

Vysoká škola ekonomická v Praze

Fakulta informatiky a statistiky

Katedra informačních technologií

Student : **Martin Vlčko**
Vedoucí bakalářské práce : **Ing. Libor Gála**
Recenzent bakalářské práce: **Ing. Jan Tichý**

TÉMA BAKALÁŘSKÉ PRÁCE

**NEKALÉ PRAKTIKY V SEARCH ENGINE
OPTIMIZATION**

ROK : 2008

Poděkování

Na tomto místě bych rád poděkoval vedoucímu mé bakalářské práce Ing. Liboru Gálovi za cenné rady a připomínky při zpracování zvoleného tématu a PhDr. Radoslavovi Rohaľovi, PhD. za praktické připomínky.

Prohlášení

Prohlašuji, že jsem bakalářskou práci zpracoval samostatně a že jsem uvedl všechny použité prameny a literaturu, ze kterých jsem čerpal.

V Praze dne 15.12.2008

.....
podpis

Abstrakt

Optimalizace pro vyhledávače patří v současné době mezi základní stavební prvky webových stránek. Bez ní se stává web nekonkurenceschopným a nemá šanci uspět v tržním internetovém prostředí. Základní prvky optimalizace webových stránek podporují marketingovou propagaci.

Cílem této práce je popsat nekalé praktiky optimalizace pro vyhledávače a nastínit možný boj proti nim. První kapitola se zabývá principy optimalizace a také důvody, proč je nutné webové stránky optimalizovat. Další kapitola popisuje nejdůležitější faktory optimalizace. Následující část práce se detailně zabývá popsáním nekalých praktik a možné obrany proti nim z několika různých pohledů.

Přínos této práce spočívá především v sepsání všech nejznámějších nekalých SEO metod, včetně jejich bližšího popisu a návrhu obrany proti jejich zneužívání. Praktické ukázky obsažené v práci mohou dále sloužit i jako základní návod na co si dát při optimalizaci pozor a čeho se zcela vyvarovat.

Abstract

A search engine optimization is the basic milestone of web sites creation nowadays.

Without a process of search engine optimization the web sites are usually not competitive and can not find their place in the internet market environment.

Additionally, the elements of web sites optimization support and develop marketing and promotion.

Main goal of our bachelor thesis is a black hat description for explorers and also the proposal of protection against black hats.

First part of our thesis develops the principles of optimization and also the reasons why it is necessary to optimize the web sites.

Following chapter describes the most important factors of optimization. A detailed description of black hats from different points of view can be found in the third chapter.

The added value of this bachelor thesis is mainly created by SEO black hat methods including detailed descriptions and proposals of protection against them. Moreover, the practical examples can be used as basic instructions how to avoid black hats when optimizing the web sites.

Obsah

1	Úvod.....	1
1.1	Téma práce.....	1
1.2	Cíl práce.....	1
1.3	Přístupy k řešení.....	1
1.4	Používaná notace programových kódů.....	2
2	Účel SEO a přístupy.....	3
2.1	Pozice SEO včetně vazeb na životní cyklus webu.....	3
2.2	Přístupy.....	6
2.2.1	Zařazení.....	6
2.2.2	Fáze webu.....	7
2.2.3	Chování účastníků v optimalizačních procesech.....	7
2.2.4	Metody optimalizace.....	8
2.2.5	Povaha praktik SEO.....	9
2.3	Shrnutí.....	10
3	Analýza nekalých praktik.....	11
3.1	Přístup k analýze.....	11
3.2	Analýza nekalých praktik.....	11
3.2.1	Opakující se klíčová slova v textu (repeated words, keyword stuffing).....	11
3.2.2	Malý text.....	12
3.2.3	Skrytý text (hidden text).....	14
3.2.4	Doorway pages (entry pages).....	16
3.2.5	Cloaking.....	18
3.2.6	Automaticky generovaný obsah (Automatic Content Generation).....	19
3.2.7	Doménový spam.....	20
3.2.8	Spamování diskuzí.....	22

3.2.9	Klamné přesměrování (deceptive redirects)	26
3.2.10	Odkazové farmy (link farms)	27
3.2.11	Patičkové odkazy	29
3.2.12	Duplikované stránky (mirror sites)	31
3.3	Shrnutí výsledků analýzy	34
4	Závěr	37
	Seznam použitých zkratk a termínů	41

1 Úvod

1.1 Téma práce

Optimalizace pro vyhledávače představuje obor zabývající se úpravou webových stránek za účelem jejich lepšího ohodnocení ze strany vyhledávačů a tím získání vyšších pozic ve výsledcích vyhledávání na relevantní klíčová slova.

Nekalé praktiky optimalizace jsou prvky, které jsou vyhledávači buďto zakázané, nebo veřejností eticky neuznávané. Jejich využívání vede k nerelevantním a nežádoucím výsledkům vyhledávání, kterým se vyhledávače brání penalizováním internetových stránek.

Již několik let se zabývám optimalizací webových stránek, čímž jsem získal potřebný přehled a zkušenosti, které mi pomohou objektivně zpracovat téma nekalých praktik v optimalizaci pro vyhledávače.

1.2 Cíl práce

V této práci si kladu za cíl přiblížit nekalé praktiky optimalizace pro vyhledávače, jejich rozpoznání, dopad na výsledky vyhledávání a možnou obranu proti nim i běžným uživatelům internetu, kteří nejsou odborníky na danou problematiku. Správcům webových stránek chci předložit přehled všech technik, kterých by se měli vyvarovat a nastínit jejich závažnost, možné následky a způsob obrany proti spamu. Dalším cílem je navrhnout vyhledávačům možnosti identifikace jednotlivých nekalých praktik.

Nejdříve má práce za úkol uvést do problematiky optimalizace pro vyhledávače a přiblížit, proč je nutné se optimalizací webových stránek zabývat. Poté se dočtete o zvolených metodách a technikách, které byly zvoleny při zpracovávání této práce. Dále vám poskytnu základní přehled a orientaci v problematice a ve stěžejní části seznámí se zkoumanými hledisky jednotlivých nekalých optimalizačních praktik.

Na závěr poskytnu přehledné shrnutí a celkové zhodnocení nekalých praktik.

1.3 Přístupy k řešení

Při zpracovávání bakalářské práce jsem postupoval formulováním klíčových faktorů, které mají vliv na úspěšnou či neúspěšnou optimalizaci, jejich rozdělení do kategorií a sumarizaci dle vytyčených kritérií.

Metodu, kterou se práce řídí, je popis. Tato metoda dovoluje přiblížit zpracovávané téma, poskytnout základní informace a zorientovat se v problematice.

Práce je určena nejen pro běžné uživatele internetu, kterým dává bližší teoretické a praktické znalosti optimalizačních technik, ale také pro správce webových stránek, kteří zde naleznou výčet jednotlivých nekalých SEO faktorů a důvody, proč se jim vyhýbat ve svých webových prezentacích.

Problematika je zpracována s důrazem na popis nekalých praktik. Uvést výčet a popis jednotlivých technik, které využívají doporučených optimalizačních technik, by bylo nad rámec této práce, proto uvádím odkaz na elektronickou podobu diplomové práce Ing. Radima Smíčky, ve které se lze blíže dočíst o jednotlivých SEO faktorech.¹

¹ Viz zdroj 0

1.4 Používaná notace programových kódů

V názorných ukázkách zdrojového kódu budu využívat následujícího označení. Samotný příklad zdrojového kódu je uveden odlišným stylem písma. ***Tučně a kurzívou je pak zvýrazněný text***, který označuje důležité prvky pro jejich lepší čitelnost, jak je vidět na následujícím příkladu.

```
<meta HTTP-EQUIV="Refresh" CONTENT="0"; URL=http://ABC
```

V příkladu jsem označil 3 výrazy, které jsou nejpodstatnější.

- Refresh – jedná se o přesměrování.
- 0 – přesměrování se provede za 0 vteřin (ihned) po načtení kódu.
- ***http://ABC*** – tímto termínem označuji odkaz na smyšlenou doménu ABC.

2 Účel SEO a přístupy

Tato kapitola má za úkol přiblížit základní principy optimalizace pro vyhledávače. Vysvětlit jednotlivé součásti, z kterých se skládá, a popsat její význam tak, aby byly zřejmé dopady nekalých praktik, popisovaných v navazující kapitole.

2.1 Pozice SEO včetně vazeb na životní cyklus webu

Optimalizace pro vyhledávače, neboli SEO (Search Engine Optimization), je obecně sada postupů, technik a doporučení, jak dosáhnout vyšších pozic webových stránek ve výsledcích vyhledávání (SERP – Search Engine Result Page) na požadovaná klíčová slova (KW - Keywords).

Bez optimalizovaných webových stránek se nemůže obejít žádný moderní web, který se snaží stát nejnavštěvovanějším ve svém oboru. V komerční oblasti je pak optimalizování pro vyhledávače nástrojem, který podle serveru Marketingové noviny [9] dokáže:

- zvýšit návštěvnost
- zlepšit konverzní poměr
- snížit náklady na získání zákazníka
- rozšířit tržní podíl v daném oboru
- zvýšit příjmy a zisky z internetových stránek
- budovat a rozšiřovat povědomí o značce
- provádět public relations online.

Všechny tyto výše uvedené faktory v důsledku znamenají vyšší zisky a to je také důvod, proč je SEO optimalizace nezbytná pro úspěšný konkurenceschopný web.

Vyhledávače:

Mluvíme-li o vyhledávačích v českém prostředí, máme na mysli následující (viz. Tabulka 1), kteří zastupují, dle serveru Navrcholu.cz [2], největší podíl na trhu.

Vyhledávač	Červenec 2006	Říjen 2006	Leden 2007	Říjen 2007	Červen 2008
Seznam	61,10%	63,39%	62,53%	62,30%	62,80%
Google	23,48%	23,85%	24,75%	28,87%	30,16%
Centrum + Atlas	-	-	-	-	4,14%
Centrum	5,37%	4,72%	4,84%	3,39%	2,86%
Atlas	2,86%	2,38%	2,58%	1,53%	1,28%
Jyxo	0,79%	0,57%	0,42%	0,24%	0,16%

Tabulka 1 – Vývoj podílu jednotlivých vyhledávačů (červenec 2006 – červen 2008) [2]

Ve výsledcích vyhledávání jsou weby na vyšších pozicích daleko navštěvovanější, než weby na pozicích nižších. Důkazem je tzv. oční mapa, která znázorňuje pohled uživatele na výsledky vyhledávání ve vyhledávači Google. Červená místa na mapě značí oblast, která je nejsledovanější ve výsledcích vyhledávání. Obsadit přední příčky znamená oslovit největší množství potenciálních zákazníků.



Obrázek 1 – Oční mapa z výsledků vyhledávání Google [3]

Celý proces řazení výsledků funguje na činnostech vyhledávacího robota, který prochází internetové stránky a každou z nich ohodnotí na základě kvality jednotlivých faktorů. Hodnocení těchto faktorů následně spočítá a přiřadí stránce tzv. rank.

Každý vyhledávač má svůj vlastní způsob hodnocení webových stránek, který je neveřejný a přísně střežený. Prozrazením vzorečku, na kterém funguje řazení stránek ve výsledcích vyhledávání, by znamenalo přemíru optimalizace a ztrátu kvalitního čtivého obsahu. Webové stránky by se tvořily za účelem nejlepšího odhodnocení vyhledávači, a cíl poskytnout uživatelům relevantní informace, by ustoupil do pozadí.

PageRank

Zakladatelé vyhledávače Google nicméně zveřejnili vzoreček [4], kterým počítali PageRank v době jeho úplného začátku. PageRank je hodnota, která vypovídá o kvalitě a počtu zpětných odkazů vedoucích na danou stránku.

PageRank stránky A lze spočítat pomocí vzorečku:

$$PR(A) = (1-d) + d * (PR(T_1)/C(T_1) + ... + PR(T_n)/C(T_n)) \quad [4]$$

$PR(A)$ = PageRank stránky A

d = tlumicí faktor v rozmezí 0 až 1; výchozí nastavení je nejspíš 0,85

$PR(T_n) = \text{PageRank stránky } n$

$C(T_n) = \text{počet odchozích odkazů stránky } n$

Výsledný PageRank stránky je ovšem neveřejné číslo, které se dále převádí do veřejné stupnice GTPR (Google Toolbar Page Rank), jak ukazuje Tabulka 2.

GTPR	Skutečný PageRank	
	od	do
0/10	0,15	0,9
1/10	0,9	5,4
2/10	5,4	32,4
3/10	32,4	194,4
4/10	194,4	1 166,4
5/10	1 166,4	6 998,4
6/10	6 998,4	41 990,4
7/10	41 990,4	251 942,4
8/10	251 942,4	1 511 654,4
9/10	1 511 654,4	9 069 926,4
10/10	9 069 926,4	$0,85 * N + 0.15$

Tabulka 2 – Přepočítání PR na GTPR [5]

PageRank je pouze jeden z faktorů, který má vliv na řazení výsledků ve vyhledávači Google. Vedle něho existuje ještě algoritmus pro počítání kvality prvků webové stránky, který je také neveřejný a v kombinaci s PageRankem a dalšími faktory mají vliv na celkové řazení výsledků.

Řazení výsledků na Centrum.cz

Dušan Janovský na svém blogu [6] detailně popsal řazení výsledků vyhledávače Centrum.cz. Rozhodl se tak zveřejnit tuto zásadní informaci několik dní před nahrazením vyhledávacího enginu Morfeo enginem „Googlovským“.

Z následující tabulky (viz. Tabulka 3) je patrné, které prvky webové stránky vyhledávače všeobecně rozpoznávají a jakou váhu jim přisuzují.

Číslo v závorkách znamená rozptýl při zadání klíčového slova v jiném tvaru, než je na stránce v jednotlivých prvcích uvedeno.

Body	Slovo se nachází v
110	katalogovém titulku
30	katalogovém popisku
100	titulku stránky (tag <title>)
80	nadpisu h1, h2, h3
50 (-10)	nadpisu h4, h5, h6
20 (-70)	normálním textu
50 (-40)	tagu em, strong, i a b (zvýraznění)
10	tagu small
10	tagu alt (u obrázků)
80	textu odkazů mívajících na stránku
0	meta descriptions
90	doméně druhého řádu
25	Subdoméně

Tabulka 3 – hodnocené prvky stránky vyhledávačem Centrum.cz na enginu Morfeo [6]

SRank

Vyhledávač Seznam nazývá své hodnocení SRank, jehož hodnota se nachází v rozmezí od 0 do 100.

Samotná společnost Seznam popisuje ve své nápovědě pro uživatele [7], S-rank jako „váženou nelineární kombinací různých veličin, v nichž výrazně převažují off-page faktory. Výpočet hlavního zdroje ranku se podobá známému algoritmu Hubs & Authorities [8], ale je upraven tak, aby dával smysl i pro netematické množiny stránek.“

2.2 Přístupy

K problematice optimalizace webových stránek můžeme přistupovat z několika hledisek.

Obecně je výhodné optimalizovat webové stránky téměř vždy, obzvlášť tehdy, když se jedná o komerční web, který má přilákat potenciální návštěvníky k jednorázovému, nebo i opakovanému používání webových stránek.

2.2.1 Zařazení

Search Engine Marketing (SEM) je pojem používaný pro marketingové nástroje vyhledávačů. Mezi tyto nástroje můžeme zařadit PPC (Pay Per Click) a SEO (Search Engine Optimization).

PPC je forma placených textových, ale i grafických inzerátů, videí či animací, které vyhledávače zobrazují ve výsledcích vyhledávání (placené pozice), nebo v obsahové síti a snaží se uživatele internetu zaujmout ke kliknutí na reklamu, a tím ho odkázat na propagované zboží či službu.

SEO je jedním z nástrojů, jakým lze na webové stránky přivést více návštěvníků i potenciálních zákazníků a zvýšit jeho konkurenceschopnost na poli internetového trhu, neboť jeho využitím lze dosáhnout předních pozic přirozených výsledků vyhledávání.



Obrázek 2 – Rozmístění výsledků vyhledávání na Google

2.2.2 Fáze webu

Nejvhodnějším způsobem je navrhnout optimalizační techniky ještě před začátkem tvorby nových webových stránek. Na základě přesné specifikace lze vytvořit a spustit stránky, se kterými nebudou mít vyhledávače problémy při indexování. Výhodou je takzvaný „čistý účet“, neboť vyhledávače indexují nové stránky domény. Nevýhodou je naopak právě nová doména, které vyhledávače nepřikládají takovou důležitost, jako doméně zavedené.

Naproti tomu stojí optimalizace již indexovaných stránek zavedené domény, kde je nevýhodou historie nekvalitních stránek a v případě nešetření přesměrování upravených URL stránek také riziko vytvoření duplicitního obsahu.

2.2.3 Chování účastníků v optimalizačních procesech

Účastníky optimalizačních procesů **rozdělují** do tří následujících skupin podle jejich chování v daném procesu.

Poctivý konkurent

Jednu z modelových rolí lze pojmenovat „Poctivým konkurentem“. Jeho cílem je dlouhodobě obsazovat přední pozice ve výsledcích vyhledávání a zvyšovat návštěvnost a konverzní poměr svého webu. Těmito způsoby se snaží získat nové, ale i udržet stálé zákazníky, pro dosahování stále vyšších zisků a posilování obchodní značky společnosti.

Aby obstál v boji s konkurencí, využívá všechny dostupné poctivé SEO praktiky. Sleduje kroky konkurence a snaží se být vždy o krok napřed. Důraz při tom klade na důvěryhodnost a objektivnost svého webu.

Nekalý konkurent

Na opačné straně poctivého konkurenta stojí nekalý konkurent, který se snaží dostat svůj web za každou cenu na přední pozice. Neobává se využít nekalých SEO praktik i za cenu možné penalizace vyhledávačů. Jeho cíl je krátkodobý, proto ho penalizace, v podobě vyřazení webových stránek z indexu, příliš negativně nezasáhne. Nejde mu o dobrou pověst, ani o důvěryhodnost ze

strany vyhledávačů. Rychle obsadí přední pozice ve výsledcích vyhledávání na zvolená klíčová slova a v krátkém čase získá velký počet návštěvníků, kteří přijdou z vyhledávačů.

Většinou se jedná o případy, kdy je potřeba informovat o jednorázové akci a v rekordně krátké době tak oslovit velké množství potenciálních návštěvníků. Webové stránky jsou nejčastěji vytvořeny za účelem finančního zisku.

Robot vyhledávače

Robot vyhledávače stojí na třetím vrcholu pomyslného trojúhelníku. Má za úkol procházet webové stránky a analyzovat jejich obsah. Na základě neveřejného algoritmu ohodnotí jednotlivé analyzované prvky webových stránek a určí jim hodnotu (rank), která je závislá na jejich kvalitě. Tyto vypočítané hodnoty nakonec vyhledávače sestupně seřadí a zobrazí relevantní výsledky vyhledávání.

Pokud robot odhalí praktiky nekalého konkurenta, smaže z indexu všechny stránky, které jeho doména obsahuje. Tato penalizace také sníží, či úplně vynuluje, rank všech stránek v rámci penalizované domény.

Odhalit nekalé praktiky může robot několika způsoby. Samotným odhalováním jednotlivých nekalých praktik se zabývá kapitola 3.2. Některé praktiky robot ale není schopen odhalit, proto existují tzv. spam reporty, kterými vyhledávače dávají uživatelům internetu možnost nahlásit podezřelou stránku, kterou následně administrátoři prověří a v případě odhalení nekalých praktik manuálně penalizují.

Přítomnost robota na stránkách lze odhalit pomocí user-agenta. Robot vyhledávače Seznam se jmenuje Seznambot [10], vyhledávač Google pojmenoval svého robota Googlebot [11].

2.2.4 Metody optimalizace

Metody optimalizace lze rozdělit na základě povahy jednotlivých faktorů do dvou základních skupin.

Onpage faktory

Jedná se o metodu optimalizace, která se týká pouze prvků umístěných na unikátní stránce. Takovým prvkem je například titulek, popis, nadpisy, tučné písmo, odkazy či popisky obrázků. Účinnost těchto elementů je také závislá od vhodně zvolených klíčových slov, souvisejících s tématem celé stránky a na jejich hustotě, v jaké se v textu vyskytují.

Offpage faktory

Další optimalizační metoda se nazývá offpage faktory. Na rozdíl od onpage faktorů pracuje tato metoda s prvky, které se na stránce nenachází. Jsou to především zpětné odkazy směřující na webovou stránku.

Zpětné odkazy můžeme dále rozdělit na externí a interní. Externí zpětné odkazy pochází z jiné domény, než na které se nachází stránka, kam odkazy směřují. Interní zpětné odkazy jsou pouze v rámci stejné domény.

Obecně lze říci, že čím více zpětných odkazů stránka získá, tím lepší bude její hodnocení. Samotný odkaz na tuto stránku znamená jakýsi hlas pro ni. To, že uživatelé internetu na ni rádi odkazují, je pro vyhledávače známka jejího kvalitního obsahu.

Při hodnocení zpětných odkazů dávají vyhledávače větší význam odkazům:

- pocházejícím z webu s vyšším rankem
- z webu s tematicky příbuzným obsahem

- jejichž anchor text² obsahuje výstižná klíčová a tematicky příbuzná slova (LSI – Latent Semantic Indexing)
- umístěným v textu článku, méně potom samostatným odkazům v patičce

2.2.5 Povaha praktik SEO

Praktiky optimalizace pro vyhledávače lze z hlediska jejich povahy rozdělit do několika skupin. Vyhledávače se snaží odhalovat záměr jednotlivých praktik a podle toho k nim přistupují. Pokud shledají, že jsou vhodně využívány, stránku ohodnotí kladně. V opačném případě můžou stránce ohodnocení odebrat zčásti, nebo úplně celé. Na hraně mezi těmito dvěma případy stojí praktiky, které nelze shledávat jako nekalé, ale ani příliš etické.

Povolené SEO praktiky (White hat)

Do této skupiny patří techniky optimalizace, které jsou jednotlivými vyhledávači doporučovány pro získání lepšího ohodnocení. Dva nejvýznamnější vyhledávače pro český trh – Seznam [12] a Google [13] doporučují v podstatě stejné techniky, přesto budou výsledky v těchto vyhledávacích zcela odlišné. Důvodem je odlišný algoritmus a nastavené váhy technik.

Chewy Trehwella pracující na pozici Developer Advocate [14] u společnosti Google tvrdí, že existuje více než 200 faktorů, které Google rozeznává pro ohodnocení kvality webové stránky. Ty nejdůležitější z nich byly oficiálně zveřejněny 13. 11. 2008 v Google's Search Engine Optimization Starter Guide [13].

Nekalé SEO praktiky

Nekalé praktiky jsou všechny ty, které vyhledávače penalizují smazáním ze svých seznamů indexovaných stránek a snížením či odstraněním hodnot ranku.

a) Zakázané SEO praktiky (Black Hat)

Zakázané praktiky, neboli spam techniky, jsou metody k dosažení vyšší pozice ve výsledcích vyhledávání, které vyhledávače považují za nekalé, a proto je penalizují. Použití těchto technik je velmi riskantní a nevede k dlouhodobě vysokým pozicím ve výsledcích vyhledávání. Web, který je používá, bývá po odhalení penalizován, což má za následek snížení hodnoty ranku, nebo dokonce úplné smazání indexovaných stránek, a to často i na několik měsíců.

Vyhledávače jsou nuceni rozpoznávat povahu a záměr webových stránek, aby předešly zobrazování těch, které nejsou vytvořeny za účelem poskytování relevantních informací, ale za účelem zisku.

Cílem každého vyhledávače je poskytnout uživateli odkazy na webové stránky, které se vztahují k zadaným klíčovým slovům. Aby eliminovaly ty odkazy, které jsou pro uživatele nerelevantní, používají k tomu dvě metody.

- První metoda je založená na schopnosti identifikovat zakázané SEO praktiky pomocí algoritmu. Webových stránek je nespočetné množství a každým dnem vznikají nové. Není možné manuálně kontrolovat každou stránku zvlášť. Algoritmus vyhledávačů

² Text odkazu anchor text

dokáže některé odhalit na základě jistých předpokladů, které jsou blíže popsány v kapitole 3.2.

- Druhá metoda přichází na řadu, až pokud selže metoda první. Vyhledávače nabízí, pro zkvalitnění výsledků vyhledávání, možnost nahlásit některé zakázané praktiky, které administrátoři osobně zkontrolují a na základě jejich uvážení rozhodnou o jejich penalizování.

Spam report – Seznam

Vyhledávač Seznam reaguje na spam report [15], podle mých vlastních zkušeností, nejčastěji do jednoho týdne. V případě konkrétního příkladu nahlášení doménového spamu v kombinaci s malým textem a nesmyslným opakováním klíčových slov Seznam odstranil do dvou dnů celou doménu z indexu. SRank před nahlášením reportu měl hodnotu 73/100. Po šesti dnech od nahlášení spamu spadl na hodnotu 4/100 a po třinácti dnech na hodnotu 0/100. Celá doména se během dvou dnů od nahlášení dostala na listinu nežádoucích webových stránek (blacklist).

Spam report - Google

Vyhledávač Google také umožňuje nahlásit spam [16], ovšem zkušenosti se zpětnou vazbou jsou dle mých zkušeností horší, než u českého vyhledávače Seznam. Při nahlásování nekalých praktik výše zmíněného příkladu, ani po 37 dnech vyhledávač Google nereagoval vymazáním webu ze svého indexu, ani nesnížil PageRank. Ani po druhém upozornění jsem nezaznamenal žádnou změnu. V ostatních případech Google reagoval většinou do dvou týdnů.

Důvodem může být velké množství nahlášených stránek, využívajících nekalé prvky, a kapacitní neschopnost nahlášené spamy penalizovat jednotlivě. Lze tedy očekávat, že pracovníci vyhledávače Google jednotlivé spamy analyzují a navrhuji změny algoritmu, který by je dokázal sám rozpoznat a při dalším indexování stránky doménu automaticky penalizoval.

b) Neetické SEO praktiky

Do této skupiny lze zařadit ty faktory, které nejsou vyhledávači penalizovány, ale z hlediska všeobecné etiky jsou nemorální a veřejností odsuzované.

Jako příklad lze uvést chování v diskuzích. Vydávání se za spokojené uživatele vlastního výrobku či služby, či uvádění zpětných odkazů, neboť se jedná se o formu PR propagace.

2.3 Shrnutí

V této kapitole jsem vysvětlil, proč je optimalizace pro vyhledávače důležitým marketingovým nástrojem a jakých cílů s ní lze dosahovat. Zároveň jsem uvedl přehled faktorů, které ji přímo, či nepřímo ovlivňují a přiblížil tuto problematiku z několika různých pohledů, pro vytvoření komplexnější představy o jejích možnostech.

3 Analýza nekalých praktik

Cílem této kapitoly je popsat metody hodnocení jednotlivých nekalých SEO praktik, stanovit si váhy dílčích hodnocení, na konkrétních příkladech uvést způsoby zneužití k dosažení vyšších pozic ve výsledcích vyhledávání a nastínit možnou obranu proti těmto praktikám z pohledu vyhledávače, uživatele stránek, či jejich správce.

3.1 Přístup k analýze

Předmětem mé analýzy jsou nekalé praktiky optimalizace, které budu zkoumat ze čtyř pohledů:

- vliv na hodnocení
- obtížnost provedení
- možná obrana
- riziko penalizace

„Vliv na hodnocení“ značí, jaký dopad má nekalá praktika na posun webových stránek ve výsledcích vyhledávání.

„Obtížnost provedení“ vysvětluje, jak jsou jednotlivé praktiky vytvářeny, za účelem bližšího pochopení souvislostí nekalé SEO optimalizace.

„Možná obrana“ popisuje, jakým způsobem se proti této praktice mohou bránit samotní uživatelé internetu (pokud je praktika namířena proti nim), nebo správci webových stránek (vývojáři, webmasteři), kterým navíc také dává doporučení, čeho se při vyvíjení webových stránek vyvarovat.

Pohled „riziko penalizace“ vysvětluje, jaké možnosti mají vyhledávače, aby tuto praktiku mohli snáze odhalit.

3.2 Analýza nekalých praktik

Pokud se blíže podíváme na stránky obsahující nekalé praktiky, zjistíme, že většinou obsahují kombinaci několika nekalých praktik dohromady. Abychom lépe porozuměli, jak tyto praktiky fungují, popíšu všechny z nich jednotlivě a uvedu příklady použití.

3.2.1 Opakující se klíčová slova v textu (repeated words, keyword stuffing)

Popis:

Jedná se o metodu, která je jednou z nejčastějších spamovacích praktik. Bývá používána zároveň s metodami, které popíšu v dalších podkapitolách. Celkový propojení nekalých praktik lze vidět na obrázku číslo 16.

Tato technika je založená na vložení velkého počtu klíčových slov do běžného textu stránky, za účelem zvýšení relevance.

Vliv na hodnocení:

Obsah stránek má vysokou váhu při hodnocení stránky, a proto sledávám tuto techniku, mající velký vliv na výsledky vyhledávání, jako nejvíce účinnou.

Obtížnost provedení:

Tuto techniku lze vytvořit jak manuálně, tak automaticky. Existují skripty i celé programy, které dokážou po zadání klíčového slova a procenta hustoty, rovnoměrně poskládat slova do textu a vygenerovat tak unikátní obsah stránky bohatý na požadovaná klíčová slova. Jedním z takových programů je například Website Content Wizard [17].

I pro začínajícího nekalého konkurenta je tedy velmi snadné zneužít tuto techniku pro potřeby optimalizace.

Možná obrana:

Metoda využívající opakující se klíčová slova je primárně zaměřená proti vyhledávačům, proto se ji běžní uživatelé nemusí obávat.

Pro správce webových stránek z toho vyplývá, že mají tvořit snadno čitelný text stránek, a to výhradně pro uživatele, nikoliv psát texty pro vyhledávače. Důvod je velice prostý. Pokud uživatel nenajde v textu požadované informace, odejde na jiný web, kde je snadněji najde. Existuje ovšem velmi tenká hranice pro určení ještě vhodného poměru klíčových slov k ostatnímu textu, kterou může správce webových stránek snadno překročit.

Riziko penalizace:

Při odhalování této praktiky si mohou vyhledávače spočítat procentuální výskyt klíčového slova (KW) obsaženého v textu, a pokud tato hustota překročí stanovenou mez, mohou automaticky penalizovat doménu, na které se stránka nachází.

Je to velice jednoduchý, účelný a přesný způsob automatického rozpoznání nekalé praktiky, který lze aplikovat na neomezené množství webových stránek. Na druhou stranu je relativně výpočetně náročný.

Příklad:

Pohled návštěvníka stránek	Pohled robota
KW Abc, def, KW, ghi, jkl, mno, KW, pqr, stu, KW, vwx, yza, bcd, KW	<body> <h1>KW</h1> Abc, def, KW, ghi, jkl, mno, KW, pqr, stu, <a>KW, vwx, yza, bcd, KW </body>

Obrázek 3 – Opakující se klíčová slova (KW)

3.2.2 Malý text

Popis:

Technika malý text je pokročilejší metodou opakujících se klíčových slov, která založená na metodě, při které je malý text pro uživatele webových stránek téměř nečitelný, a proto je tento uměle vytvořený text nebude mást.

Vyhledávače zjišťují obsah stránky ze zdrojového kódu, proto pro ně bude tento text stejně dobře čitelný.

Vliv na hodnocení:

Jelikož se jedná o rozšíření techniky „opakující se klíčová slova“, pouze s jistým stupněm maskování, má tato metoda stejně velký vliv na hodnocení.

Obtížnost provedení:

Vytvořit malý text na webových stránkách lze jednoduše hned několika způsoby:

- Opakovaným využitím tagu `<small>`
`<small><small><small>Text text</small></small></small>`
- Využitím tagu `<font>` s atributem velikosti písma
`<font size="1">Malý text</font>`
- S využitím CSS (Cascading Style Sheet)
`<p style="font-size: 20%">Malý text</p>`

Možná obrana:

Běžní uživatelé internetových stránek nejsou obětí této nekalé metody, proto se jí nemusí obávat.

Správci webových stránek by si měli dát pozor na velikost písma a sami posoudit, do jaké míry je velikost jimi používaného písma, čitelná.

Riziko penalizace:

Riziko penalizace je v tomto případě o něco menší, než u techniky využívajících pouze opakujících se slov, neboť se zde snižuje možnost udání pouze pomocí spam reportu. Je zde reálná šance, že ne každý si malého písma všimne.

Vyhledávače sice mohou ve zdrojovém kódu rozpoznávat tag `<small>` a atributy „size“ a „font-size“, ale tím neeliminují veškeré možnosti použití malého písma. Pokud bude zápis malého písma v připojeném externím souboru kaskádových stylů, vyhledávače nebudou schopny jednotlivé soubory procházet z důvodu nedostatečné výpočetní schopnosti.

Příklad:

Pohled návštěvníka stránek	Pohled robota
<pre>KW Abc, def, KW, ghi, jkl, mno, KW, pqr, stu, KW, vwx, yza, bcd, KW</pre>	<pre><style> body {font-size: 0.1em} </style> <body> <h1>KW</h1> Abc, def, KW, ghi, jkl, mno, KW, pqr, stu, <a>KW, vwx, yza, bcd, KW </body></pre>

Obrázek 4 – Malý text + opakující se klíčová slova

3.2.3 Skrytý text (hidden text)

Popis:

Text, ve kterém se vyskytují klíčová slova a který je nějakým způsobem před uživatelem, či vyhledávačem ukryt. Je to pokročilejší metoda ukrývání obsahu, než například zmenšováním písma, které je i tak pro vyhledávače čitelné. Ukrývání textu lze rozdělit do dvou kategorií.

Ukrytí textu před uživateli

Cílem této metody je schovat nežádoucí obsah stránky (nejčastěji opakující se klíčová slova) před uživateli, aby je nemohli vyhledávačům reportovat jako spam.

Ukrytí textu před vyhledávači

Na druhou stranu lze tuto techniku využít v případě, kdy je žádoucí zobrazit text pouze uživatelům webových stránek a naopak skrýt ho robotům vyhledávačů. V tomto případě se využívá technika, kterou roboti neumí odhalit. Je to například javascriptový kód, či text zobrazený jako obrázek. Takto prezentovaný text bývá především zaměřen na nevhodný obsah (pornografii, hazard, drogy, násilí, rasovou diskriminaci a další nezákonný obsah).

Vliv na hodnocení:

I zde je vliv na hodnocení relativně vysoký, a to z důvodu využití prvků, mající vliv na relevantnost tematického obsahu.

Obtížnost provedení:

Docílit skrytého textu lze několika různě náročnými způsoby, které mají odlišnou obtížnost odhalitelnosti.

Skrytí písma uživatelům webových stránek:

- HTML – nejjednodušší způsob, ale nejsnáze odhalitelný
 - Nastavením stejné barvy písma, jako barvy pozadí
`<font color="white">Skrytý text</font>`
- CSS
 - `<style> p {visibility: hidden;}</style>`
`<p>Skrytý text</p>`

- `<style> p {display: none;}</style>`
`<p>Skrytý text</p>`
- Překrytím písma libovolným obrázkem

Skrytí písma robotům vyhledávačů:

- JavaScript
 - Prosté vypsání textu

```
<script type="text/javascript">
var text = "Skrytý text";
document.write(text);
</script>           // vypíše „Skrytý text“
```
 - Napsání textu za využití znakové sady Unicode

```
<script type="text/javascript">
document.write(String.fromCharCode(83,107,114,121,116,253,32,
116,101,120,116));
</script>           // vypíše „Skrytý text“
```
 - Dalšími nesčetnými způsoby, které závisí pouze na programátorově fantazii.

Možná obrana:

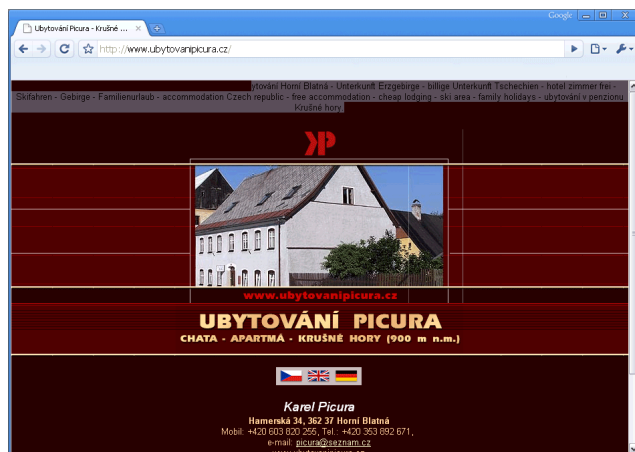
Jelikož se také jedná o rozšíření metody opakování klíčových slov, která není namířena proti uživatelům webových stránek, ani z této techniky jim neplyne žádné nebezpečí. Přesto by si měli tvůrci webových stránek dát pozor na volbu barev pozadí a samotného textu, neboť zvolením totožných, nebo málo kontrastních barev, mohou velmi snadno text skryt běžným uživatelům.

Riziko penalizace:

Vyhledávače v podstatě nemají možnost skrytý text odhalit, neboť nemají dostatečnou výpočetní schopnost procházet externí připojené soubory s kaskádovými styly. Javascriptových způsobů, jak napsat text, je obrovské množství a roboti nebudou nikdy schopni se všechny naučit.

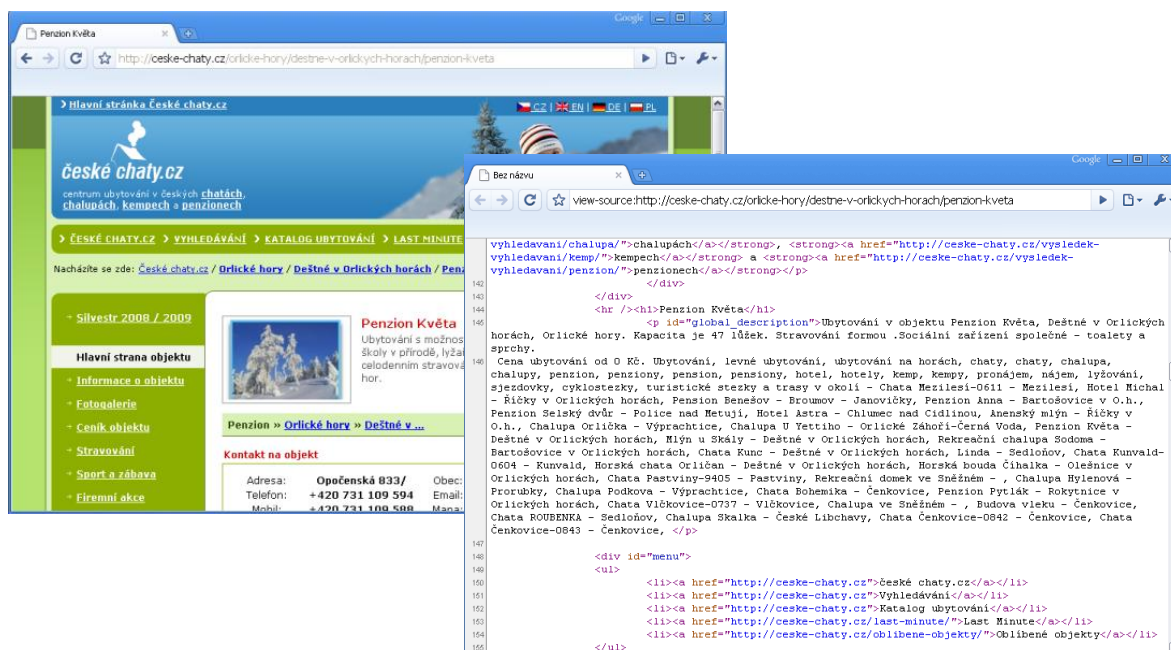
Příklad:

Na následujícím příkladu je, po označení kurzorem horní části webu, vidět skrytý text. Barva pozadí je totožná s barvou písma.



Obrázek 5 – Skryté písmo na <http://www.ubytovanipicura.cz/>³

Jedním z nejrafinovanějších způsobů je skrytí klíčová slova pod obrázek tak, jak tuto techniku praktikuje například server <http://ceske-chaty.cz/>⁴.



Obrázek 6 – Ukázka překrytí opakujících se klíčových slov obrázkem

3.2.4 Doorway pages (entry pages)

Popis:

Doorway page je nekalá praktika typu MFM (Made For Money). Jedná se o stránku, která je vytvořená za účelem získání vysokého hodnocení ve vyhledávačích.

Většinou se skládá z odstavce klíčových slov, doplněného o kontextové reklamy, či vnořného objektu se stránkou zadavatele. V extrémnějším případě je na této spamové stránce umístěno okamžité, javascriptové přesměrování na „zdravou“ stránku nekalého konkurenta.

Vliv na hodnocení:

Unikátní obsah stránek s relevantními klíčovými slovy, vyskytující se v přiměřené hustotě, je důvěryhodnou známkou pro vyhledávače, že se text daným tématem opravdu zabývá. Proto ho vyhledávače sledávají relevantním pro uživatele a nabídnou mu ho ve výsledcích vyhledávání na předních pozicích.

Obtížnost provedení:

K vytvoření doorway page existuje mnoho dostupných nástrojů, které během chvilky dokážou, po zadání zvoleného klíčového slova a procentuální hustoty s jakou se má v textu objevovat, vygenerovat unikátní obsah stránky. Tato stránka bývá umístěna na doménu (nejlépe

³ Dne 7. prosince 2008

⁴ Dne 29. listopadu 2008

na doménu vytvořenou pouze za tímto účelem), ze které se JavaScriptem automaticky přesměruje na doménu nekalého konkurenta.

K aplikování této praktiky však není zapotřebí využívat sofistikované programy, neboť ji lze snadno vytvořit nakopírováním klíčových slov do, nejlépe tematického, textu a na začátek stránky vložit javascriptové přesměrování. Vše lze vytvořit během několika málo minut.

Možná obrana:

Uživatelé se proti tomuto nechtěnému přesměrování mohou bránit vypnutím JavaScriptu v prohlížeči. Tím se z vyhledávače dostanou na doorway page, kde jim již ze samotného obsahu stránky bude patrné, že se jedná o podvodnou stránku a tlačítkem „zpět“ v prohlížeči se dostanou zpátky na výsledky vyhledávání.

Správci webových stránek by se měli vyvarovat používání JavaScriptového přesměrování a v případě nutnosti spíše využít stavových kódů serveru 301 Move permanently a 302 Move temporarily.

Riziko penalizace:

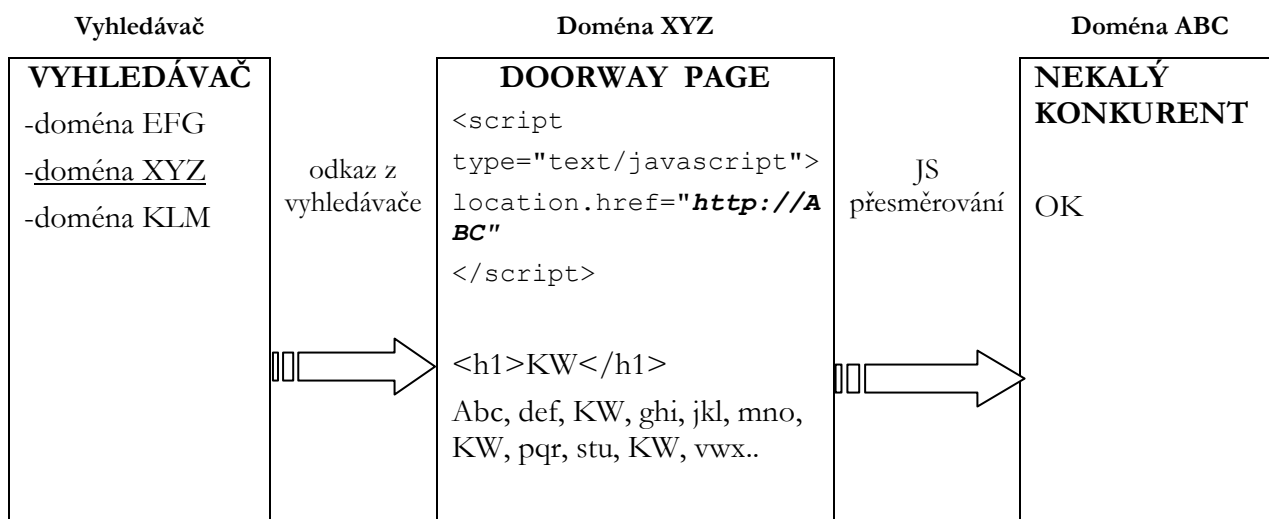
Vyhledávače jsou velice snadno schopny odhalit vysokou hustotu klíčových slov na stránce i samotnou klauzuli „location.href“ v tagu <script>. Pokud se ale doorway page nebude nacházet na stejné doméně, jako stránka na kterou vede přesměrování, nebudou moci automaticky doménu penalizovat. Důvodem je, že nebude prokazatelné, kdo takovou stránku vytvořil a zda to není jen způsob k odstavení konkurence.

Jediné co mohou vyhledávače udělat, je penalizovat samotnou doorway page, aby se ve výsledcích vyhledávání nezobrazovala na předních pozicích. To ovšem nepostihne nekalého konkurenta.

Zde zaniká riziko penalizace, neboť v případě dvou různých domén je dokázání této techniky v podstatě nemožné.

Jelikož se proti této technice dá jen velmi obtížně bránit a je těžce postihnutelná, riziko penalizace je nejmenší.

Příklad:



Obrázek 7 – Smysl a funkce doorway page

3.2.5 Cloaking

Popis:

Cloaking je výraz pro podstrkávání jiného obsahu webových stránek robotům, kteří je indexují a jiný běžným návštěvníkům stránek. Cílem této praktiky je oklamat vyhledávače a vydávat obsah stránek za obsahově kvalitnější, než je ve skutečnosti nabízen uživatelům.

Vliv na hodnocení:

Touto technikou lze vytvořit velmi kvalitní stránku, poskládanou z několika různých nekalých metod optimalizace, která bude vysoce hodnocená roboty vyhledávačů.

Obtížnost provedení:

Zde je velmi dobře vidět, že pro podstrčení různého obsahu stránek není zapotřebí žádného složitého mechanismu. Na následujícím příkladu, který je uveden v programovacím jazyku PHP, vidíte podmínku, která nejdříve ověří, zda se skutečně jedná o robota vyhledávače Google a na základě toho nabídne samotný obsah stránek.

```
<?php
    if ($_SERVER['HTTP_USER_AGENT'] == "Goog lebot") {

        echo "Obsah stránky pro robota vyhledávače Google.";

    } else {

        echo "Obsah stránky pro běžné návštěvníky.";

    }
?>
```

Možná obrana:

Uživatel se této techniky nemusí obávat, protože není namířena proti němu. Běžným procházením internetu ji ani není schopen odhalit.

K této praktice může docházet také nedopatřením při „podstrkávání“ jazykové verze návštěvníkům na základě jejich IP adresy, či nastavení jazyka v prohlížeči. Správci webových stránek by měli být opatrní při nadměrnému vycházení vstříc uživatelům stránek a nechat je si zvolit jazykovou verzi webu podle svých preferencí.

Riziko penalizace:

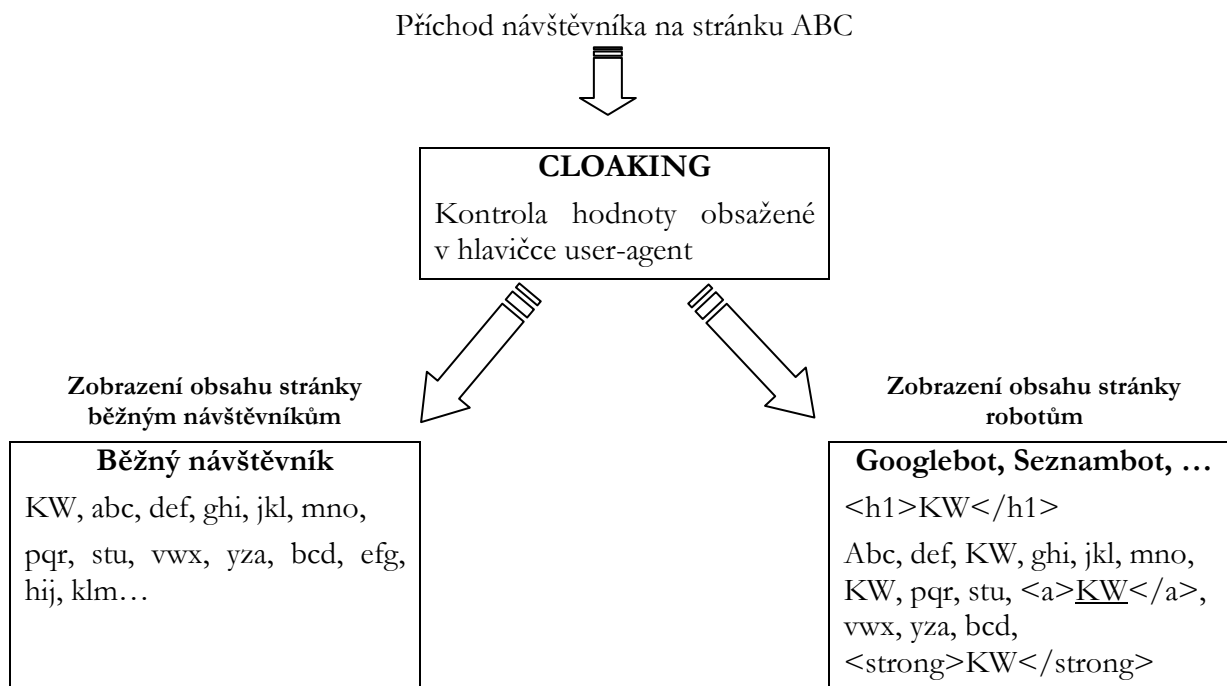
Přestože je tato metoda velmi těžce odhalitelná, najdou si i případy, kdy konkurence objeví tyto nekalé techniky a pomocí spam reportu informuje vyhledávač o jejím používání. Vhodným nástrojem pro manuální změnu hlavičky user-agenta je například rozšíření prohlížeče Mozilla Firefox s názvem User Agent Switcher [18].

Vyhledávač má možnost odhalit tuto techniku v případě, že na stránku alespoň dvakrát pošle stejného robota. Jednou však s klasickým veřejným označením (Goog lebot, Seznambot, apod.) a podruhé se bude vydávat za běžného uživatele a v hlavičce user-agenta bude zobrazovat označení libovolného prohlížeče webových stránek. Následné porovnání obou výsledků indexování by odhalilo případný cloaking a automaticky by mohla být doména penalizována.

Aby robot zamezil penalizování stránky, která bude nabízet uživatelům jazykovou verzi webu dle jejich IP adresy, může na stránky přijít z různých IP adres. Pokud dva roboti z různých IP

adres a se stejným označením user-agent⁵, najdou ve stejný čas dva různé obsahy stránky, rozpoznají tak, že se jedná o nabídnutí jazykové verze.

Příklad:



Obrázek 8 – Cloaking v praxi

3.2.6 Automaticky generovaný obsah (Automatic Content Generation)

Popis:

Automaticky generovaný obsah je název pro službu, která vygeneruje unikátní obsah podle zadaných klíčových slov. Tento obsah je následně vložen do uceleného webu a brzy od vyhledávačů získává obrovské hodnocení on-page faktorů.

Vliv na hodnocení:

Na automaticky generované webové stránky je většinou velmi obtížné získávat zpětné odkazy, a proto se nezobrazují na jednoslovná klíčová slova na vyšších pozicích ve výsledcích vyhledávání. Naproti tomu ale získávají značnou návštěvnost na konkrétnější fráze (Long Tail).

Obtížnost provedení:

K vytvoření automaticky generovaných stránek je nevyhnutelně zapotřebí použít skript, nebo aplikaci, která vygeneruje obsah. Podle zvolených klíčových slov najde obsah na internetu (v encyklopedii Wikipedia, tematicky blízkých článků, v RSS [Really Simple Syndication], na YouTube atp.) a tento vloží do stránky takovým způsobem, že pro vyhledávače je poměrně náročné zjistit, že se jedná o výtah z více cizích webů. Mezi velmi využívané aplikace k tomuto účelu patří především: MyGen, Yacg a také RSSGM.

⁵ S každým dotazem na webový server se v hlavičce také posílá označení webového klienta.

(např. User-Agent: Mozilla/5.0 (Windows; U; Windows NT 5.1; cs; rv:1.9.0.4) Gecko/2008102920 Firefox/3.0.4)

Možná obrana:

Automaticky generovaný obsah je pro zkušeného uživatele poměrně jednoduché rozeznat. Většinou se jedná o extrémně rozsáhlé stránky, (obsahující i několik desítek tisíc optimalizovaných podstránek) s velmi jednoduše vytvořenou informační architekturou. Tyto stránky mají skoro vždy vytvořené dynamické prolinkování v rámci celého webu, kde např. z patičky jedné podstránky vede 99 odkazů na jiné podstránky. Účelem je lepší indexace takového obsahu.

Riziko penalizace:

Riziko penalizace je velmi vysoké. Pro vyhledávače je velmi jednoduché zjistit, že nová doména od prvního zaindexování má například 30 000 podstránek. Dále se např. Google naučil pracovat s varovnými znaky přemíry zpětných odkazů (overlinking), které je charakteristické pro automaticky generovaný obsah. Nestává se ovšem často, aby Google tyto stránky automaticky smazal z indexu. Penalizace přichází až po několika týdnech i měsících. Otázkou zůstává, zda to není práce konkurence, která tyto praktiky nahlásila pomocí spam reportu.

Příklad:

Existuje velké množství automaticky generovaného obsahu, který je pro uživatele poměrně náročné odhalit, jelikož se tyto stránky zobrazují ve výsledcích vyhledávání na velmi nespecifické a málo vyhledávané fráze. Demo, jak vypadá práce s Yacg, najdete na adrese

<http://yacgmassinstaller.com/screencast/> .

3.2.7 Doménový spam

Popis:

Takto se označuje web, který má několik desítek až stovek domén druhého řádu se stejným designem a hlavně s málo odlišným obsahem. Nejčastěji bývá vytvořen za účelem, pokrýt co nejvíce tematických oborů, pro získání vysoké návštěvnosti.

Vliv na hodnocení:

Doménový spam se dá vytvořit různými způsoby, z nichž některé jsou pro vyhledávače i uživatele takřka nerozpoznatelné.

Příklad

Subdomény s různými klíčovými slovy, které jsou navzájem odlišné i ve vztahu LSI. Pokud se využije relevantní obsah k dané doméně, je tento způsob často využíván jako affiliate kanál (poskytování reklamního prostoru za provizi).

Dalším možným případem je hrubě vytvořený doménový spam, který se využívá například na příbuzná klíčová slova napříč subdoménami.

Příklad

Hrubě vytvořený doménový spam, který se nesnaží o netematický obsah na subdoménách, může vypadat například takto:

- letenky.ABC.cz
- levne-letenky.ABC.cz
- letenky-levne.ABC.cz
- online-letenky.ABC.cz

atd.

Tento doménový spam je vyhledávači lehce odhalitelný, a proto se na rozdíl od prvního příkladu, vyskytuje velmi málo na vyšších pozicích ve výsledcích vyhledávání.

Obtížnost provedení:

Vytvořit doménový spam se dá relativně snadně. V podstatě existují dva nejrozšířenější způsoby:

- Manuálně se vytvoří samostatná stránka pro každou subdoménu
- Subdomény na zvolená klíčová slova se vytvoří z adresářů. K tomu například poslouží zápis v konfiguračním souboru .htaccess.

Problémem doménového spamu je nemožnost provést plnohodnotný linkbuilding, protože je velmi snadno odhalitelný, a uživatel, stejně jako katalogový administrátor, ihned odhalí neužitečnost dané stránky. Doménový spam se nejčastěji využívá jako způsob optimalizace za účelem získání vyšší návštěvnosti (traffic). Typickým znakem těchto webů je existence reklamy, která přináší majiteli webu zisk, za jehož účelem byl web vytvořen. Nejčastěji se jedná o obsahovou síť reklam PPC (AdSense).

Možná obrana:

Uživatelé internetových stránek snadno rozpoznají, že pro ně webové stránky na různých subdoménách tohoto webu nemají žádný přínos a proto je brzy opouštějí. Aby nepodporovali tyto nekalé praktiky, doporučuji jim z takové stránky rovnou odejít a hlavně neklikat na reklamu, která je hlavním výnosem těchto stránek. Doménový spam je však primárně namířen proti vyhledávačům, proto se ho uživatelé internetových stránek nemusí obávat.

Správci webových stránek by měli dodržovat pravidlo odděleného tematického obsahu na různých subdoménách, umístěných na jedné doméně.

Riziko penalizace:

Problém doménového spamu mají vyhledávače jen omezenou možnost odhalit. Pokud nebudou na subdoménách tematicky podobné obsahy, vyhledávače nejsou schopni tuto metodu identifikovat.

Naopak, velmi snadno mohou odhalit web, jehož subdomény jsou velmi blízké, či naprosto stejně tematicky zaměřeny. Stačí jim k tomu porovnání jednotlivých subdomén na základě metody LSI v rámci samotného webu.

Jelikož jsou subdomény běžným způsobem, jak oddělit tematicky odlišné obsahy v rámci webu, je subdoménový spam poměrně náročné odhalit. Existuje totiž pravidlo, že na subdomény by se měl umísťovat obsah odlišný od tematiky webu. Když je doménový spam vytvořený sofistikovaně a neokaté, je skoro nemožno ho odhalit. Pokud však webmaster vytvoří např. 100 subdomén, které jsou navíc tematicky blízké, vyhledávače nemají problém takový spam odhalit.

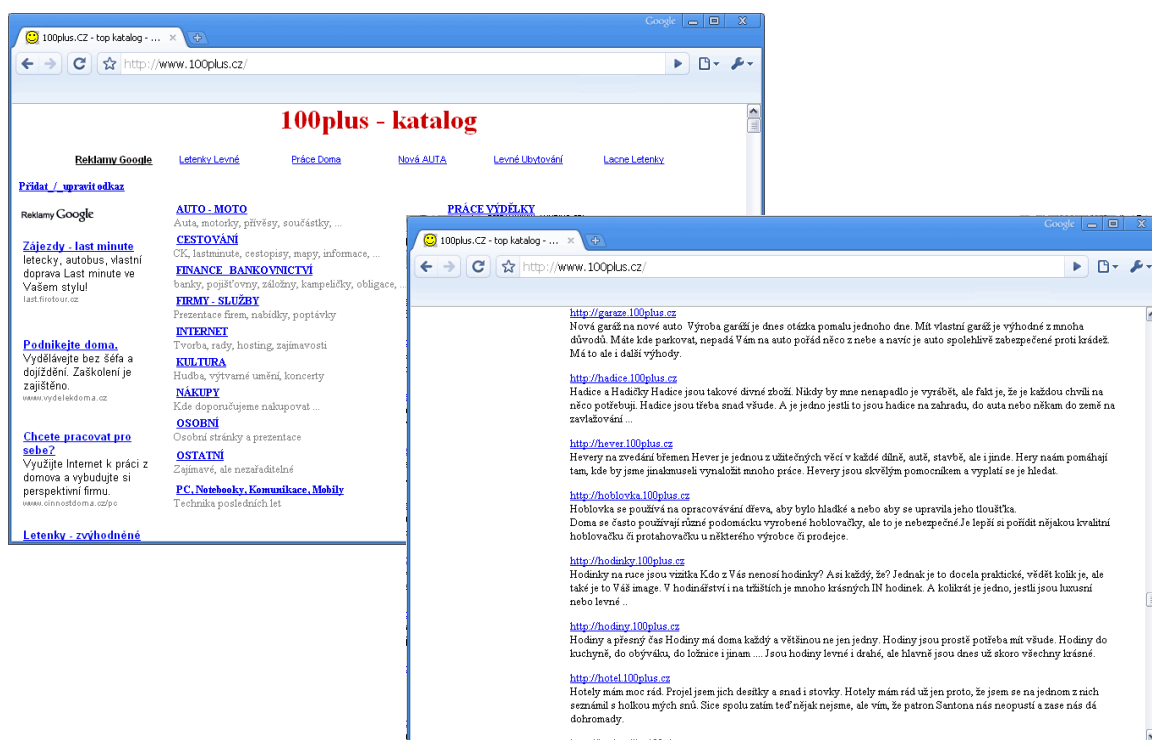
Příklad:

Názorným příkladem je doména www.100plus.cz⁶, která obsahuje více než **170 subdomén** druhého řádu s rozdílným tématem na jednotlivých subdoménách.

Např. <http://cestovani.100plus.cz>, <http://www.dane.100plus.cz>, <http://dieta.100plus.cz>, <http://hotel.100plus.cz> atd.

Na hlavní stránce lze v levém sloupci i těsně pod nadpisem „100plus – katalog“ nalézt reklamní systém Google AdSense, kvůli kterému byl celý katalog vygenerován.

Nejdeálnější případ pro majitele webu může vypadat například takto. Uživatelé, kteří přijdou na tuto spam stránku, zde sice nenaleznou kvalitní obsah, proto je nejspíš zaujme obsahová reklama Google AdSense a odejdou z tohoto webu skrze reklamní systém. Majitel takto vytvořených stránek posléze dostane zaplacen od společnosti Google za přivedení zákazníků inzerujícím internetovým stránkám. Nejen, že se tvůrci tohoto spamu podařilo oklamat robota vyhledávače, který všechny spamové stránky indexuje, ale dokonce i samotnou společnost Google, která s ním spolupracuje při provozování reklamních inzerátů v obsahové síti.



Obrázek 9 – Příklad doménového spamu na webu www.100plus.cz⁷.

3.2.8 Spamování diskuzí

Popis:

Nejedná se o zakázanou techniku, ale přinejmenším neetickou. Na internetu ji nejčastěji můžeme najít na serverech, které umožňují svým uživatelům zapojovat se do dění na webu a skrze textové formuláře reagovat na napsané články. Typickým příkladem jsou blogy či

⁶ Dne 29. listopadu 2008

⁷ Dne 4. prosince 2008

zpravodajské portály, které bývají tematicky zaměřeny a oslovují uživatele, mající k danému oboru bližší vztah. Oslovit tuto skupinu lidí, znamená oslovit potenciální zákazníky a to ve větší míře, než by tomu bylo u nediverzifikované cílové skupiny. Cílem nekalého konkurenta je psát do diskuze pod tematicky spřízněné články a šířit mezi ostatními čtenáři takzvaný PR efekt. Neboli vychválit vlastní zboží či službu jako nezávislý pozorovatel a podnítit k nákupu. Často bývá připojený i odkaz na stránku nekalého konkurenta. Dalším cílem je získat zpětné odkazy a posílit tak hodnocení odkazované stránky.

Spamování diskuzních fór se vyskytuje v několika různých variantách s různými cíly a efekty. Probíhá buďto automaticky, za pomoci nástrojů vytvořených speciálně za tímto účelem, nebo manuálním vypisováním polí formuláře. Nyní popíšu, jak se od sebe tyto formy spamování liší.

Automatizované spamování

Nejčastějším způsobem spamování diskuzí je využívání spamovacího robota. Pod tímto pojmem si můžeme představit program, který za nás vloží a odešle požadovaný text do formulářů webových stránek. Využívání tohoto nástroje má výhodu v odeslání mnoha formulářů za velmi krátkou dobu. Nevýhodou je menší úspěšnost odeslání formuláře, než je to u manuální metody, v nemožnosti reagovat na předešlé příspěvky a fakt, že všechny rozeslané komentáře budou vypadat naprosto stejně, tudíž je relativně snadné je identifikovat jako spam.

Manuální spamování

Také jsou známy případy, kdy si provozovatelé webových stránek najdou tematicky spřízněné diskuze a tam si na své otázky, zaslané z jedné IP adresy, sami odpovídají z jiné IP adresy a uvádějí odkaz na své internetové stránky. Tento způsob je o mnoho pomalejší, než využití spamovacích robotů, je ovšem daleko flexibilnější. Manuálním vkládáním komentářů lze relevantně reagovat na předešlé příspěvky a úspěšnost odeslání formuláře je téměř stoprocentní.

Vliv na hodnocení:

V případě, že se podaří najít nezabezpečené diskuze proti spamování, je tento způsob účinnou zbraní při vytváření zpětných odkazů a posilování hodnocení stránek nekalého konkurenta. Jelikož se ale zabezpečení diskuzí stále zdokonaluje, šance najít nějaké nezabezpečené se stále snižuje.

Obtížnost provedení:

Existuje mnoho dostupných robotů, kteří dokážou procházet internet a vkládat do diskuzí požadovaný obsah. Jejich úspěšnost je ovšem s přibývajícím ochrannou formulářů stále menší.

Samotné manuální vkládání je velmi neefektivní, v poměru vynaloženého času na získání zpětného odkazu či potenciálního zákazníka, a proto je jen zřídka využíváno.

Možná obrana:

Spamování diskuzních fór se v poslední době velmi rozšířilo a je potřeba se proti němu bránit. Zabránit automatickému vyplňování formulářů se dá různými způsoby. Asi nejrozšířenější technikou obrany je captcha. Před odesláním vyplněného formuláře je uživatel povinen opsat z obrázku zobrazovaný text. To robotům velmi znesnadňuje práci a velké množství z nich tuto překážku nezdolá. Existují ovšem i automatické způsoby, jakým roboti obrázek přečtou a odešlou formulář.

Dalším častým zabezpečením je nastavení určitého časového limitu, který musí uplynout mezi

načtením stránky diskuze a odesláním polí formuláře. Roboti totiž nejčastěji odesílají formuláře okamžitě ihned po načtení stránky. Tímto způsobem se dá ve většině případů také odlišit uživatele od robota.

Nejvhodnějším způsobem zabezpečení formuláře by ovšem mělo být unikátní řešení, které nevychází z těch nejběžnějších, ale je kombinací několika různých ochranných mechanismů. Nepřekonatelné řešení ale neexistuje.

Pokud se již robot dostane přes bezpečnostní prvky, je potřeba mu zabránit v dosažení jeho cíle. Je potřeba zabezpečit diskuzi proti zveřejnění odkazu na stránku nekalého konkurenta. V této situaci nechceme zobrazovat žádný odkaz, proto nerozlišujeme mezi robotem a „živým“ uživatelem.

Odkaz uvedený v diskuzi, znamená předání hodnocení stránky diskuze odkazované stránce nekalého konkurenta. Aby diskuze nepředávaly hodnocení stránky dál, lze uvést u odkazu (tag <a>) atribut rel=“nofollow“, a ten tímto zamezí předání hodnocení odkazu. Přesto samotný odkaz umožní čtenářům dostat se snáze na propagovanou stránku, než kdyby byli nuceni, si podle názvu samotný web dohledat sami.

Některé diskuze využívají již hotová řešení pro vkládání příspěvků, která pouze implementují do své stránky. Jedná se o takzvané WYSIWYG editory, které umožňují formátovat vkládaný text. Existuje jich celá řada a používají různé zápisy formátovaných prvků. Vedle formátování písma pro tučný text či kurzívu, lze pomocí kódu zapsat zobrazení odkazu. Správci webových stránek by neměli řešit zákaz vkládání odkazů pouze tím, že z WYSIWYG editoru odeberou ikonku pro vložení odkazu, ale měli by také smazat zdrojový kód, který formátuje zápis ve formuláři. Pokud tak neučiní a nekalý konkurent vyzkouší vložit několik možných způsobů⁸ formátování, které nejběžnější WYSIWYG editory používají, s největší pravděpodobností některý z nich vloží odkaz do diskuze.

Riziko penalizace:

Přestože mohou vyhledávače tuto techniku relativně spolehlivě identifikovat, penalizace za ni v současné době v podstatě nehrozí. Nelze totiž prokázat, kdo je za ni zodpovědný. Zda majitel samotné stránky, kam odkaz míří, či nekalý konkurent, který se touto technikou snaží poškodit svého konkurenta.

Postihování mohou řešit správci webových stránek samotných diskuzí a to podle vlastního uvážení. Mohou uživateli zakázat pouze přispívání do diskuze, na základě jeho identifikace podle IP adresy a user-agenta, nebo dokonce zakázat přístup k celému webu. Toto řešení nebude ovšem nikdy stoprocentní, neboť tentýž uživatel může ve spamování diskuzí pokračovat z jiného počítače.

Diskuze plné spamu odrazuje uživatele k přispívání a snižuje spokojenost uživatelů s webem. Nejen proto by měli správci webových stránek dbát na jejím zabezpečení.

V případě, že vložený odkaz bude odkazovat na webové stránky nevhodného obsahu (např. pornografie), může být doména diskuze penalizována za (bad neighbour). Jelikož napadání diskuzí je především výsadou webů nevhodného obsahu, je riziko penalizování přímo úměrné jejímu zabezpečení.

Příklad:

Konkrétní příklad vytváření dobrého jména společnosti nalezneme v diskuzi na serveru Jak psát web [19]. Zde se majitel domény, a zároveň provozovatel společnosti Z, pan XY, vydával

⁸ Viz Příklad: Různé metody vložení odkazu do diskuzí.

pod anonymním uživatelským jménem za spokojeného zákazníka a uživatele svých vlastních služeb. Odhalen byl moderátorem diskuze, neboť IP adresa anonymního uživatele se shodovala s IP adresou pana XY, který v diskusi vystupoval i pod svým pravým jménem.

Různé metody vložení odkazu do diskuzí.

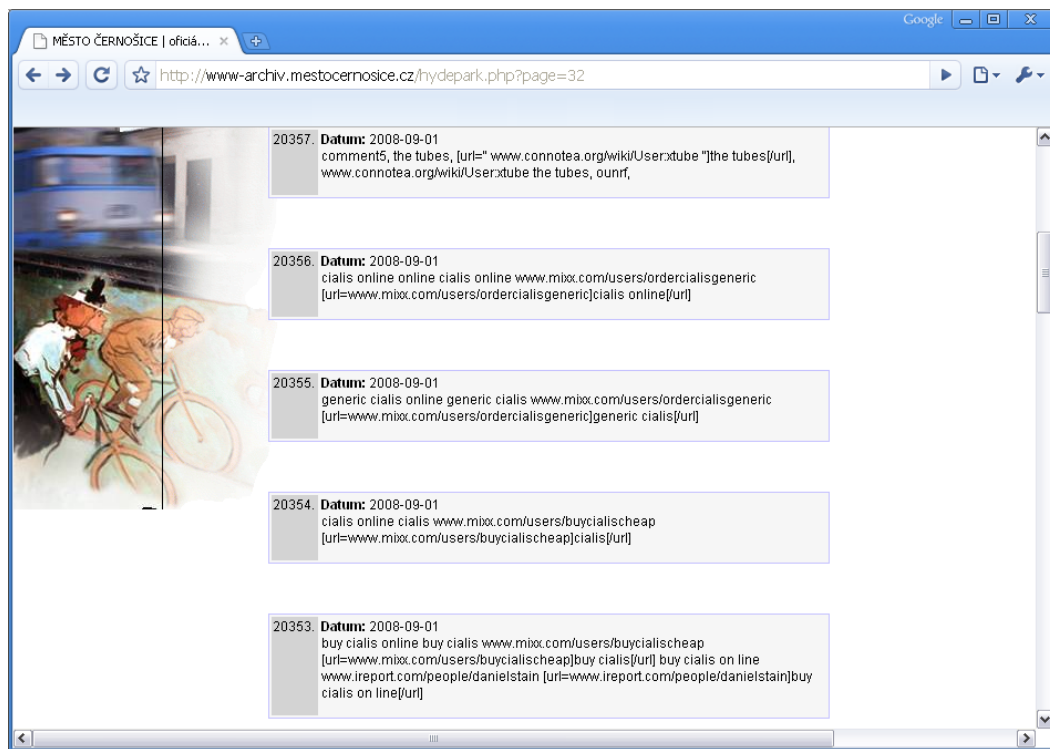
```
http://www.stranka-nekaleho-konkurenta.cz/nejlepsi-vyrobek.html  
[http:// www.stranka-nekaleho-konkurenta.cz nejlepší výrobek]  
"nejlepší výrobek": www.stranka-nekaleho-konkurenta.cz  
[url]http://www.stranka-nekaleho-konkurenta.cz[/url]  
[url=http://www.stranka-nekaleho-konkurenta.cz] nejlepší výrobek [/url]  
[link http:// www.stranka-nekaleho-konkurenta.cz]nejlepší výrobek[/link]  
[link=http://www.stranka-nekaleho-konkurenta.cz] nejlepší výrobek[/link]  
<a href="http://www.stranka-nekaleho-konkurenta.cz">nejlepší výrobek</a>
```

Několik příkladů vložení kódu do stránek diskuze, který automaticky přesměruje na stránku nekálního konkurenta.

```
[img]http://www.stranka-nekaleho-konkurenta.cz/presmerovani.php[/img]  

```

Pokud je formulář proti zobrazování odkazů vhodně zabezpečen, ovšem nedostatečně oproti robotům, můžeme vidět na následujícím příkladu.



Obrázek 10 – Ukázka spamových příspěvků v diskusi stránek města Černošice⁹

⁹ Dne 2. prosince 2008

3.2.9 Klamné přesměrování (deceptive redirects)

Popis:

Tato metoda bývá využívána k přesměrování uživatele ze stránky, na které se nachází, na stránku novou. Ta většinou ani neodpovídá vyhledávanému obsahu. Cílem této techniky je oklamat vyhledávače i uživatele webových stránek. Jedná se o jednu z nejvyužívanějších metod nekalých SEO praktik.

Vliv na hodnocení:

Klamné přesměrování nemá přímý vliv na hodnocení stránek ve výsledcích vyhledávání, ale je jednou z metod, která dokáže spamovou stránku před vyhledávači ukrýt.

Obtížnost provedení:

Tato metoda se nejčastěji provádí javascriptovým přesměrováním, které uniká pozornosti vyhledávačů. Lze ji rozdělit do dvou skupin.

Bez vědomé akce uživatele:

- HTML přesměrování po zvolené době
 - `<meta HTTP-EQUIV="Refresh" CONTENT="0; URL=http://ABC">`
- Spuštění s načtením stránky
 - `<body onLoad="location.href='http://ABC'">`
 - `<script type="text/javascript">`
`location.href="http://ABC"`
`</script>`
- Přjetím prvku myši
 - `<a href="#" onMouseOver="location.href='http://ABC'">`
`nepřejíždět myši`
`</a>`
- A další techniky.

Na základě vědomé akce uživatele:

- Kliknutím na odkaz
 - `<a href="http://www.velka-zajimavost.cz"`
`onMouseDown="location.href='http://ABC'">podívej se`
`</a>`
 - Tímto zápisem lze docílit zobrazení adresy, smyšleného serveru „Velká zajímavost“, vlevo dole na stavovém řádku prohlížeče a odkázání na stránku ABC.

Javascriptových funkcí, vyvolávajících akci, lze využít daleko víc. Např. `onBlur()`, `onError()`, `onFocus()`, `onChange`, `onReset()`, `onSelect()`, `onSubmit()`, `onUnload()` atd.

Možná obrana:

Uživatelé internetových stránek se mohou bránit vypnutím podpory javascriptu ve svém webovém prohlížeči, čímž předejdou většině způsobů nechtěného přesměrování.

Správci webových stránek by měli využívat přesměrování co nejméně, a to pouze v případě, který je nezbytný pro zajištění správné funkčnosti webu a pokud návštěvníka neodkazují na stránku, kam si sám nepřeje.

Samotné přesměrování by neměli provádět pomocí jazyka javascript, ale měli by využívat stavových kódů serveru 301 Moved Permanently, nebo 302 Moved Temporarily, které vyhledávače dokážou rozpoznat a vyhodnotit tak povahu přesměrování.

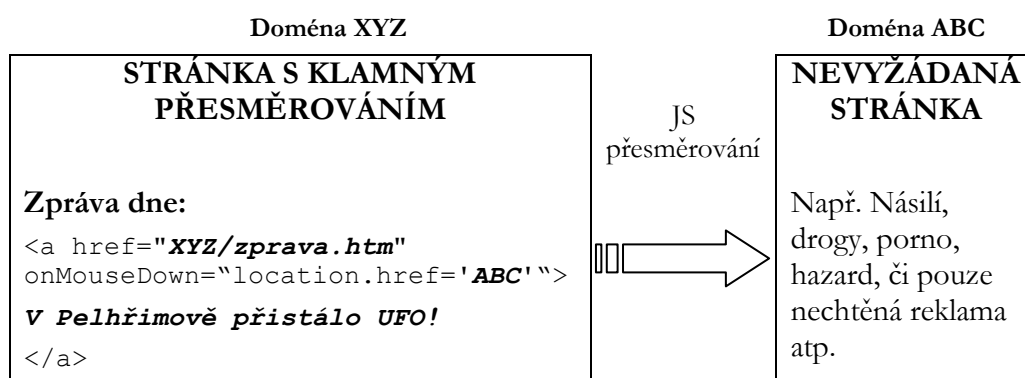
Správci stránek by měli dbát na zabezpečení své webové prezentace proti podstrčení nebezpečného zdrojového kódu (XSS – Cross Site Scripting) a zamezit tak útočníkům využít bezpečnostních děr a přesměrovat stránku kamkoliv si jen dokážeme představit. Pokud na svém webu provozují některý z volně dostupných redakčních systémů, měli by ho udržovat aktualizovaný, aby snížili riziko útočnicko napadení.

Riziko penalizace:

Na základě přesměrování, využívajících stavových kódů serveru, jsou vyhledávače velmi dobře schopny procházet odkazy a analyzovat cílovou stránku přesměrování. Naopak, je-li přesměrování provedeno přes javascript, vyhledávače na něj nejsou schopny přijít a zhodnotit tak povahu odkazu.

Dokud roboti nebudou mít dostatečnou výpočetní schopnost k procházení a parsování javascriptu, nikdy nebudou schopny tuto nekalou praktiku odhalit a automaticky penalizovat. Samotné odhalení formule „location.href“ je pro vyhledávače velmi snadné, ale pokud je spojena s nějakou funkcí vyvolávací akcí, nejsou schopny určit, zda se skutečně jedná o nekalou praktiku.

Příklad:



3.2.10 Odkazové farmy (link farms)

Popis:

Odkazová farma je název pro službu, která sdružuje několik stovek, či tisícovek, navzájem odkazujících stránek. Jejich prolinkováním mezi sebou mohou získat vysoké hodnocení (např. PageRank) od vyhledávačů.

Vliv na hodnocení:

Počet zpětných odkazů má veliký vliv na hodnocení stránky, proto je vyhledávači velmi poctivě sledován. Čím více zpětných odkazů povede na stránku, tím bude mít stránka větší důvěru u vyhledávačů, neboť samotné odkazy na ni vypovídají o její kvalitě.

Obtížnost provedení:

Zapojení se do výměnného systému zpětných odkazů může být různě obtížné. Většina farem funguje za úplaty, neboť samotné naprogramování systému je u dobře navržených farem relativně složitá záležitost. Naopak jednoduché farmy, které nedbají na důmyslnější maskování před vyhledávači, bývají volně dostupné a velmi rychle penalizované.

Možná obrana:

Správci webových stránek by měli dbát na vytváření zpětných odkazů přirozenou cestou (kvalitní obsah, na který budou uživatelé sami odkazovat).

Ve zcela hypotetickém případě, v rámci konkurenčních bojů, může nekalý konkurent koupit zpětné odkazy pro svého soupeře, kterého mohou vyhledávače, na základě automatického vyhodnocení, penalizovat. Ten potom nemá jinou možnost, než podat vysvětlení a věřit v jeho vyslyšení. Např. vyhledávač Google [20] dává možnost obětem těchto praktik vysvětlit, jak k této situaci mohlo dojít.

Veškerá tvrzení ovšem nelze ve většině případů věrohodně podložit důkazy, proto se vyhledávače řídí jakousi mírou důvěry (trust rank) v napadený web. Pokud by napadený web patřil velké společnosti s dlouholetou tradicí, je zde malá pravděpodobnost, že by vyhledávače trvaly na jeho penalizování. Naopak je penalizování pravděpodobnější v případech nových, nezavedených webových stránek, které nemají dlouhou historii.

Riziko penalizace:

Pro vyhledávače je velmi snadné porovnat mezi sebou index jedné stránky v rozmezí např. jednoho týdne. Pokud vyhledávač zaznamená nepřiměřený nárůst zpětných odkazů, může automaticky stránky penalizovat. Velký skok, v nárůstu počtu zpětných odkazů, upozorňuje na nekalou manipulaci s odkazy, jejich nákupem, či výměnou. Umělé získání zpětných odkazů samo o sobě nehovoří o kvalitě stránky, proto ji vyhledávače kontrolují a postihují.

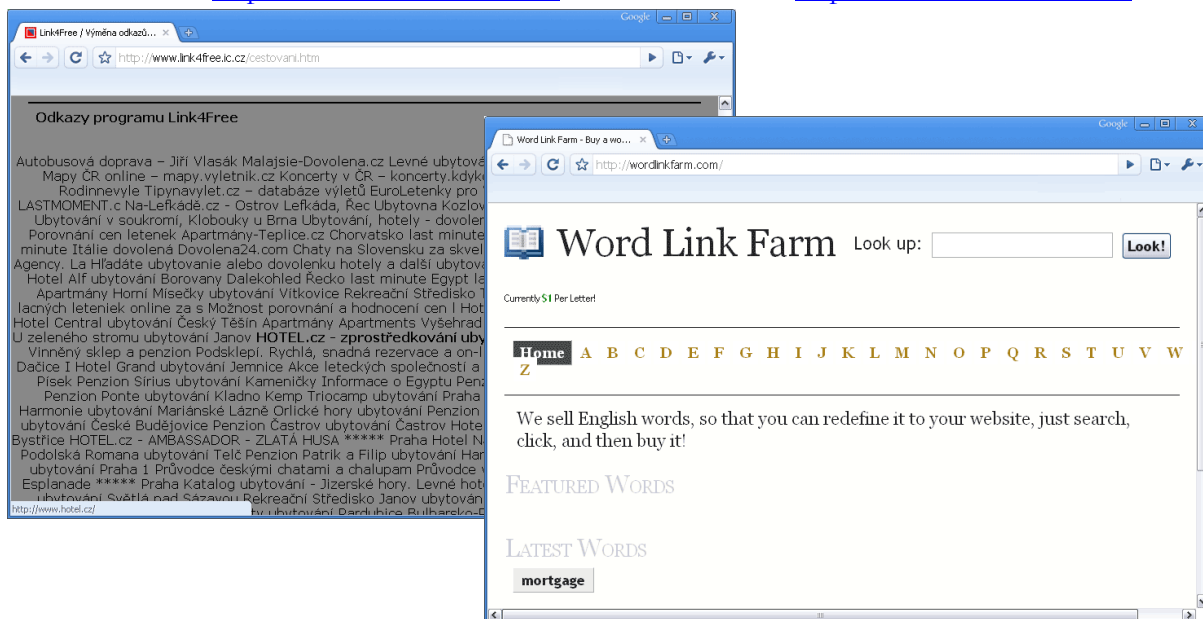
Podle mých zkušeností vyhledávače tuto techniku, jako jednu z mála, automaticky penalizují, neboť její rozpoznání není těžké a nekalé praktiky jsou velmi snadno prokazatelné.

Příklad:

Existuje velké množství linkových farem, které se od sebe liší svou funkcí, rozsahem zapojených webů, způsobem používání, vlivem na hodnocení stránek a maskováním před vyhledávači.

<http://www.link4free.ic.cz/>¹⁰

<http://wordlinkfarm.com/>¹¹



Obrázek 11 – Ukázka seznamu stovek tematicky seřazených odkazů a webu, zabývajícím se placeným nákupem odkazů

3.2.11 Patickové odkazy

Popis:

Nejedná se o zakázanou techniku, ale vyhledávače by ji měly do budoucna umět rozpoznat a vyřazovat při počítání hodnocení stránky, neboť uváděním zpětných odkazů v patičce stránky dochází ke snižování relevantnosti a smyslu odkazovat na zajímavé a tematicky spřízněné weby.

Jak již sám název napovídá, patickové odkazy byly pojmenovány, podle umístění zpětných odkazů, do patičky webové stránky. Pro uživatele však uvedení samotného odkazu nepřináší žádnou relevantní informaci o obsahu odkazované stránky, proto se je například vyhledávač Google učí rozpoznávat a vyřazovat při počítání zpětných odkazů.

Patickové odkazy jsou přímým nástrojem odkazových farem, které této metody využívají.

Vliv na hodnocení:

Odkazy mají doposud velký vliv na hodnocení stránky například vyhledávačem Google. Čím více odkazů povede na webovou stránku, tím větší bude mít hodnocení a důvěru vyhledávačů, že je opravdu kvalitní. To je důvodem výměny těchto odkazů mezi internetovými stránkami.

Čím vyšší hodnocení samotná stránka má, tím vyšší hodnocení předává stránce, na kterou odkazuje. Více o počítání hodnocení stránky jste se mohli dočíst v kapitole 2.1.

Obtížnost provedení:

¹⁰ Dostupné dne 30. listopadu 2008

¹¹ Dostupné dne 30. listopadu 2008

Existují různé metody výměny zpětných odkazů uváděných např. v patičkách. Od přímé výměny odkaz za odkaz (A->B, B->A), až po sofistikovanější výměny mezi čtyřmi weby dvou správců webových stránek (A->B, C->D), které jsou v podstatě skoro neodhalitelné.

Proces manuálního získávání zpětných odkazů je ovšem poněkud zdoluhavý, proto jsou správci stránek často zlákáni k využití odkazových farem, které jim během chvilky mohou získat velké množství zpětných odkazů.

Možná obrana:

Správci webových stránek by měli dbát na to, aby na svém webu uváděli kvalitní informace, kterými sami přimějí provozovatele ostatních webů na ně odkazovat. Tím by si web získával přirozené zpětné odkazy tak, jak vyhledávače předpokládají. Měli by také odkazovat na tematicky podobně zaměřené weby.

Riziko penalizace:

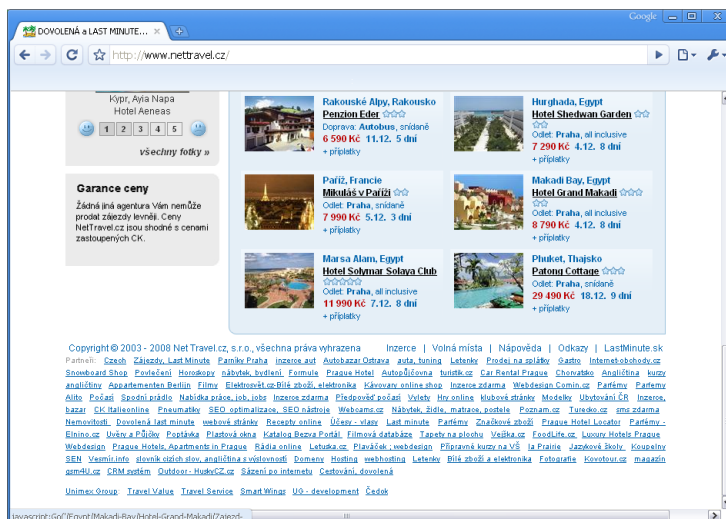
Vyhledávače by se měly snažit odlišit zpětné odkazy, které jsou vytvořeny pouze za účelem posílení pozice ve výsledcích vyhledávání a upřednostňovat ty přirozeně vytvořené na základě jednostranného odkázání (A->B). Porovnáním odkazů z webu A s odkazy webu B mohou zjistit, že weby na sebe vzájemně odkazují, a tím nezapočítat tyto odkazy do celkového hodnocení konkrétní stránky.

Vždy se ovšem najde způsob výměny zpětných odkazů ($A \rightarrow B$, $C \rightarrow D$), které vyhledávače nebudou schopny spolehlivě odhalit. Proto sledávám obranu proti umělé výměně zpětných odkazů za středně nesnadnou.

Stejně, jako u odkazových farem, i u této metody platí, že rychlý nárůst zpětných odkazů mohou vyhledávače rozpoznat a penalizovat. S největší pravděpodobností samotnou výměnu zpětných odkazů zatím vyhledávače nepenalizují, neboť nemají dostatečně spolehlivý způsob rozpoznání.

Příklad:

Typickým příkladem zpětných odkazů v patičce, které nemají pro uživatele žádnou vypovídající hodnotu, můžeme najít například na webu cestovní kanceláře Net Travel.



Obrázek 12 – Patičkové odkazy na webu <http://www.nettravel.cz>¹².

¹² Dne 2. prosince 2008

3.2.12 Duplikované stránky (mirror sites)

Popis:

Mirror site je název pro stránky, kde se duplikuje obsah nejčastěji přes tag <iframe>, s dodáním relevantních klíčových slov, či už ve smysluplných větách, anebo ve výpisu klíčových slov. Uživatel se potom stránka jeví stejně, ačkoliv robot ji identifikuje jako zcela unikátní

Vliv na hodnocení:

Technika mirror sites je poměrně úspěšná díky výborné možnosti optimalizace a také získání zpětných odkazů na různých doménách se stejným obsahem pro uživatele.

Obtížnost provedení:

Vytvoření mirror sites je v porovnání s ostatními technikami relativně zdoluhavý proces, přesto není použití této praktiky nijak obtížné. Stačí vlastnit pár domén, na kterých se duplikuje obsah stránek skrz tag <iframe>, přičemž se do stránky vloží několik odlišných vět, které vyvolají u robota zdání, že se jedná o diametrálně odlišné weby. Vodítkem pro vyhledávače mohou být doprovodné, nekalé techniky, které jsou s touto spojeny. Jde velmi často např. o malý nebo skrytý text. Důvodem je zobrazování důležitých informací v iframu. Vložený text zde slouží pouze k odlišení jednotlivých mirror sites. Jelikož je těžké psát smysluplný a přínosný text, spameri se často uchylují k neetickým technikám – skrytý text, malý text, opakování klíčových slov atd.

Možná obrana:

Bojovat proti mirror sites se dá jen velmi obtížně. Samozřejmě lze využít spam report, ale z vlastních zkušeností vím, že bezúspěšně. Hlavně když se dané mirror sites vyskytují v českém jazyku.

Riziko penalizace:

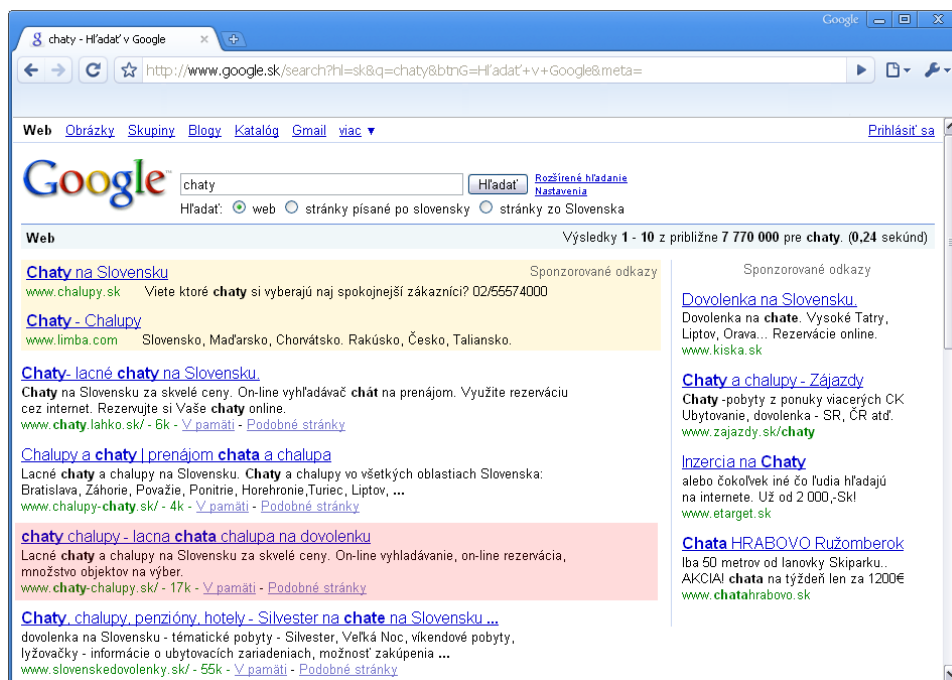
Pro vyhledávače je zatím poměrně obtížné zjistit, zda se jedná o mirror sites. Jelikož je vkládání obsahu do stránek pomocí tagu <iframe> legitimní způsob, nelze tedy rozlišit, jaké povahy je vkládaný obsah a zda se nejedná o duplicitu. Je velmi obtížné rozpoznat neškodný affiliate (který je svým způsobem tak trochu mirror site) od čistokrevných neetických mirror sites. Hlavním kritériem pro odlišení nebezpečnosti těchto webů by měla být přidaná hodnota zrcadlového webu, ale tuto zatím robot nedokáže určit.

Riziko penalizace je tedy velmi nízké.

Příklad:

Jako příklad uvádím slovenskou stránku <http://dovolenky.sk/> a její mirror stránky: <http://yoyam.sk/>, <http://chaty-chalupy.sk/>, <http://www.silvestre.sk/> a další. Přičemž jak hlavní stránka, tak i mirror stránky se umísťují ve výsledcích vyhledávání na vyšších pozicích.

Zde můžete vidět jednu z mirror sites na 3. místě ve výsledcích vyhledávání Google.sk na klíčové slovo „chaty“:



Obrázek 13 - Výsledky vyhľadávání Google.sk na klíčové slovo „chaty“¹³.

Hlavní stránku, ktorou podporujú spam techniky na ostatných webech, najdeme ve zdrojovém kódu každej spamovej stránky. Vkládá se zde pomocí tagu iframe:

```
<iframe src="http://www.dovolenky.sk" width="100%" height="1400" marginwidth="0"
marginheight="0" frameborder="0">
```

¹³ Dne 4. prosince 2008



Obrázek 14 – Mirror 1: www.chaty-chalupy.sk, Mirror 2: www.dovolenky.sk, Mirror 3: www.yoynam.sk ¹⁴.

Ve spodní části můžete vidět další nelegální techniky – takměř neviditelné písmo. Další neetickou metodu, kterou tento web využívá, je název domény yoynam.sk. Majitel této domény získává návštěvnost z překlepů uživatelů, kteří si nevšimnou nastavené anglické klávesnice na svém počítači a pokusí se zadat do adresního řádku URL vyhledávače Zoznam.sk.

Dále se ve spodní části textu nachází přexponované anchor texty, naplněné všemi klíčovými slovy relevantními pro podstránky v rámci webu chalupy-chaty.sk na které vedou:

¹⁴ Dne 4. prosince 2008

Dovolenka a ubytovanie na Slovensku - chaty a chalupy na Slovensku - www.CHATY-CHALUPY.sk

[Bratislava a okolie](#): Chtelnica, Harmónia, Jahodník, Senec, Svätý Jur, Zochova chata | [Záhorie](#): Šaštín, Zlatnícka dolina | [Považie](#): Dubník, Košariská, Modrová, Piešťany, Zelená Voda | [Horné Považie](#): Besné, Čičmany, Dolná Poruba, Horná Tižina, Kľak, Rajecké Teplice, Kamenná Poruba, Kúnerad, Lehota pod Vtáčnikom, Lutiše, Lysá pod Makytou, Rajecká Lesná, Rajecké Teplice - Stránske, Súľov, Terchová, Varín, Veľká Čierna, Vrátna dolina, Veľké Rovné, Žilina | [Ponitrie](#): Bojnica, Čavoj, Duchonka, Nitrianske Rudno, Opatovce nad Nitrou, Podhradie, Ráztočno | [Podunajsko](#): Horné Saliby, Moča, Mužla, Nírad, Patince, Podhájska, Pozba - Podhájska, Radava, Semerovo, Šurany, Veľký Meder | [Pohronie](#): Hodruša, Hrabiny, Kľak, Levice, Malá Lehota, Nová Baňa, Rudno nad Hronom, Vyhne | [Novohrad](#): Sebachľaby | [Podpoľanie](#): Hriňová, Kováčová, Látka | [Horehronie](#): Bacúch, Brezno, Bravčovo, Čierny Balog, Donovaly, Horná Lehota, Jarabá, Jelenec, Kordíky, Krpáčovo, Myto Pod Dúmbierom, Moštenica, Nemecká, Osrblie, Pohorelská Maša, Podbrezová, Šachtičky, Staré Hory, Tajov, Telgárt, Tále, Turecká, Valaská | [Turiec](#): Čremošné, Jasenská dolina, Lipovec, Martin, Martinské hole, Mošovce, Pribovce, Šútovská dolina, Turčiansky Michal, Trusálová, Turčianske Teplice, Valča, Valčianska dolina, Vrútky | [Liptov](#): Bobrovník, Bodice, Bešeňová, Bobrovec, Demänovská Dolina, Dúbrava, Ivachnová, Jasná pod Chopkom, Korytnica, Kľačany, Lazisko, Lúčky, Liptovský Hrádok, Lisková, Liptovský Ján, Liptovská Kokava, Liptovská Ondrášová, Liptovská Teplá, Liptovské Revúce, Liptovská Sielnica, Liptovský Trnovec, Ľubela, Liptovské Vlachy, Martinček, Malinô Brdo, Liptovský Michal, Liptovské Matiašovce, Nižná Boca, Nižné Malatíny, Liptovská Osada, Podtureň, Partizánska Ľupča, Podsúchá, Prosiek, Potok, Potok - Bešeňová, Pavčina Lehota, Ružomberok, Liptovské Sliače, Štiavnička, Svarín, Trstené, Vážec, Vyšná Boca, Veterná Poruba, Žiar | [Orava](#): Oravský Biely Potok, Brezovica, Dolný Kubín, Habovka, Istebná, Kubínska Hôľa, Krivá, Leštiny, Oravská Lesná, Oravská Poruba, Oravice, Oravské Veselé, Podbiel, Oravská Priehrada, Štefanov, Tvrdošín, Veličná, Zázrivá, Zakamenné, Zuberec | [Kysuce](#): Makov, Nižný Keľčov, Oščadnica, Ostré, Raková-Korchán, Skalité, Staškov, Turzovka, Vysoká nad Kysucou, Zákopce | [Vysoké Tatry](#): Gerlachov, Starý Smokovec, Hranovnica, Horný Smokovec, Lopošná Dolina, Mengusovce, Mlynčeky, Nová Lesná, Nový Smokovec, Podbanské, Poprad-Veľká, Stará Lesná, Štrbské Pleso, Štrba, Štôla, Tatranská Lomnica, Tatranské Matliare, Tatranská Štrba, Vernár, Veľká Lomnica, Vrbov, Veľký Slavkov, Ždiar | [Zemagurie](#): Haligovce, Jezersko, Lesnica, Spišské Hanušovce, Stará Ľubovňa, Vyšné Ružbachy | [Spiš](#): Čingov, Dedinky, Harakovce, Hnilec, Košíarny Briežok, Mlynky, Poráč, Smižany, Spišský Štiavnik, Stratená, Spišská Teplica, Závada | [Gemery](#): Dobšinská ľadová Jaskyňa, Dobšiná, Lániho huta, Rejdová, Slavec, Štítnik | [Košice a okolie](#): chaty pri Košiciach | [Šariš](#): chaty Šariš | [Horný Zemplín](#): Domaša | [Dolný Zemplín](#): Zemplínska šírava, Zemplínska šírava - Kamenec, Zemplínska šírava - Medvedia Hora | [Silvester, silvestrovské pobyty](#)

Ubytovacie zariadenia sú ponúkané v spolupráci s WWW.DOVOLENKY.SK
Silvester, silvestrovské pobyty a ubytovanie nájdete na WWW.SILVESTRE.SK

Obrázek 15 – Odkazy s relevantními klíčovými slovy umístěnými v anchor textu

Ostatní mirror stránky uvádět nebudu, vzhledem k jejich velkému počtu a charakteru této práce.

3.3 Shrnutí výsledků analýzy

Jednotlivé techniky, popsané v předešlé kapitole, jsem zhodnotil z několika různých pohledů, které jsou významné pro samotnou realizaci nekalých praktik a optimalizování pro vyhledávače. Každému hledisku jsem přiřadil hodnotu v rozmezí stupnice jedna až pět. Následně sečtu hodnocení jednotlivých pohledů, vydělím počtem metod a porovnáím výsledky mezi sebou. Nekalá praktika s nejnižší hodnotou bude vypovídat o metodě, která má největší vliv na hodnocení výsledků vyhledávání, nejmenší riziko penalizace, nejmenší obtížnost provedení a proti které se bude dát nejhůře bránit.

Pro posouzení jednotlivých metod nekalých SEO faktorů jsem si stanovil následující kritéria, se kterými jsem následně pracoval při vyhodnocování závěrů.

Pohledy	Stupnice				
	1	2	3	4	5
vliv na hodnocení	největší	velký	střední	malý	nejmenší
obtížnost provedení	nejmenší	malá	střední	velká	největší
možná obrana	nejtěžší	těžká	střední	lehká	nejlehčí
riziko penalizace	nejmenší	malé	střední	velké	největší

Tabulka 4 – Stanovené hodnocení nekalých SEO faktorů

Abych byl schopen určit metodu, kterou lze těžko odhalit, snadno zrealizovat, těžce se proti ní bránit a aby měla největší vliv na hodnocení, využil jsem relativních vah mezi ukazateli. To znamená, využil jsem celý rozsah stupnice pro jednotlivé zkoumané pohledy.

Zkoumaným faktorům jsem, na základě údajů z předešlé tabulky, přiřadil následující

hodnocení:

Název nekalé praktiky	Sledované faktory				Celkové hodnocení
	Vliv na hodnocení	Obtížnost provedení	Možná obrana	Riziko penalizace	
Opakující se klíčová slova v textu	1	1	2	5	2,25
Malý text	5	1	2	4	3
Skrytý text	5	1	2	3	2,75
Doorway pages	1	2	1	1	1,25
Cloaking	1	2	3	5	2,75
Automaticky generovaný obsah	3	2	3	5	3,25
Doménový spam	2	3	3	2	2,5
Spamování diskuzí	4	4	2	3	3,25
Klamné přesměrování	5	2	5	2	3,5
Odkazové farmy	1	3	4	4	3
Patičkové odkazy	1	2	3	1	1,75
Duplikované stránky	1	2	1	2	1,5

Tabulka 5 – Zhodnocení jednotlivých nekalých SEO praktik

Celkové hodnocení značí součet hodnot sledovaných pohledů pro jednotlivé nekalé praktiky, vydělený počtem pohledů. Tyto hodnoty značí vhodnost nekalé praktiky pro její zneužití. Hodnoty blíže jedné označují metody nejvhodnější a naopak hodnoty blíže pěťce, metody nejméně vhodné.

Jak je patrné z výsledků celkového hodnocení, mezi nejúčinnější nekalé praktiky optimalizace patří metody:

- 1) Doorway pages
- 2) Duplikované stránky (Mirror sites)
- 3) Patičkové odkazy

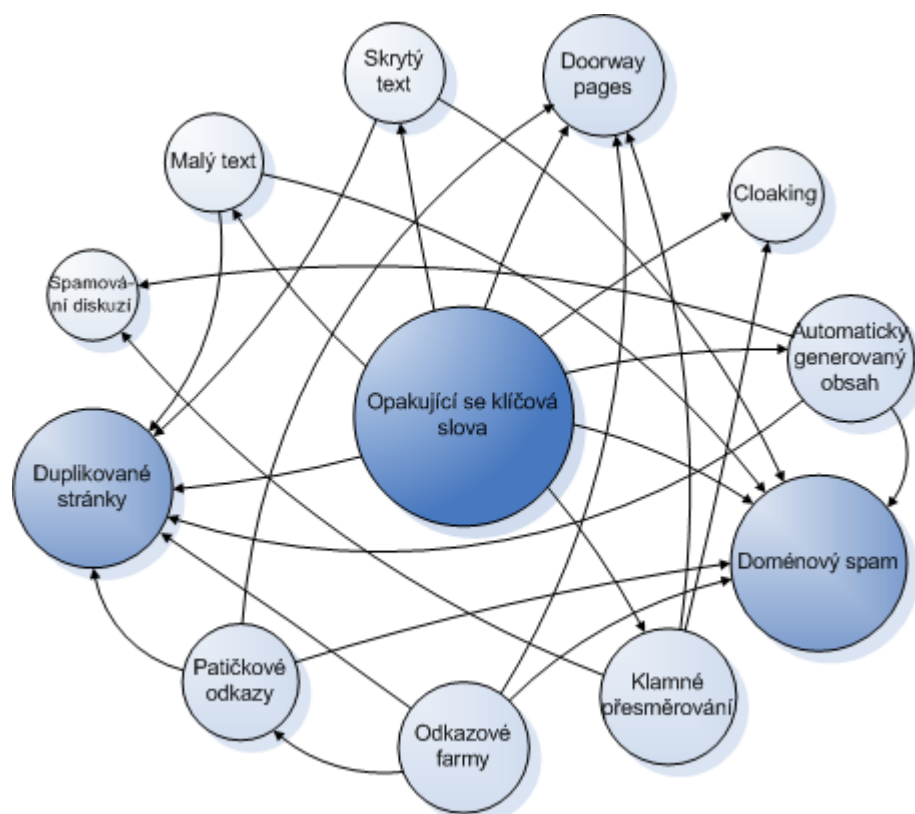
Naopak mezi nejméně účelné nekalé SEO praktiky patří:

- 1) Klamné přesměrování
- 2) Automaticky generovaný obsah
- 3) Spamování diskuzí
- 4) Odkazové farmy
- 5) Malý text

Výše uvedená tabulka počítá pouze s hodnocením samotných nekalých praktik. Neuvažuje již kombinaci více praktik dohromady. Proto například spojení dvou či více metod dohromady může vytvořit ještě silnější nekalou techniku.

Na následujícím obrázku (Obrázek 16) lze vidět, jaké praktiky se nejčastěji vyskytují společně, k dosažení ještě většího efektu. Barvy a velikosti bublin poukazují na praktiky, které lze nejčastěji zkombinovat s ostatními. Jak z obrázku vyplývá, přemíra opakujících se klíčových slov je jednou

z nejpoužívanějších metod.



Obrázek 16 – Vztahy mezi jednotlivými nekalými praktikami

4 Závěr

Na samém začátku jsem popsal, proč se vůbec optimalizací webových stránek zabývat a jaká jsou její úskalí. Uvedl jsem nejzákladnější faktory, které se dotýkají oboru optimalizace pro vyhledávače, abych seznámil čtenáře s pojmy, které budu v průběhu práce používat. Jelikož by bylo popsání všech SEO faktorů nad rámec této práce, odkázal jsem čtenáře na literaturu, která se jimi zabývá, aby si mohli dostudovat všechny potřebné znalosti k pochopení bližších souvislostí této práce a já se mohl detailněji zaměřit právě na nekalé praktiky.

Při analyzování jednotlivých nekalých praktik se mi podařilo popsat principy a způsoby jaké se využívají k oklamání vyhledávačů za účelem dosažení vyšších pozic ve výsledcích vyhledávání. Správci webových stránek zde naleznou teoretické i praktické příklady zabezpečení proti jednotlivým nekalým praktikám. Jsou upozorněni, čeho se mají vyvarovat při tvorbě a provozování webových stránek, aby se nestali obětí spamu, či se sami nestali nekalým konkurentem.

Ze samotné analýzy také vyplývá, že existuje velmi slabá hranice mezi vhodnou optimalizací, využívající přiměřený poměr klíčových slov ke zbytku textu, a mezi přemírou optimalizace, která již uživateli nedává příjemný pocit ze čtivých stránek. Není snadné určit, kam až zajít při konkurenčním boji o přední pozice ve výsledcích vyhledávání. Toto rozhodnutí záleží na samotném správci webových stránek.

Jako hlavní cíl této práce jsem si stanovil zhodnocení nekalých praktik, abych mohl blíže proniknout do problému a vyvodit z něj určité závěry. Samotným analyzováním jsem dospěl k závěru, že existuje mnoho technik, které vyhledávače nejsou schopny rozpoznat a tím dávají velký prostor nekalým konkurenčním praktikám, kterých bude dle mého názoru stále více přibývat, což bude mít za následek snižování relevantnosti výsledků vyhledávání.

Aby vyhledávače mohly této skutečnosti zabránit, hledají stále nové způsoby, jak lépe vyhodnocovat kvalitu webových stránek. Kromě současných snah často pozměňovat, upravovat a zdokonalovat vyhodnocovací algoritmy, přišel Google také s nápadem nechat vyhodnotit relevanci každé stránky samotné uživatele [21]. Otázkou zůstává, do jaké míry získá zpětnou vazbu a bude schopen na základě získaných údajů vyhodnotit míry relevance. Je to ale jedna z možností, kam se bude tato problematika dále vyvíjet.

Na příkladech doménového spamu, malého textu, skrytého textu, opakování klíčových slov a patičkových odkazů se mi podařilo dokázat, že vyhledávač Google spam buď nerozpozná, ignoruje, nebo mu velmi dlouho trvá, než spamové stránky penalizuje. To dává prostor nekalým konkurentům obsazovat přední pozice ve výsledcích vyhledávání a vytvářet nerelevantní výsledky pro konečného uživatele. Nekalé praktiky nejsou využívány samostatně, ale pro zvýšení efektu a snížení rizika penalizace jsou různě spojovány, čímž vytváří hůře odhalitelné spamové stránky. Celková úspěšnost vyhledávání se tak ještě více snižuje.

Český vyhledávač Seznam, na rozdíl od Google, reaguje celkem rychle na spam reporty a nahlášené spamové praktiky penalizuje. Také tím zajišťuje relevantní výsledky na jednotlivá klíčová slova a úspěšnost celkového hledání. Může to být také jeden z faktorů, proč je zatím stále nejpoužívanějším vyhledávačem na českém trhu.

Doporučení, jakým způsobem lze rozpoznat nekalé praktiky, jsem uvedl při jejich bližším analyzování. Správcům webových stránek jsem popsal, jakým způsobem si mohou svůj web zabezpečit a jak se chovat, aby sami sebe nepostavili do pozice penalizovaného webu. Na druhou stranu jsem vyhledávačům načrtnul možnosti, jak jednotlivé praktiky odhalovat.

Velmi doufám, že jsem ve své práci dokázal popsat nekalé SEO praktiky takovým způsobem, aby je byli schopni odhalit i uživatelé, kteří nejsou odborníky na tuto problematiku. Zároveň doufám, že jsem je přesvědčil o nevhodnosti využívání těchto praktik, neboť vede

k nerelevantním výsledkům ve vyhledávání, což ztěžuje hledání nám všem. Už jen pro naše vlastní zájmy bychom měli proti nekalým praktikám všichni usilovně bojovat.

Seznam použité literatury a zdrojů

- [1] SMÍČKA, R.: Optimalizace pro vyhledávače – SEO. Jasminka.cz [online]. 2004. Dostupný na WWW <<http://seo.jasminka.cz/seo-kniha.pdf>> [21.11.2008].
- [2] INTERNET INFO: Navrcholu.cz: Spojení Atlasu s Centrem pozici Seznamu ve vyhledávání neohrozilo. Internet info [online]. Dostupný na WWW <<http://www.iinfo.cz/tiskove-centrum/tiskove-zpravy/navrcholu-cerven-vyhledavace/>> [7.12.2008].
- [3] SEARCH ENGINE LAND: Eye Tracking On Universal And Personalized Search Third Door Media [online]. Dostupný na WWW: <<http://searchengineland.com/eye-tracking-on-universal-and-personalized-search-12233.php>> [22.11.2008].
- [4] STANDFORD.EDU: The Anatomy of a Large-Scale Hypertextual Web Search Engine, The Stanford University InfoLab [online]. Dostupný na WWW: <<http://infolab.stanford.edu/~backrub/google.html>> [22.11.2008].
- [5] EFACTORY.DE: The Implementation of PageRank in the Google Search Engine, eFactory Internet-Agentur [online]. Dostupný na WWW: <<http://pr.efactory.de/e-pagerank-implementation.shtml>> [23.11.2008].
- [6] JAK PSÁT WEB: Podle čeho řadí vyhledávání na Centrum.cz, Yuhůův weblog o webu [online]. Dostupný na WWW: <<http://weblog.jakpsatweb.cz/d/1224512100-podle-ceho-radi-vyhledavani-na-centrumcz.html>> [23.11.2008].
- [7] SEZNAM.CZ, a.s.: Algoritmus, Seznam nápověda [online]. Dostupný na WWW: <<http://napoveda.seznam.cz/cz/hledani-fulltext-algoritmus-vyhledavani-razeni-vysledku-faq-dotazy.html>> [23.11.2008].
- [8] THE STANFORD NATURAL LANGUAGE PROCESSING GROUP: Hubs and Authorities, [online]. Dostupný na WWW: <<http://nlp.stanford.edu/IR-book/html/htmledition/hubs-and-authorities-1.html>> [25.11.2008].
- [9] MARKETINGOVÉ NOVINY: Co dokáže optimalizace pro vyhledávače, Marketingové noviny – váš průvodce marketingem [online]. Dostupný na WWW: <http://www.marketingovenoviny.cz/index.php3?Action=View&ARTICLE_ID=250> [23.11.2008].
- [10] SEZNAM.CZ, a.s.: Komunikace s vyhledávači – robots.txt, Seznam nápověda [online]. Dostupný na WWW: <<http://napoveda.seznam.cz/cz/hledani-fulltext-komunikace-s-vyhledavaci-robots-txt.html>> [22.11.2008].
- [11] GOOGLE: Verifying Googlebot. Google webmaster help center [online]. Dostupný na WWW: <<http://www.google.com/support/webmasters/bin/answer.py?answer=80553>> [22.11.2008].
- [12] SEZNAM.CZ, a.s.: Jak lépe optimalizovat, Seznam nápověda [online]. Dostupný na WWW: <<http://napoveda.seznam.cz/cz/hledani-fulltext-jak-lepe-optimalizovat-stranky-seo.html>> [22.11.2008].
- [13] GOOGLE: Search engine optimization starter guide, Google webmaster help center [online]. Dostupný na WWW: <<http://www.google.com/webmasters/docs/search-engine-optimization-starter-guide.pdf>> [22.11.2008].
- [14] GOOGLE: Search Friendly Web Development, Google developer day 2008 [online]. Dostupný na WWW: <<http://sites.google.com/site/developerday2008prague/google-developer-day-prague-2008/search-friendly-web-development>> [22.11.2008].
- [15] SEZNAM.CZ, a.s.: Nahlášení spamu, Search seznam [online]. Dostupný na WWW: <<http://search.seznam.cz/nahlasit-spam>> [22.11.2008].
- [16] GOOGLE: Oznámit spam v našem indexu, Nástroj pro webmastery [online]. Dostupný

- na WWW: <<https://www.google.com/webmasters/tools/spamreport?hl=cs>> [22.11.2008].
- [17] WEBSITE CONTENT WIZARD.COM: David Watson's Website Content Wizard [online]. Dostupný na WWW: <<http://www.websitecontentwizard.com/wbctwz/14SecondReport.php>> [7.12.2008].
- [18] MOZZILA.ORG: User Agent Switcher 0.6.11, Firefox Add-ons [online]. Dostupný na WWW: <<https://addons.mozilla.org/en-US/firefox/addon/59>> [7.12.2008].
- [19] JAK PSÁT WEB: Ebola.cz – zkušenosti II, diskuze jak psát web [online]. Dostupný na WWW: <<http://diskuse.jakpsatweb.cz/index.php?action=vthread&forum=6&topic=71357>> [22.11.2008].
- [20] GOOGLE.COM: Požádat o nové posouzení, Nástroje pro správce webu, [online]. Dostupné po přihlášení na WWW: <<https://www.google.com/webmasters/tools/reconsideration?hl=cs>> [7.12.2008].
- [21] BBC.CO.UK: Google unveils customised search, BBC News, [online]. Dostupné na WWW: <<http://news.bbc.co.uk/2/hi/technology/7740815.stm>> [8.12.2008].

Seznam použitých zkratk a termínů

Bad neighbour	Výraz pro weby s nelegální či nevhodnou tematikou, na které se nedoporučuje odkazovat.
Black list	Seznam zakázaných webových stránek trvale či dočasně vyřazených z indexu.
Captcha	Je způsob zabezpečení formulářových polí webových stránek, proti jejich zneužití automatickými systémy (roboty). Nejčastěji je ve formě obrázku, na kterém je umístěný text. Pro odeslání formuláře je potřeba opsat text z obrázku do formulářového pole a odeslat k vyhodnocení. Pokud se bude shodovat text obrázku s opsaným textem ve formulářovém poli, dojde k úspěšnému odeslání formuláře. Tato metoda se hojně využívala z důvodu vysoké spolehlivosti odolávat vůči robotům. V současné době již tak úspěšná není a nedoporučuje se spoléhat pouze na ni.
CSS	„Cascading Style Sheets“, neboli „kaskádové styly“ je označení pro technologii zpracování grafického designu webových stránek.
Crawler	Robot, který na základě odkazů umístěných na stránkách prochází internetem a indexuje nové webové stránky.
IP adresa	„Internet Protocol“ adresa je číselná identifikace počítače uvnitř počítačové sítě. Příklad IP adresy: 192.168.0.1
KeyWords	Česky „klíčové slovo“ je pojem používaný pro frázi, která je typická pro tematiku webu. Uživatelé budou tuto frázi zadávat do vyhledávače za účelem zjištění relevantních výsledků.
Konverzní poměr	Např. poměr nákupů v eshopu na počet návštěvníků, nebo počet zaregistrovaných uživatelů na počet návštěvníků.
Long Tail	Neboli česky „dlouhý ocas“ je pojem označující návštěvnost webových stránek na velmi specifická, tří a víceslovné fráze, přes které uživatelé naleznou web ve výsledcích vyhledávání.
LSI	„Latent Semantic Indexing“ je metoda indexování webových stránek, která podle hustoty klíčových slov na stránce zjistí o čem stránka je a dokáže jednotlivá klíčová slova přiřadit do tematických skupin. Na tomto základě posléze umí zjistit, který další dokument se zabývá stejným tématem.
PPC	„Pay Per Click“ je reklamní systém, který zobrazuje textové, grafické, video a jiné reklamy, ať již v obsahové síti, nebo na stránce výsledků vyhledávání. Mezi nejznámější pro české prostředí patří Sklik a AdWords.
PR	„Public Relation“ je skupina metod a technik pro vztahy s veřejností, kterými si firma zajišťuje popularitu a „dobré jméno“ společnosti.
Rank	Vyhledávače hodnotí kvalitu jednotlivých stránek takzvaným rankem. Např. Seznam nazývá své hodnocení SRank. Google nazývá své

hodnocení PageRank.

Robot	Je, ve smyslu k metodě SEO, automatický systém procházející internetem, který na základě svého algoritmu, zaznamenává jednotlivé faktory webových stránek. Ze zjištěných faktorů se posléze vyhodnocuje kvalita samotných stránek.
RSS	„Really Simple Syndication“ je označení pro informační kanál, na kterém weby mohou informovat o svých novinkách. Uživatelé si mohou tyto informace zobrazit například v tzv. RSS čtečce a dozvědět se tak prakticky ihned o různých novinkách, aniž by museli navštívit samotný web.
SEM	„Search Engine Marketing“ je označení pro marketing ve vyhledávacích a formu, jak zvýšit viditelnost webových stránek ve výsledcích vyhledávání.
SEO	„Search Engine Optimization“ neboli optimalizace pro vyhledávače, je pojem používaný pro činnosti zabývající se úpravou webových stránek za účelem dosahování vyšších pozic ve výsledcích vyhledávání.
SERP	„Search Engine Result Page“ je pojem používaný pro výsledky vyhledávání.
Spam	Spam je označení pro nevyžádaný obsah.
Spam report	Webový formulář, jehož vyplněním a zasláním administrátorům vyhledávače lze nahlásit nekalé praktiky konkrétní stránky.
WYSIWYG	„What You See Is What You Get“ je metoda zápisu obsahu webových stránek, která umožňuje zobrazit výsledek zápisu již ve stádiu jeho psaní.
XSS	„Cross Site Scripting“ je označení pro bezpečnostní díru webových stránek, která útočníkovi umožňuje vložit jeho vlastní skript do stránek „oběti“ a možnost získat nad jeho webem kontrolu.