Description</a>	Onlinovky4u.cz - online hry zdarma! Můžete si zahrát onlinovky, které jinde nenajdete, např. hry pro holky, skákačky, bojovky, oblíkačky, závodní hry
<a href="#">Author</a>	Roman Peček
<a href="#">Robots</a>	index,follow
<a href="#">Keywords</a>	onlinovky, online hry, skákačky, hry zdarma, online hry zdarma, flash hry

Table 7 Meta tags

The final change which was made on the website was that all the important texts were inserted into headlines. The most interesting keywords were inserted into H1 tags, other important elements are names of categories these names are in H2 tags, the less important categories names are in H3 tags.

H1									
Onlinovky, online hry, skákačky, hry zdarma   Onlinovky4u.cz									

H2									
Bojové	Závodní	Logické	Pro dva	Dívčí	Erotické	Oblíkačky	Srandičky	Sportovní	Animace

H3			
Nejnovější onlinovky	Nejhranější onlinovky	Nejdiskutovanější onlinovky	Nejlépe hodnocené hry
Náhodné onlinovky	Dívčí hry	Hry pro dva	Logické
Závodní	Bojové	Erotické	Oblíkačky
Srandičky	Sportovní	Animace	Doporučené hry

Picture 26 Headlines texts

After, all of these changes had been made, the overall score increased from 43% to 96%. It was a significant improvement, because before the changes were made, there were categories, but they were not in the headlines structure, so when the styles were disabled in the browser, it was not possible to distinguish what are the most important elements of the website.

## CONCLUSION

This thesis deals with website optimization. There were described techniques which could help improve overall performance of the website it means decrease the time needed for loading the page. These was not all the described techniques, the other techniques dealt with importance of website optimization for search engines, they show how it is possible to gain visitors from these engines, what it is necessary to do to be listed in search engines results.

In the theoretical part there was shown how some changes could ruin all the work, if they are made the wrong way. It is necessary to be very careful when these changes are making. Otherwise the resources which have been spent on previous optimization would be wasting of resources such as time, money and people as well. There are only few “disasters” bigger than angry visitor. These visitor is a lost visitor and it is very to persuade him to get back to the website.

In the practical part there were shown techniques in practice. It was shown how to find out whether the domain is available for registration or not and other necessary steps before the website is launched.

There was also shown how to decrease the time needed for loading of the website on a real world example. The applied approaches decreased the size of the website from 1224.1KB to 937.3KB – when the cache was empty. When the cache was primed the before the changes the size of the data was 51.6KB and after the changes 50.8KB, the difference is not so big, but there is another significant improvement when the cache is primed – the number of request for data was decreased from 89 to 9. The biggest change which could be seen by users is that the time needed for downloading the page was decreased from 7.02 seconds to 3.041 seconds, it means that the sites is fully loaded after 3 seconds, not 7.

The second part of the practical part deals with on-site optimization. The tool which was used for checking the improvement had been found at <http://seo-servis.cz/power-sila-stranky/> . Before the changes were made the overall score was 43% but after the changes it was increased to 96%. It was a big improvement.

**BIBLIOGRAPHY**

- [1] SOUDERS, S., *High performance web sites essential knowledge for front-end engineers*, 1<sup>st</sup> ed., O'Reilly Media, September 11<sup>th</sup> 2007, ISBN: 978-0-596-52930-7
- [2] SOUDERS, S., *Even Faster Web Sites: Performance Best Practices for Web Developers*, 1<sup>st</sup> ed., O'Reilly Media, June 10<sup>th</sup> 2009, ISBN: 978-0-596-52230-8
- [3] ENGE, E., SPENCER, S., FISHKIN, R., STRICCHIOLA, J., *The art of seo - mastering seo*, 1<sup>st</sup> ed., O'Reilly Media, October 21<sup>st</sup> 2009, ISBN: 978-0-596-51886-8
- [4] *Search engine optimization starter guide* [online], 2010 [cit. 2010-10-20], <[http://static.googleusercontent.com/external\\_content/untrusted\\_dlcp/www.google.com/cs/webmasters/docs/search-engine-optimization-starter-guide.pdf](http://static.googleusercontent.com/external_content/untrusted_dlcp/www.google.com/cs/webmasters/docs/search-engine-optimization-starter-guide.pdf)>
- [5] Tenni Theurer, *Performance Research, Part 2: Browser Cache Usage – Exposed* [online], 2010 [cit. 2010-10-20], <<http://yuiblog.com/blog/2007/01/04/performance-research-part-2>>
- [6] Seomoz.com, *The beginners guide to SEO, 2<sup>nd</sup> edition*, [online], 2010 [cit. 2011-04-21], <<http://www.seomoz.org/dp/download-the-pdf-version-of-the-beginners-guide-to-seo>>
- [7] Seomoz.com – posted by “randfish”, *Illustration the Long Tail*, [online], November 24<sup>th</sup> 2009 [cit. 2011-04-05], <<http://www.seomoz.org/blog/illustrating-the-long-tail>>
- [8] JONES, K., *Search Engine Optimization: Your visual blueprint for effective Internet marketing*, Wiley Publishing, Inc., 2008, ISBN: 978-0-470-22448-9
- [9] Google.com, *Search engine optimization starter guide*, [online], 2010 [cit. 2011-04-28], <[http://static.googleusercontent.com/external\\_content/untrusted\\_dlcp/www.google.com/en/webmasters/docs/search-engine-optimization-starter-guide.pdf](http://static.googleusercontent.com/external_content/untrusted_dlcp/www.google.com/en/webmasters/docs/search-engine-optimization-starter-guide.pdf)>
- [10] Seznam.cz, *Čeho se při optimalizaci vyvarovat*, [online], 2011 [cit. 2011-05-07], <<http://napoveda.seznam.cz/cz/fulltext-hledani-v-internetu/ceho-se-pri-optimalizaci-vyvarovat/>>
- [11] lidovky.cz, *Zemětřesení v českých vyhledávačích. Google předstihl Seznam.cz*, [online], January 11<sup>th</sup> 2011 [cit. 2011-05-10], <[http://www.lidovky.cz/zemetreseni-v-ceskych-vyhledavacich-google-predstihl-seznam-cz-pwm-/ln-media.asp?c=A110111\\_095334\\_ln-media\\_pks](http://www.lidovky.cz/zemetreseni-v-ceskych-vyhledavacich-google-predstihl-seznam-cz-pwm-/ln-media.asp?c=A110111_095334_ln-media_pks)>

**LIST OF ABBREVIATIONS**

FTP	File Transfer Protocol
URL	Uniform Resource Locator
WWW	World Wide Web
NLP	Natural Language Processing
PPC	Pay Per Click
SEO	Search Engine Optimization
HTML	HyperText Markup Language
XML	Extensible Markup Language
IP	Internet Protocol
AJAX	Asynchronous JavaScript and XML
CDN	Content Delivery Network
CSS	Cascading Style Sheet
HTTP	HyperText Transfer Protocol
Gzip	GNU Zip
PHP	Hypertext Preprocessor
MySQL	My Structured Query Language

**LIST OF FIGURES**

Picture 1 The first Seznam's webdesign from the year 1996.....	15
Picture 2 Search engines market share.....	16
Picture 3 Facebook's home page – easy to navigate.....	19
Picture 4 Google's home page – no disturbing elements.....	19
Picture 5 Long tail terms [7] .....	25
Picture 6 Google's AdWords Keyword Tool.....	28
Picture 7 CSS sprite [1] .....	36
Picture 8 Components loading.....	40
Picture 9 Google's images search .....	45
Picture 10 Links normal behavior.....	53
Picture 11 Fake links behavior.....	53
Picture 12 Czech search engines market shares [11] .....	56
Picture 13 Google AdWords Keyword Tool – recommended keywords.....	57
Picture 14 Seznam's keyword tool .....	57
Picture 15 Web4u.cz domain name tool – domain is not available.....	58
Picture 16 Web4u.cz domain name tool – domain is available .....	59
Picture 17 Website structure .....	59
Picture 18 Example of a category URL .....	60
Picture 19 Example of a game URL .....	60
Picture 20 YSlow's performance score before the changes were made .....	61
Picture 21 YSlow's statistics before the changes were made .....	61
Picture 22 YSlow's performance score after the changes were made .....	62
Picture 23 YSlow's statistics after the changes were made .....	62
Picture 24 w3.org syntax validator result – validator.w3.org .....	64
Picture 25 Game's page title .....	65
Picture 26 Headlines texts.....	66

**LIST OF TABLES**

Table 1 The biggest players' CDNs [1] .....	35
Table 2 Page weight savings for different levels of compression [1].....	37
Table 3 Number of searches per day.....	57
Table 4 YSlow ranked parameters before and after the changes .....	63
Table 5 Chosen keywords .....	64
Table 6 Titles .....	65
Table 7 Meta tags.....	65

## APPENDICES

## **APPENDIX P I: APPENDIX TITLE**