Univerzita Pardubice
Fakulta ekonomicko-správní

Moderní metody vyhledávání v internetu

Michal Caha

Bakalářská práce

2008

Univerzita Pardubice
Fakulta ekonomicko-správní
Ústav systémového inženýrství a informatiky
Akademický rok: 2007/2008

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Michal CAHA**

Studijní program: **B6209 Systémové inženýrství a informatika**

Studijní obor: **Informatika ve veřejné správě**

Název tématu: **Moderní metody vyhledávání v internetu**

Zásady pro vypracování:

1. Rozdělení a popis internetových vyhledávačů
2. Metody vyhledávání internetových vyhledávačů
3. Popis metody SEO (Search engine optimization)
4. Vliv optimalizace na vyhledávání

Rozsah grafických prací:

Rozsah pracovní zprávy:

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

Seznam odborné literatury:

[1] SMIČKA, Radim. Optimalizace pro vyhledávače - SEO, 2004. 120 s. ISBN 80-239-2961-5

[2]WALL, Aaron Mathew, Search Engine Optimization Book, 2005. 220s

[3]Interval.cz [online]. 2004 [cit. 2007-10-23]. Dostupný z WWW: <http://interval.cz/serialy/seo-search-engine-optimization/>. ISSN 1212-865.

[4]LUPA [online]. 2007 [cit. 2007-10-23]. Dostupný z WWW: <http://www.lupa.cz/n/seo/>. ISSN 1213-070.

[5]SEO Optimalizace pro vyhledávače [online]. 2004 [cit. 2007-10-23]. Dostupný z WWW: <http://tvorba-webu.zdarek.com/seo/>.

[6]Jak psát web [online]. 2007 , 12. října 2007 [cit. 2007-10-23]. Dostupný z WWW: <http://www.jakpsatweb.cz/google.html>.

[7]SEOmoz.org [online]. 2007 , 12. října 2007 [cit. 2007-10-23]. Dostupný z WWW: <http://www.seomoz.org/article/beginners-guide-to-search-engine-optimization>.

Vedoucí bakalářské práce: **Ing. Milan Tomeš**
Ústav systémového inženýrství a informatiky

Datum zadání bakalářské práce: **25. října 2007**

Termín odevzdání bakalářské práce: **19. května 2008**

prof. Ing. Jan Čapek, CSc.
děkan

L.S.

doc. Ing. Pavel Petr, Ph.D.
vedoucí ústavu

V Pardubicích dne 25. října 2007

SOUHRN

Tato práce se zabývá problematikou spojenou s metodami vyhledávání na internetu a metodami optimalizace webových stránek z hlediska vyhledávání. Práce je rozdělena do čtyř kapitol. V nich jsou rozebrány druhy internetových vyhledávačů a jejich metody vyhledávání, popsána metoda SEO a vyhodnocen vliv optimalizace na konkrétních internetových stránkách.

KLÍČOVÁ SLOVA

internet, internetové vyhledávače, metody vyhledávání, Google, optimalizace webových stránek

TITLE

Modern methods of web searching

ABSTRACT

This bachelor thesis discusses problems joined with methods of web searching and methods of web site optimization in light of web searching. The thesis is divided into four main chapters. Types of web search engines and their methods of web searching are analysed, SEO-method is described and the influence of optimization on particular web site is examined.

KEYWORDS

internet, web search engines, methods of web searching, Google, web site optimization

OBSAH

1 Úvod 5
2 Rozdělení a popis internetových vyhledávačů 6
2.1 Katalogové vyhledávače 6
2.2 Fulltextové vyhledávače 7
2.2.1 Fultextové vyhledávače ve světě 8
2.2.2 Fultextové vyhledávače v Česku 8
3 Metody vyhledávání internetových vyhledávačů 9
3.1.1 Centralizovaná architektura 9
3.1.2 Metavyhledávání 10
3.1.3 Distribuované vyhledávání 11
3.2 Problémy webových vyhledávacích strojů 12
3.3 Sémantický web 13
3.3.1 Hlavní prvky sémantického webu 13
4 Metoda SEO 14
4.1 On-page Faktory 14
4.1.1 URL Adresa 14
4.1.2 Titulek 15
4.1.3 Metainformace 15
4.1.4 Vlastní obsah stránky 17
4.1.5 Rámce 19
4.2 Off-page faktory 20
4.2.1 PageRank 20
4.2.2 Získávání zpětných odkazů 21
5 Zakázané metody optimalizace 22
5.1 Skrytý text nebo odkazy 22
5.2 Malý text 22
5.3 Cloaking 22
5.4 Odkazové farmy 23
5.5 Duplicitní obsah na více URL 23
5.6 Komentářový spam 23
5.7 Klamná či opakující se slova 23
6 Vliv optimalizace na vyhledávání 23

6.1 Představení testovaných stránek 24
6.2 Optimalizace on-page faktorů 24
6.2.1 Optimalizace URL adres 25
6.2.2 Změny doménové adresy 25
6.2.3 Odstranění rámců 25
6.2.4 Optimalizace titulků 26
6.2.5 Optimalizace metainformací 27
6.2.6 Optimalizace nadpisů 28
6.2.7 Optimalizace obrázků 28
6.3 Optimalizace off-page faktorů 29
6.4 Představení testovacích kritérií 31
6.5 Výsledky vyhledávání neoptimalizovaných stránek 31
6.6 Výsledky vyhledávání stránek s optimalizovanými on-page faktory 32
6.7 Výsledky vyhledávání stránek s optimalizovanými off-page faktory 33
6.8 Výsledky vyhledávání plně optimalizovaných stránek 33
6.9 Zhodnocení vlivu optimalizace 35
7 Závěr 36

1 Úvod

V dnešní době se vyhledávání informací, produktů a služeb prostřednictvím internetu stává čím dál běžnějším a rozšířenějším jevem napříč celou populací. Je logické, že čím větší je význam Internetu z hlediska vyhledávání pro jeho uživatele, tím větší význam má tedy i pro ty, kteří tyto informace či produkty chtějí co nejširší veřejnosti poskytnout. Každým okamžikem tak na Internetu vznikají nové webové stránky, v jejichž rapidně narůstajícím množství je stále těžší něco nejen nalézt (z pohledu příjemce informací), ale také být nalezen (z pohledu poskytovatele informací). Jedním ze základních prostředků tvůrce internetových stránek, kterým lze zvýšit šanci na upoutání pozornosti internetového uživatele právě jeho webovým obsahem, by měl být Search Engine Optimization. Search Engine Optimization (SEO) je způsob úpravy (optimalizace) obsahu a zdrojového kódu jednotlivé stránky tak, aby se objevila na předních místech ve fulltextových vyhledávačích, které jsou v dnešní době klíčovým hráčem ve vyhledávání informací na Internetu. Čím lepší umístění taková stránka ve výsledcích zaujímá, tím víc je na očích a tím větší má šanci, že bude uživatelem, v určitých případech i potenciálním zákazníkem, navštívena. Proto se v dnešní době do SEO a SEM (Search Engine Marketing) investují nemalé finanční prostředky.

Cílem této bakalářské práce je využít teoretických poznatků o způsobech vyhledávání internetových vyhledávačů a metodách optimalizace webových stránek k ověření, zda a do jaké míry ovlivňuje tato optimalizace jejich postavení ve vyhledávačích.

Bakalářská práce je rozdělena do 4 kapitol. První část je zaměřena na rozdělení a popsání základních vyhledávacích nástrojů na internetu – vyhledávačů katalogových a fulltextových.

Druhá kapitola popisuje metody vyhledávání internetových vyhledávačů, s jakými problémy se vyhledávače setkávají a v závěru nastiňuje koncepci sémantického webu, který by vedl k zefektivnění vyhledávání na internetu.

Třetí, a zároveň stěžejní kapitola se věnuje metodě SEO a je rozčleněna do tří hlavních částí, v nichž jsou popsány všechny způsoby optimalizace webových stránek. Tedy optimalizace on-page faktorů, on-page faktorů, ale i zakázané (penalizované) způsoby optimalizace.

V poslední, čtvrté kapitole se bakalářská práce věnuje posouzení vlivu optimalizace na výsledky vyhledávání stránek vyhledávači. Toto posouzení spočívá ve vytvoření

neoptimalizovaných a optimalizovaných stránek a následném porovnání výsledků jejich vyhledávání.

2 Rozdělení a popis internetových vyhledávačů

Současný stav vyhledávacích nástrojů je ovlivněn jejich historickým vývojem. Oba základní typy současných vyhledávačů (katalogové vyhledávače a fulltextové vyhledávače) vycházejí z původních potřeb získávání informací z Internetu. První způsob byl vytváření seznamů dokumentů podle jejich tematiky na základě ručního sběru dat. Úplně na začátku byly často záložky (angl. bookmarks) v internetových prohlížečích. Tento způsob shromažďování informací se stal s obrovským nárůstem informací dále neudržitelný. To vedlo ke vzniku a k rozšíření služeb využívajících pro sběr a vytváření databází s využitím automatizovaných prostředků. V současnosti jsou právě tyto služby nejúčinnějšími vyhledávacími nástroji, neboť využívají vysoce výkonné technologie pro fulltextové indexování a vyhledávání informací.[3]

2.1 Katalogové vyhledávače

Katalogový vyhledávač (katalog) je web, který obsahuje velké množství odkazů na jiné webové stránky. Tyto odkazy jsou řazeny do tematických oblastí, podkategorií a to podle zaměření (např. doprava, kultura, sport, restaurace atd.). V katalogu by měl mít web v jedné kategorii pouze jeden záznam. Pokud web obsahuje podstránky, jejichž obsah je tematicky odlišný od zaregistrované stránky v jiné kategorii, je možné tyto podstránky do jiné kategorie přidat. Přidání odkazu do katalogu probíhá tak, že se nejprve stránka v příslušném katalogu zaregistruje, poté je zkontrolována lidským faktorem - editorem a následně je do katalogu záznam přidán.

V katalozích je možné vyhledávat dvěma způsoby. Buď listováním jednotlivými sekcemi, nebo zadáním dotazu do vyhledávacího pole. Při vyhledávání pomocí dotazu dochází k porovnávání zadaného dotazu s URL, nadpisem a popiskem stránky, které byly v katalogu zaregistrovány. Má-li být zajištěna co nejlepší pozice stránky při vyhledávání v katalogu, měl by být kladen největší důraz na volbu klíčových slov v nadpisu stránky, které nejvěrněji popisují obsah stránek. Důležité je také vyhnout se slovům jako „nejlepší" a „nejrychlejší", protože nepřinášejí žádnou užitečnou informaci a jsou editory odstraňovány. Klíčová slova je třeba, jak v popisku, tak i nadpisu, psát v takovém tvaru, v jakém je nejpravděpodobněji uživatel zadá, protože naprostá většina katalogových

vyhledávačů neumí vyhledávat podobná slova, či skloňovat. Popisek je také třeba psát jako smysluplnou větu a ne jako prostý výčet klíčových slov, protože jinak ji editor do katalogu nezařadí.[3],[13]

Mezi nejrozšířenější české katalogy patří dle [17] například:

- Seznam (http://www.seznam.cz)
- Atlas (http://www.atlas.cz)
- Centrum (http://www.centrum.cz)
- O2 Active (http://vyhledavani.o2active.cz/)

Mezi nejpoužívanější zahraniční katalogové vyhledávače patří:

- DMOZ (http://dmoz.org)
- Yahoo (http://www.yahoo.com)
- Excite (http://www.excite.com)
- Google (http://directory.google.com)

Význam katalogových vyhledávačů v současnosti vzhledem k obrovskému počtu nově vznikajících stránek klesá a zajímavější jsou z pohledu vyhledávání na internetu vyhledávače fulltextové.

2.2 Fulltextové vyhledávače

Fulltextový vyhledávač je takový program, který vyhledává a stahuje dokumenty (webové stránky, textové dokumenty, PDF soubory, aj.), ty zpracovává a ukládá do své databáze. Tomuto postupu se také říká indexování. V této databázi potom na základě uživatelských, více či méně složitých, dotazů vyhledává a vrací uživateli výsledky. Hledání tedy neprobíhá v reálném čase na internetu, ale v databázi na serveru vyhledávače. Fulltextový vyhledávač se skládá ze dvou částí. Robot (nebo-li také spider-pavouk, či crawler-lezec) prochází zdrojový kód stránek, který indexuje, a pomocí odkazů v něm nalezených se přesouvá na další stránky. Druhou částí je webové rozhraní, se kterým pracuje samotný uživatel.

Do fulltextových vyhledávačů je možné registrovat stránky i manuálně, ale tento způsob postrádá smysl, protože by je robot našel sám. Navíc ty stránky, které si najde robot sám, se umisťují na lepších pozicích díky množství odkazů vedoucích na ně. Ruční registrace má smysl pouze u vyhledávačů katalogových.[3],[16]

2.2.1 Fultextové vyhledávače ve světě

V současné době drží největší podíl na trhu Google. Hovoří se o tom, že Google nemůže kralovat věčně, ale prozatím není vyhledávač, který by byl populárnější. Dohnat Google se snaží hlavně Yahoo!, které koupilo celou řadu vyhledávačů (Altavista, Inktomi, Alltheweb), aby z nich vyladilo jednu technologii, kterou začalo používat na svých stránkách na začátku roku 2004.[6]

Podíly vyhledávačů na vyhledávání v prosinci roku 2007 zobrazuje následující tabulka.

Tab. 1 Podíly vyhledávačů ve světě; zdroj: comScore [14]

Vyhledávač	počet vyhledávání (v milionech)	podíl (v %)
Google Sites	41,345	62,4
Yahoo! Sites	8,505	12,8
Baidu.com Inc.	3,428	5,2
Microsoft Sites	1,940	2,9
NHN Corporation	1,572	2,4
eBay	1,428	2,2
Time Warner Network	1,062	1,6
Ask Network	728	1,1
Yandex	566	0,9
Alibaba.com Corporation	531	0,8

2.2.2 Fultextové vyhledávače v Česku

V Česku se pro vyhledávání v současnosti nejvíce používá Seznam. Pro vyhledávání v českých stránkách Seznam používá vlastní technologii Seznam Fulltext a pro vyhledávání v zahraničních stránkách technologii Googlu. Druhý nejpoužívanější vyhledávač u nás je zahraniční Google, který však neustále posiluje a ubírá podíly na vyhledávání jak Seznamu, tak ostatním vyhledávačům, jejichž podíly jsou zanedbatelné (pod 5%).[16]

Tab. 2 Podíly vyhledávačů u nás; zdroj: Lupa [16]

Vyhledávač	Červenec 2006	Říjen 2006	Leden 2007	Říjen 2007	Absolutní rozdíl*	Relativní rozdíl**
Seznam	61,10%	63,39%	62,53%	62,30%	-0,23	-0,4
Google	23,48%	23,85%	24,75%	28,87%	4,12	16,6
Centrum	5,37%	4,72%	4,84%	3,39%	-1,45	-30
Atlas	2,86%	2,38%	2,58%	1,53%	-1,05	-40,7
Jyxo	0,79%	0,57%	0,42%	0,24%	-0,18	-42,8

* rozdíl mezi říjnem a lednem 2007, v %

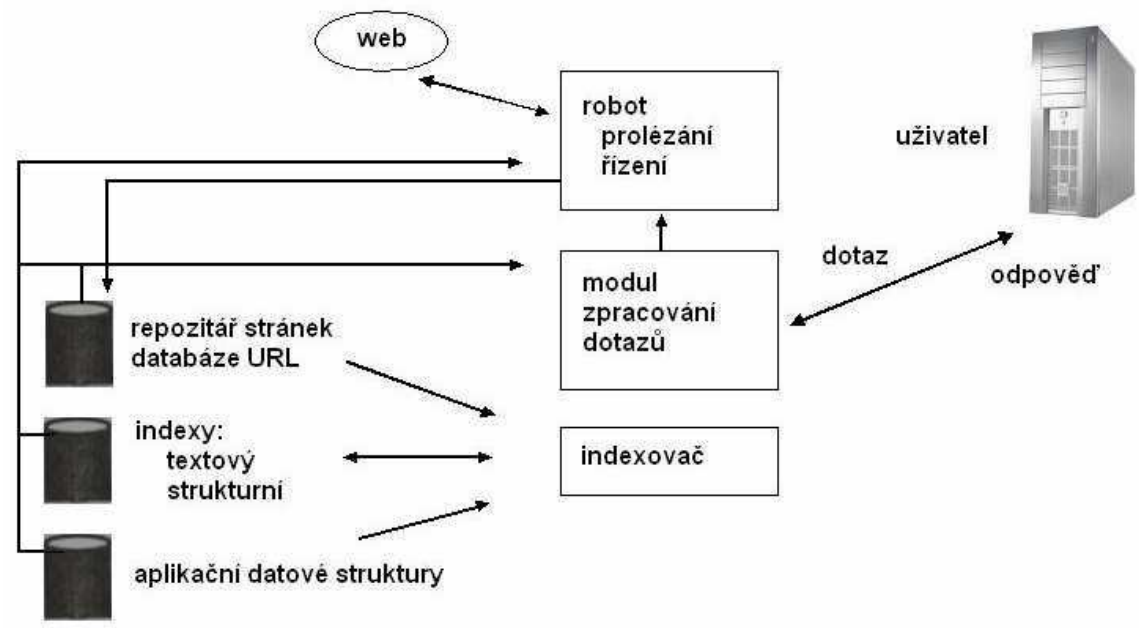
** rozdíl mezi říjnem a lednem 2007, v %

3 Metody vyhledávání internetových vyhledávačů

Rozlišujeme tři základní metody vyhledávání na webu založených na architektuře vyhledávacích strojů. Těmito metodami jsou: tradiční centralizované vyhledávání, metavyhledávání a distribuované vyhledávání. Vyhledávací stroj může také být součástí obecnějších architektur, jako jsou webové vyhledávací služby, nebo ho lze využít ve speciálních aplikacích, jako jsou např. webové portály. Webový portál je chápán jako místo, ve kterém jsou organizovány informace podle témat. Na rozdíl od univerzálních vyhledávacích strojů, které indexují informace na webu, portály jsou udržované člověkem s cílem usnadnit navigaci listováním strukturou.[10]

3.1.1 Centralizovaná architektura

Cílem univerzálních vyhledávacích strojů je indexovat dostatečně velkou část webu, nezávisle na jejím tématickém zaměření či aplikační doméně. Každý takový stroj se skládá z komponent zobrazených na obrázku 1.



Obrázek 1: Centralizovaná architektura; zdroj: [10]

Robot je program, řízený modulem pro řízení robota, který "prolézá" web a sbírá dokumenty rekurzivním pohybem přes odkazy po spojích ze startovacích stránek. Vybrané stránky nebo jejich části jsou komprimovány a ukládány do repozitáře stránek. URL a jejich vazby jsou přenášeny do modulu pro řízení robota, který rozhoduje o pohybu po grafu webu. Každý robot má odlišnou strategii, po kterých spojích a jak hluboko se v grafu webu bude pohybovat. Problémem tohoto robotu je, že nikdy nemůže z množiny startovacích stránek nalézt všechny webové stránky, protože graf webu netvoří jednu souvislou grafovou komponentu. Dalším důvodem je, že nové stránky jsou přidávány s větší rychlostí, než je rychlost robota. Navíc mohou být stránky aktualizovány dříve, než k nim robot dorazí.

Indexovač zpracovává stránky posbírané robotem. Nejprve rozhoduje, které stránky má indexovat. Např. může vynechávat duplikáty obsahu stránek.

Modul zpracování dotazů zpracovává uživatelské dotazy a vrací odpovědi na shodu v pořadí určeném nějakým řadícím algoritmem např. podle klesající hodnoty ohodnocení odpovědi.[10]

3.1.2 Metavyhledávání

Jeden ze způsobů, jak poskytnout přístup k informacím v textových databázích neviditelného webu, je realizován prostřednictvím metavyhledávačů, které mohou být použity pro dotazování nad více databázemi najednou. Metavyhledávač provádí tři hlavní

úkoly. Po obdržení dotazu nalézá pro jeho vyhodnocení ty nejlepší databáze (proces výběr databáze), pro každou databázi překládá dotaz do vhodné formy (proces překlad dotazu), vybírá a slučuje výsledky z různých databází (proces vytváření výsledků) a poté konečný výsledek vrací uživateli. Komponenta výběru databáze je u metavyhledávačů velmi důležitá z hlediska pružnosti a efektivnosti zpracování dotazu.

Pro vyhledávání webu se u metavyhledávačů používají tři metody:

- Vyhledávání s daným seznamem vyhledávacích strojů
- Sekvenční vyhledávání
- Souběžné vyhledávání

V prvním případě řadí metavyhledávač získané výsledky podle určitých parametrů, jako je popularita vyhledávacího stroje, termy dotazu apod. Uživatel nemusí vědět, o jaké vyhledávací stroje jde. V druhém případě uživatel vybírá sám některé zdroje z nabídnutého seznamu a obdrží výsledky najednou v pořadí tak, jak je vracely jednotlivé stroje. Nevýhodou tohoto postupu je, že je závislý na nejpomalejším stroji z použitého seznamu. V posledním případě se výsledky okamžitě zobrazují uživateli, což spoří čas uživatele strávený při provádění dotazu.

Problematická místa metavyhledávačů se týkají překládání dotazů, při kterém se díky různosti jednotlivých vyhledávačů může ztrácet původní záměr uživatele, dále pak vytváření konečného výsledku, tj, problému, jak dostat dopředu ty nejlepší dokumenty a neztrácet čas na zpracovávání těch nerelevantních.[9],[10]

Příklady metavyhledávačů ve světě dle[18]:

- Search.com (http://www.search.com)
- Apollo 7 (http://www.apollo7.de)
- Meta Crawler (http://www.metacrawler.com)
- Monster Crawler (http://www.monstercrawler.com)
- Search Caddy (http://www.searchcaddy.com)

3.1.3 Distribuované vyhledávání

Jakkoliv jsou řadící algoritmy pro centralizované vyhledávací stroje úspěšné, vyskytují se dva potenciální problémy: vysoká cena výpočtu a potenciálně slabé řazení. Výlučné použití globálního kontextu a nestabilita řadících algoritmů s sebou nese přídavné sémantické problémy.

Nejobecnějším přístupem v distribuovaných vyhledávání je federace nezávisle řízených metavyhledávačů a více specializovaných vyhledávacích strojů. Tyto stroje poskytují cílené vyhledávací služby ve specifikované doméně (např. pro dané téma). Metavyhledávače zpracovávají uživatelské dotazy efektivně a účinně tak, že je distribuují pouze k nejvhodnějším vyhledávacím strojům.[10]

3.2 Problémy webových vyhledávacích strojů

Problémy webových vyhledávacích strojů jsou spojeny s každou komponentou jejich architektury a každým procesem, který uskutečňují. Vyhledávací stroje například nejsou schopny aktualizovat indexy ve stejné rychlosti, jakou se web vyvíjí. Dalším problémem je kvalita vyhledávání. Ta do jisté míry závisí na kvalitě webu. Lze diskutovat nad následujícími problémy, jak je uvedeno v [10]:

- Jak obnovovat stránky, tj. které stránky a jak často – většina vyhledávacích strojů aktualizuje stránky zhruba na měsíční bázi. Obdržený graf webu je tedy vždy neúplný a výpočet globálního řazení tedy méně přesný.
- Jak minimalizovat zatížení navštívených stránek – sběr stránek spotřebovává zdroje (diskové paměti, CPU) patřící jiným organizacím Robot by měl minimalizovat svůj vliv na tyto zdroje. Stejný problém nicméně pociťuje uživatel během svého přístupu na web, protože pro něj čas potřebný pro natažení dat z webu představuje podstatný problém.
- Jak paralelizovat proces prolézání webu – vyhledávací stroj může používat paralelně více robotů a je tak nutné zajistit, aby neprováděly stejnou práci.

Mezi problémy týkající se kvality řazení stránek, které by měly být se zvýšeným úsilím studovány a zkoumány patří:

- Spam – Aby dosáhli lepšího zařazení stránek, zkoušejí někteří autoři stránek záměrně ovlivňovat jejich umístění v pořadí zařazení pomocí zakázaných technik.
- Kvalita obsahu – Existuje mnoho příkladů ukazujících, že webové stránky obsahují protichůdné informace. Přesnost a spolehlivost dokumentu tak není garantována
- Kvalita vyhodnocení – Protože přímá zpětná vazba od uživatelů není spolehlivá, vyhledávací stroje často sbírají zpětnou vazbu od uživatele implicitně použitím webových *žurnálů dat* (weblogs). Jde o soubory, které se automaticky udržují v průběhu dotazování.

3.3 Sémantický web

Sémantický web představuje v principu relační databázi publikovanou na webu. Nikoli však pouze v podobě v ní uložených dat, ale s příslušným kontextem, se strojově čitelnou informací o významu dat. Tato technologie tedy dává obsahu webu nový rozměr. To umožní integraci dat z různých webových (ale i vnitrofiremních) zdrojů a daleko širší možnosti jejich zpracování. Sémantický web totiž nepředstavuje nic jiného, než existenci pavučiny webových stránek v takové podobě, která umožní její strojovou čitelnost. To následně přinese uživatelům daleko pohodlnější manipulaci s webovými daty včetně možnosti jejich vzájemného propojení.[9],[11]

3.3.1 Hlavní prvky sémantického webu

Jedním ze základních kroků k vytvoření sémantického webu je konceptualizace dat dostupných na internetu. Jedním z klíčových nástrojů konceptualizace jsou **ontologie**. Ontologie lze charakterizovat jako formalizované reprezentace znalostí určené k jejich sdílení a znovupoužití. Ontologie jsou často doménového (oborového) zaměření a bývají konstruovány jako pojmové (konceptuální) hierarchie nebo sítě.

(Polo)automatizované zpracování informací v sémantickém webu může být realizováno pomocí **softwarových agentů**, což jsou do určité míry autonomní inteligentní programové komponenty pohybující se obvykle v distribuovaném prostředí a schopné realizovat požadavky na vyhledávání informací, transakce apod.

Důležitým předpokladem sémantického webu je rovněž **standardizovaný popis webových zdrojů**. Zdrojem se v této souvislosti rozumí cokoliv, co je dosažitelné prostřednictvím sítě WWW, tedy textové dokumenty, obrázky, videosekvence, zvukové soubory apod. Každý zdroj by byl vybaven stejnými charakteristikami (autor, typ zdroje, klíčová slova atd.), což by umožnilo uživatelům internetu pracovat se službou WWW jako s relační databází a dotazovat se na její obsah prostřednictvím jazyků podobných SQL. Významným důsledkem by například byla velmi vysoká přesnost a relevance odpovědi na vyhledávací dotaz, což znamená, že by byl uživateli při vyhledávání určité informace vrácen seznam všech zdrojů, které se této informace týkají, a žádný zdroj navíc.[9]

Jakkoliv se zdá koncepce Sémantického webu futuristická, je jisté, že se stále vyvíjí a investují se do ní nemalé prostředky. Často se řeší projekty, které nezahrnují web jako celek, ale jeho jisté části, vedoucí ke koncepcím, jakými jsou např. tematické portály.[10]

4 Metoda SEO

Search Engine Optimization (SEO) je úprava obsahu a zdrojového kódu jednotlivé stránky tak, aby se objevila na předních místech ve fulltextových vyhledávačích.
Metody v SEO se dají rozdělit na dvě skupiny. První z nich jsou faktory, kterými optimalizujeme stránky přímo při vytváření jejich obsahu, a nazývají se on-page faktory. Druhou skupinou jsou takzvané off-page faktory, což je souhrn všech zpětných odkazů, které na naši stránku míří ze stránek mimo naši doménu, ale i z ní.[3],[12],[13]

4.1 On-page Faktory

On-page faktory jsou faktory ovlivňující výsledek vyhledávání, které můžeme na dané stránce ovlivnit. Patří mezi ně správná volba dynamických a statických URL adres a dále vlastní obsah stránky, ve kterém můžeme ovlivnit titulek, meta elementy, nadpisy, popisky u obrázků, samotný obsah stránky, hlavičky a mnohé další.

4.1.1 URL Adresa

URL adrese je přikládána velká důležitost mnoha vyhledávači, a to z různých pohledů. Čím kratší URL adresa je (tedy má co nejmenší počet zanořených adresářů), tím lepší má hodnocení, než adresa s dlouhou URL adresou. Největší význam je kladen na doménové jméno. Toto může být nevýhoda pro weby na různých freewarových serverech, které tak mohou být handicapovány tematickou nevyhraněností celé domény a přílišnou délkou URL adresy.
Dalším důležitým faktorem jsou klíčová slova obsažená v URL, kterým vyhledávače přikládají určitou váhu. V této souvislosti je dobré se ve struktuře URL vyhnout podtržítku a nahradit jej pomlčkou, protože například Google jej používá pro spojení slov a došlo by tak ke vzniku nového, nechtěného slova.
Parametry v URL (tj. URL typu http://www.domena.cz/stranka.php?id=1234) si některé (opět spíš starší) vyhledávače vykládají jako stránku s dynamickým obsahem, který nestojí za to indexovat. Vždy je proto lepší optimalizace klasické, statické HTML stránky. S optimalizací statických URL adres souvisí také fakt, že neexistuje způsob,

jakým přejmenovat stránku A.html na stránku B.html, aniž by tato stránka přišla o své vybudované postavení u vyhledávačů. Proto je důležité při plánování struktury URL vše dopředu důkladně promyslet. Ne přímo s optimalizací, ale spíše se schopností uživatele snadněji se pohybovat po webu souvisí tvorba URL adresy. Někteří se pomocí umazávání části URL snaží pohybovat po webu jednodušeji (např. z http://www.sperky/zlato/prsteny.html na http://www.sperky/zlato/ ve snaze dostat se ke zlatým šperkům) a je vhodné jim to umožnit. [1]

4.1.2 Titulek

Titulek říká vyhledávači, jaký má zobrazit text ve výsledcích hledání a je proto velmi důležitý. Každá stránka na webu by měla obsahovat jedinečný titulek popisující její obsah a pro danou stránku zvolená klíčová slova. Titulek by měl být dlouhý přibližně 40 až 60 znaků, tedy tolik, kolik zobrazují vyhledávače.[12]

4.1.3 Metainformace

Vyhledávače pracují s těmito parametry:

4.1.3.1 Description

Jedná se o stručné shrnutí obsahu stránky a měl by obsahovat dostatek klíčových slov. Většina vyhledávačů tento meta tag indexují, některé ovšem ne. Mnohé vyhledávače jej používají jako popis stránek při zobrazování výsledků vyhledávání a proto by měl být napsán natolik výstižně, aby přesvědčil uživatele ke kliknutí. Stejně jako by měl být titulek unikátní pro každou stránku, i tento tag by měl být pro každou stránku odlišný. Doporučená délka je max. 250 znaků.[3]
Zápis: *<meta name="description" content="Popis stránky">*

4.1.3.2 Keywords

Význam tohoto elementu je stejně jako u předchozího rozdílný u různých vyhledávačů. Některé mu přikládají velký význam, některé jej indexují jako běžný text a některé (jako například Google) jej ignorují úplně. Opět platí, že keywords mají být pro danou stránku unikátní. Pro metainformace obsažené v tagu Keywords, ale i Description a Title, je důležité, aby byly konzistentní s obsahem celého dokumentu. Pokud totiž budou tyto tagy obsahovat klíčová slova, která se ve zbylém textu nebudou vyskytovat, vyhledávač je vůbec nezohlední.[3]

Zápis: *<meta name="keywords" content="klíčová slova">*

4.1.3.3 Robots

Pomocí tohoto elementu můžeme nastavit, jak má indexovací robot zacházet se stránkou. Můžeme mu indexování povolit, či zakázat. Implicitní hodnota je "vše povoleno".[3]
Zápis: *<meta name="robots" content="all, follow">*

4.1.3.4 Language

Element Language zaručuje správné zobrazení textu a je proto nezbytným i bez optimalizace.[3]
Zápis: <meta http-equiv="Content-language" content="cs">

4.1.3.5 Meta element Googlebot

Google uvedl speciální meta element, aby umožnil autorům stránek ovlivňovat své chování. Zápis vypadá následovně:
<meta name="googlebot" content="snippet/nosnippet, archive/noarchive" />
Meta element Googlebot pracuje se dvěma parametry:

- Archive – zakazuje nebo povoluje ukládání stránky do Google archivu. Výchozí hodnota je "ukládání povoleno".
- Snippet – určuje, co bude zobrazeno uživateli ve výsledcích hledání. Jestli obsah elementu Description nebo text „před a za" hledaným slovem. Výchozí hodnota je "text před a za".

Archiv stránek je výborná funkce a mnoha lidem se určitě líbí. Pokud je nějaký web zrušen a přestane fungovat, můžete se přesto díky archivu na něj podívat. Pokud ale jeho autor chce web z análů vymazat kompletně, stačí, aby do hlavičky uvedl meta element archiv, kde zakáže archivování svého webu. Potom musí počkat až Google zaindexuje novou verzi stránek a jeho web z archivu zmizí.
Je ovšem možné využít zákazu archivování i při běžném provozu webu. Například u stránky, kde je generován dynamický obsah a jejíž starší obsah je úplně nerelevantní – to může být případ kurzů měny. Zde archiv stránky ztrácí smysl.
Určitě by se ale zakazování nemělo přehánět. Je totiž plno případů, kdy nám může archiv pomoci. Pokud má web výpadek, mnoho jiných způsobů, jak se k důležitým informacím dostat, není. Další případ může být web, kde je starší obsah posouván mimo hlavní

stránku a nahrazován novým obsahem. Potom by vyhledání informace mohl být problém, ale s využitím archivu je vše otázkou několika vteřin.

Snippet je část stránky kolem hledaného slova, kterou Google zobrazí v jejím popisku na stránce s výsledky hledání. Někdy je tato informace užitečná, ale někdy se zobrazuje například část menu nebo jiné nezajímavé údaje. Proto dal Google webmasterům možnost zvolit si, co se bude jako popisek zobrazovat. [2]

4.1.4 Vlastní obsah stránky

Obecně se dá říci, že čím kvalitnější je obsah webu, tím méně nutná je jeho optimalizace, protože na kvalitně napsaný text odkazuje více lidí.

4.1.4.1 Klíčová slova

Starší a jednodušší vyhledávače často hodnotí jen frekvenci klíčových slov. Čím větší, tím lepší. Vyspělejší vyhledávače již poměřují hustotu klíčových slov s celkovým počtem slov na stránce, ale stále zde platí, že čím je tento poměr větší, tím je to pro stránku z hlediska vyhledání lepší. Nic by se ovšem nemělo přehánět, protože stránky s příliš velkým poměrem mohou být vyspělými vyhledávači, které se soustředí na optimální poměr, penalizovány či dokonce ignorovány. Některé vyhledávače také zohledňují to, kde se klíčová slova v textu nacházejí. Pokud jsou na začátku textu, je jim přikládána větší váha, pokud na konci textu, tak menší, či dokonce nulová, protože vyhledávač indexuje pouze omezené množství textu a k takto vzdálenému slovu se už nedostane. Dalším faktorem při vyhledávání je vzdálenost klíčových slov od sebe. Ideální je, pokud vytvářejí vyhledávanou frázi uživatelem. Pokud uživatel hledá kombinaci těchto slov, je dobré, pokud jsou co nejblíže k sobě.

4.1.4.1.1 Analýza klíčových slov

Nejdůležitější na začátku je uvědomit si, kdo bude naše stránky navštěvovat, co hledá a jaká slova tedy nejspíš zadá do vyhledávače. Tato analýza se skládá ze dvou částí: vertikální analýzy a laterální analýzy. Vertikální analýza dodá různé tvary daného slova (např. ze slova „sport“ udělá „sportovní“, „sportovní oddíl“, „sportu“ „historie sportu“ atd.). Laterální analýza naopak vyhledá taková slova, která vyhledávají lidé, kteří hledají něco týkající se slova „sport“ (např. „tenis“, „golf“, ale i „plavky”).[7]

Základním cílem analýzy klíčových slov je nalézt co nejvíce slov a frází, která si lidé spojují s určitým tématem. V další fázi je třeba vybrat ta slova, která se pro naše potřeby hodí nejlépe. Obecně by měla klíčová slova splňovat tyto charakteristiky [8]:

- Musí být relevantní. To znamená, že musí co nejlépe vystihovat obsah a cíl našeho webu. Pokud se to podaří, budou web nacházet návštěvníci, kteří hledají přesně to, co stránky nabízejí. Pokud nebudou výsledky hledání dostatečně relevantní, značná část návštěvníků z webu zklamaně odejde.
- Musí mít co nejvyšší frekvenci hledání. Slova, která lidé ve vyhledavačích vyhledávají častěji, jsou lepší než ta, která nehledá skoro nikdo. Hlavním důvodem je samozřejmě objem dosažitelné návštěvnosti, ale není to důvod jediný. Svou roli hraje i srozumitelnost a obvyklost slov. Častěji hledaná slova jsou obvyklejší a lidé jim lépe rozumějí. Jsou tedy vhodnější i z hlediska použitelnosti webu.
- Zejména u obchodních webů je dalším důležitým kritériem míra konkurenčnosti. Pokud se vybrané slovo nachází na příliš mnoha jiných stránkách, bude obtížnější uspět s novou stránkou vlastní. Málo konkurenční a přitom frekventovaná slova jsou tak velmi výhodná.

4.1.4.1.2 Nástroje na vyhledávání slov

Na internetu je k nalezení spousta nástrojů, které mohou pomoci s rozšířením a upřesněním klíčových slov. Dle [3] jsou jimi například:

- Wordtracker

Wordtracker je nástroj, který dokáže navrhnout podobná slova a zjistit jejich potenciální návštěvnost. Dále analyzuje tisíce stránek, které na zadané slovo najdou vyhledavače, a hledá na nich další častá slova. Tím odhalí nejen všechna možná synonyma, ale i většinu laterálních slov. Jeho nevýhodou je jednak fakt, že zdarma ukazuje pouze část výsledků (zbylá část je placená), a také, že tento nástroj lze použít pouze pro hledání slov v angličtině. Pro hledání synonym v češtině lze použít například vyhledávač Morfeo.

- Google AdWords Keyword Suggestions

Pro návrh dalších slov (i v češtině) je tu Google AdWords Keyword Suggestions, nástroj primárně navržený k tomu, aby pomáhal inzerentům k propagaci jejich stránek v placeném systému Adwords. Dokáže nalézt další hledaná slova, podobná

našemu zvolenému. Jeho nevýhodou je, že neukazuje konkrétní počty hledání těchto slov, pouze je seřadí podle četnosti hledání.

4.1.4.2 Popisky u obrázků (atribut ALT u tagu <IMG>)

Další věcí, na kterou kladou vyhledávače váhu, jsou atributy ALT, které se používají u obrázků a které říkají robotu, co se na daném obrázku nachází. Dalším atributem sloužícím k popisu obrázku je volitelný atribut title. Ten ukazuje, co je na obrázku, pokud nad ním uživatel zůstane myší, a měl by tedy obsahovat obecné shrnutí obrázku. Jeho váha je nižší, než u atributu ALT.[13]

<img src="obrazek.gif" alt="obsah obrázku" title="shrnutí obsahu">

4.1.4.3 Nadpisy <H1>,<H2>,<H3>

U nadpisů je vhodné důsledně dodržovat tagování elementy h1 až h3. Robot pozná, že se jedná o nadpis a slovům v něm obsaženým přidá váhu závislou na velikosti nadpisu. H1 smí být na stránce pouze jednou, ostatní vícekrát.[3]

4.1.4.4 Zvýrazněný text <STRONG> a <EM>

Text na stránce je možné zvýraznit tagy strong pro tučný text a em pro kurzívu, ale význam mu přikládá jen vyhledávač Yahoo! a to ještě nepříliš velký.

4.1.5 Rámce

Rámce (frames) jsou dodnes používaný způsob tvorby stránek, vedle tabulek a CSS je to základní možnost jak stránku rozvrhnout. Základní soubor rámců definuje, v které části se nachází menu (u nás nejčastěji vlevo a nahoře) a v které části se načte vlastní obsah stránky. Po kliknutí na položku v menu se obsah zobrazuje v příslušné části. Rámce se začaly používat hlavně z důvodu, že kód navigačního menu byl na jednom místě a také se pomocí nich dá ušetřit na množství dat přenesených k uživateli.

Hlavní problém rámců je s vyhledávacími roboty. Některé totiž rámce úplně odmítají indexovat a zaindexují pouze část noframes. Pokud již vyhledávače zaindexují obsah jednotlivých stránek v rámcích, nastane problém při zobrazení uživateli. Nabídnou mu totiž odkaz na konkrétní stránku a ne na rámce. Uživateli se pak většinou zobrazí stránka bez navigačního menu, což značně snižuje pravděpodobnost, že takový návštěvník na stránce zůstane déle (nemá na co kliknout a ne každý umí odmazat část URL). Pokud jsou již rámce nutnost, optimalizuje se alespoň obsah značky noframes. Ten se pak

ukazuje ve vyhledávačích. Bohužel lze najít celou řadu stránek, ve kterých je použito něco podobného[3]:

```
<noframes>
  <body>
  Váš prohlížeč není podporován.
  </body>
</noframes>
```

4.2 Off-page faktory

Off-page faktory jsou takové faktory, kterými nemůžeme přímo ovlivnit naši stránku. Jedná se především o odkazy, které na ni směřují, a to jak z cizích stránek, tak i z našich vlastních. Tyto off-page faktory získávají v poslední době stále větší váhu při řazení výsledků ve vyhledávači.[7]

4.2.1 PageRank

PageRank je algoritmus sloužící k ohodnocení relativní důležitosti webových stránek, navržený Larry Pagem a Sergeyem Brinem na Standford University v Kalifornii, tvořící základ vyhledávače Google. Tento algoritmus využívá strukturu hypertextových odkazů jako vzájemné „doporučování" stránek. Předpokládá se tedy, že čím více je na stránku odkazováno z kvalitních webů, tím kvalitnější poskytuje obsah. Čím je tedy vyšší hodnota PageRanku, tím je umístění stránky ve výsledcích vyhledávání lepší.[15]

4.2.1.1 Způsob výpočtu Google PageRanku

PageRank je hodnocení stránky na základě kvality odkazů, které na ni vedou. Hodnota PageRanku představuje součet PageRanků, které jsou stránce předány ze stránek, které na ni odkazují. Pokud má stránka A PageRank 5/10 a obsahuje 50 odkazů, pak každý odkaz předá cílové stránce právě padesátinu PageRanku zdrojové stránky, tedy A50 = 0,1/10. PageRank cílové stránky B se získá součtem všech PageRanků předaných s odkazy. Aby cílová stránka B získala také PageRank 5/10, muselo by na ni vést padesát odkazů s PageRankem 0,1/10.[5]

4.2.1.2 Tématicky orientovaný PageRank

Tematický PageRank je takový PageRank, kdy vyhledávač zohledňuje nejen samotné hodnoty předávaného PageRanku, ale také zaměření obou webů (odkazujícího

a odkazovaného). Pokud tedy odkazuje stránka zabývající se prodejem zlatých šperků na článek o zpracování zlata, má takový odkaz větší váhu, než kdyby odkaz směřoval na stránky zoologické zahrady. Hlavním smyslem tematického PageRanku je tedy eliminace spamových stránek, které mají odkazy zakoupené. Druhým možným využitím tematického PageRanku je, že pokud vyhledáváme v kontextu (například tím, že na webové stránce zvýrazníme slova a ta jsou použita pro vyhledávání), je užito téma daného kontextu, v němž se vyhledávaná slova nachází. Jiné výsledky pak dostaneme, hledáme-li slovo "architektura" z dokumentu o designu CPU, a jiné výsledky budou, týká-li se dokument stavby budov. [4]

4.2.1.3 Badrank

Badrank, neboli nulový Pagerank, může mít stránka z několika důvodů. Buď je zcela nová, nebo na ni zatím vede minimum odkazů. Druhou možností jak získat PageRank 0 je penalizace, ke které vyhledávače přistupují při odhalení zakázaných metod optimalizace (viz níže). Tato penalizace může být buď automatická, nebo na základě stížnosti na danou stránku. Problém penalizace je však spojen i s BadRankem a s takzvaným špatným okolím stránky. Vezměme v úvahu dvě stránky. Stránka A je penalizovaná, stránka B nikoliv. Pokud A odkazuje na B a B neodkazuje zpět, BadRank se nepřenáší. Pokud však stránka B odkazuje zpět, dostane od penalizované A BadRank. BadRank se tedy přenáší pouze pokud je odkazováno na penalizovanou stránku zpět, a to z důvodu, že nelze ovlivnit odkaz z penalizované stránky. [3]

4.2.1.4 Rankingy a ostatní vyhledávače

Principu PageRanku využívají mnohé další vyhledávače. Jsou mezi nimi drobné rozdíly, které ale pro SEO nemají téměř žádný význam. Pro všechny webové stránky totiž platí, že čím více kvalitních odkazů na ně z ostatních webů směřuje, tím lépe. Český Seznam.cz využívá S-ranku, Atlas JyxoRanku. Malou výjimku tvoří Web Rank od Yahoo!, který je kombinací odkazů vedoucích na stránku, klikavosti ve výsledcích hledání a návštěvnosti měřené pomocí Yahoo! toolbaru.[3]

4.2.2 Získávání zpětných odkazů

Získávání zpětných odkazů je práce velmi složitá, především potom pro začínající webové stránky. Možnosti, jak získat zpětný odkaz, jsou dle [13] následující:

- registrací webu do internetových katalogových vyhledávačů,

- výměnou odkazů s ostatními (nejen) tematicky spřízněnými weby,
- napsáním článku o tématu, kterým se na webu zabýváme a který je umístěn na jiných webových stránkách,
- od návštěvníku našeho webu, které zaujal jeho obsah a odkáží na něj na svých stránkách bez našeho vědomí,
- účastí v internetových diskuzích se zanecháním zpětného odkazu,
- rozesláním tiskových zpráv v HTML s odkazy.

5 Zakázané metody optimalizace

Vedle optimalizace stránek pro vyhledávače existuje také takzvaný SEO spam, což jsou vlastně nekalé metody sloužící k umělému zlepšování pozic ve výsledcích vyhledávání. Weby využívající tyto metody jsou vyhledávači tvrdě penalizovány, v některých případech mohou být z výsledků hledání úplně vyloučeny.

5.1 Skrytý text nebo odkazy

Vzhledem k obtížnosti tvorby takového textu, který by měl optimální hustotu, resp. maximální frekvenci klíčových slov při současném zachování stylistické hodnoty se někteří optimalizátoru uchylují k použití skrytého textu, tedy takového textu, který má stejnou barvu jako pozadí stránky a je tak neviditelný. Tato technika může být účinná pouze u starších vyhledávačů, které neviditelný nepoznají a indexují jej jako viditelný. Moderní vyhledávače už tento text rozpoznají a penalizují. Pro uživatele existuje možnost také tzv. spam reportu, kterým může vyhledávač na tento "podvod" sám upozornit a ten stránku tvrdě penalizuje. Například vyhledávač Google má svůj Google Report, Seznam využívá Seznam Report. [3],[13]

5.2 Malý text

Jedná se o podobnou techniku jako u skrytého textu. Jedná se o malý, nečitelný text, umístěný většinou na konci stránky.[3]

5.3 Cloaking

Cloaking je taková technika, kdy se vyhledávači předkládá jiná stránka, než jakou vidí uživatel. Tato stránka, určená vyhledávači, je pak pro něj speciálně optimalizovaná. I tato metoda je poměrně snadno odhalitelná. Vyhledávače tuto metodu kontrolují tak, že jejich

robot se identifikuje jako běžný uživatel a poté porovnává obě indexované verze. Stránkám, jež jsou usvědčeny z tohoto podvodu, hrozí permanentní vyřazení z výsledků vyhledávání (např. Google).[3]

5.4 Odkazové farmy

Odkazové farmy neboli Link farmy, jsou speciální weby čítající několik tisíců až milionů navzájem velmi dobře propojených stránek zpětnými odkazy. Takto prolinkovaným stránkám se zvyšuje PageRank a web se tak nachází velmi vysoko ve výsledcích vyhledávání. Link farmy mohou tvořit také navzájem propojené weby. Odhalování i trestání těchto webů je poněkud složitější, přesto se nedoporučuje do těchto farem zapojovat.[3]

5.5 Duplicitní obsah na více URL

Duplicitní stránky jsou takové stránky, jejichž obsah je stejný na více URL adresách. Vyhledávače se tomuto brání, protože není v zájmu jejich, ani uživatelů, aby zobrazovaly duplicitní obsah ve výsledcích vyhledávání více než jednou. Odhalit duplicitní obsah pro ně však není příliš jednoduché a v případě nahlášení duplicitních stránek pomocí spam reportu nedochází vždy k penalizaci, ale ke kontrole algoritmu pro kontrolu zobrazení duplicitních stránek.[3]

5.6 Komentářový spam

Podstatou této metody je roznášení zpětných odkazů na stránky robotem do nejrůznějších diskusních fór. Boj vyhledávačů s tímto problémem je složitější proto, že nerozpoznají souvislost odkazů s tématem a je tak na autorech postižených stránek, aby se proti těmto robotům bránili sami, například antispamovými filtry, povinnou registrací uživatelů a dalšími metodami.[3]

5.7 Klamná či opakující se slova

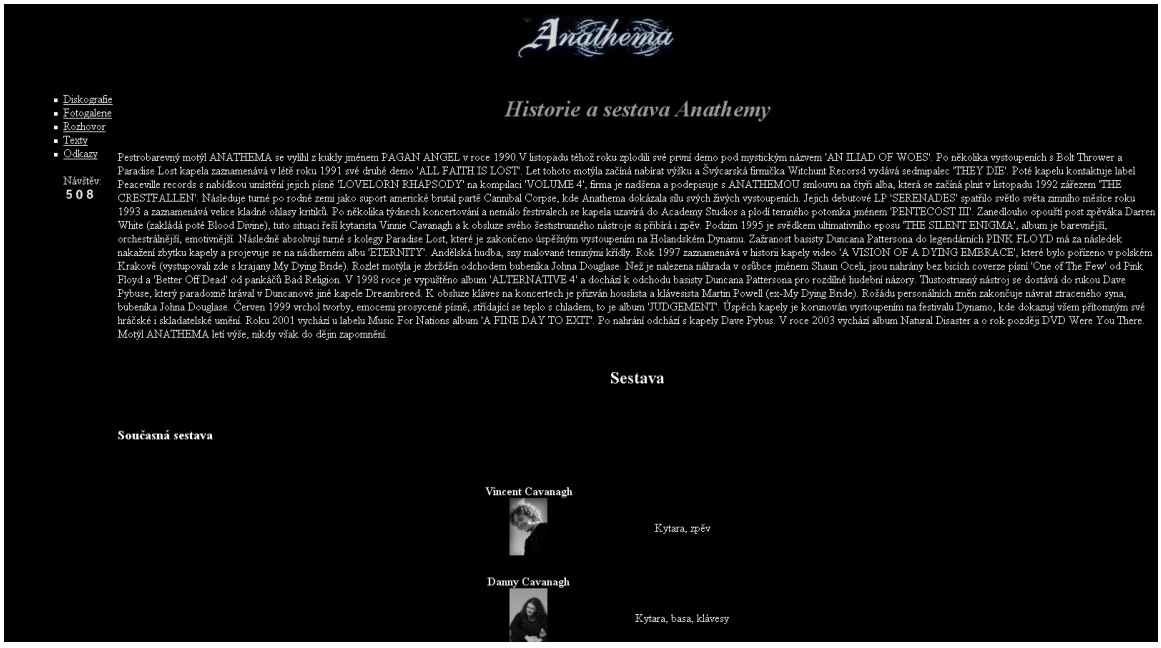
Do textu stránky jsou vkládána klíčová slova v nejrůznějších tvarech, která nemají žádnou souvislost s tématem stránky, za účelem zlepšení výsledku při vyhledávání.[3]

6 Vliv optimalizace na vyhledávání

Cílem této kapitoly bude ověřit vliv optimalizace na umístění stránek ve vyhledávačích.

6.1 Představení testovaných stránek

K posouzení vlivu optimalizace na výsledky vyhledávání internetovými vyhledávači byly použity vlastnoručně vytvořené stránky hudební skupiny Anathema. Anathema je alternativní rocková (v raných dobách metalová) skupina z Velké Británie, založená v 90. letech minulého století. Vytvořené stránky se skládají z úvodní vstupní strany a dále z jednotlivých stránek, věnujících se historii a složení této skupiny, stránky obsahující fotografie Anathemy, rozhovor s kapelou, texty písní vybraného alba, diskografii kapely a stránky s odkazy na další materiál o kapele na internetu. Všechny tyto stránky byly zpočátku zcela neoptimalizovány – neměly optimalizované ani on-page, ani off-page faktory. Byly nahrány na webový server http://www.wz.cz na adresu http://anathema.webz.cz a během doby potřebné k zařazení stránek internetovými vyhledávači do výsledků svých vyhledávání bylo přistoupeno k optimalizaci těch samých stránek v off-line podobě uložených na počítači. Na obrázku 2 je ukázka jedné ze stránek kapely.



Obrázek 2: Stránky skupiny Anathema; zdroj: vlastní

6.2 Optimalizace on-page faktorů

Aby bylo možné posoudit jak význam optimalizace on-page faktorů, tak optimalizace off-page faktorů, byly stránky pro obě tyto skupiny faktorů optimalizovány zvlášť. Nejprve bylo přistoupeno k optimalizaci on-page faktorů.

6.2.1 Optimalizace URL adres

Do www adres, odkazujících na jednotlivé stránky, bylo přidáno více klíčových slov specifických pro každou z nich. Původní a optimalizovaná struktura URL adres po optimalizaci je následující:

Původní: *diskografie.html*
Po optimalizaci: *diskografie-metalove-kapely-anathema.html*
Původní: *fotografie.html*
Po optimalizaci: *fotografie-clenu-kapely-anathema.html*
Původní: *historie.html*
Po optimalizaci: *historie-kapely-anathema.html*
Původní: *index.html*
Po optimalizaci: *index-vstupni-stranka-anathemy.html*
Původní: *odkazy.html*
Po optimalizaci: *odkazy-souvisejici-s-anathemou.html*
Původní: *rozhovor.html*
Po optimalizaci: *rozhovor-s-kapelou-anathema.html*
Původní: *texty.html*
Po optimalizaci: *texty-pisni-kapely-anathema.html.*

6.2.2 Změny doménové adresy

V rámci optimalizace stránek také několikrát došlo ke změně doménové adresy. Jak bylo zmíněno v úvodu kapitoly, původní neoptimalizované stránky byly testovány na adrese http://anathema.webz.cz. Poté, co byla provedena optimalizace on-page faktorů, byla změněna adresa těchto stránek na http://anathema.unas.cz. Stránky s optimalizovanými off-page (a posléze i on-page) faktory byly umístěny na adresu http://anathema.cz.

6.2.3 Odstranění rámců

Protože mnohé z indexovacích robotů odmítají indexovat rámce nebo indexují pouze tag <noframes>, byly ze všech testovaných stránek odstraněny postranní rámce s menu obsahujícím další odkazy a horní rámce s logem kapely. Postranní rámec byl nahrazen následujícím:

<table>
<tr>

<td align="left" width="750" valign="top">

<ul type="square">

* Historie a sestava
*

* Diskografie
*

* Fotogalerie
*

* Rozhovor
*

* Texty
*

* Odkazy
*

*
*

</td>

</tr>

</table>

Horní rámec nahradil následující HTML kód:

* <img src="anathema.JPG"> *

6.2.4 Optimalizace titulků

Každá stránka byla opatřena unikátním titulkem popisujícím, co daná stránka obsahuje.

Původní: *Diskografie*

Po optimalizaci: anathema.unas.cz – *Diskografie kapely Anathema*

Původní: *Fotogalerie*

Po optimalizaci: anathema.unas.cz – F*otografie členů kapely Anathema a fotky z jejich vystoupení*

Původní: *Historie*

Po optimalizaci: anathema.unas.cz – *Historie Anathemy, její současní a minulí členové*

Původní: I*ndex*

Po optimalizaci: anathema.unas.cz – *Vstupní stránka Anthemy*

Původní: O*dkazy*

Po optimalizaci: anathema.unas.cz – *Odkazy související s kapelou Anathema*

Původní: Texty

Po optimalizaci: anathema.unas.cz – *Anathema: Texty písní skupiny Anathema - album Alternative 4*

Původní: Rozhovor

Po optimalizaci: anathema.unas.cz – *Anathema: Rozhovor s kapelou Anathema - Vincent Cavannagh*

6.2.5 Optimalizace metainformací

Meta tagy v neoptimalizované verzi zcela chyběly, proto musely být ke každé stránce vytvořeny nové.

6.2.5.1 Description

Pro stránku: historie-kapely-anathema.html

<meta name="description" content="Historie Anathemy od počátku až po současnost. Seznam současných i bývalých členů a hostů..">

Pro stránku*:* fotografie-clenu-kapely-anathema.html

Zápis: *<meta name="description" content=" Zde se nacházejí fotografie Anathemy, promo fotografie členů, ale i celé kapely, dále fotografie z koncertů a vystoupení.“>*

Pro stránku: diskografie-metalove-kapely-anathema.html

Zápis: *<meta name="description" content="Diskografie kapely Anathema – CD a DVD The Crestfallen, Serenades, Pentecost, Silent Enigma, Eternity, Alternative, Judgement, Resonance, Natural Disaster, Fine Day To Exit, Were You There Live.">*

Pro stránku: texty-pisni-kapely-anathema.html

Zápis: *<meta name="description" content="Texty písní skupiny Anathema z desky Alternative 4. ">*

Pro stránku: rozhovor-s-kapelou-anathema.html

Zápis: *<meta name="description" content="Zde je obsáhlý rozhovor s kapelou Anathema, kterou v něm reprezentuje Vincent Cavannagh. ">*

Pro stránku: odkazy-souvisejici-s-Anathemou.html

Zápis: *<meta name="description" content="Další zajímavé odkazy související s Anathemou. Oficiální stránky kapely a další. ">*

6.2.5.2 Keywords

Pro stránku: historie-kapely-anathema.html

<meta name="keywords" content="Anathema, historie, sestava, současná sestava Anathemy, bývalí členové Anathemy, hosté">

Pro stránku: diskografie-metalove-kapely-anathema.html

<meta name="keywords" content="Anathema, diskografie, CD, DVD, alba Anathemy, disk, deska, desky ">

Pro stránku*:* fotografie-clenu-kapely-anathema.html

<meta name="keywords" content="Anathema, fotografie, fotky, promo fotografie, živé fotografie Anathemy, fotky z vystoupení">

Pro stránku: texty-pisni-kapely-anathema.html

<meta name="keywords" content="Anathema, texty Anathemy, písně, slova písní, Alternative 4">

Pro stránku: rozhovor-s-kapelou-anathema.html

<meta name="keywords" content="Anathema, rozhovor s Anathemou, Vincent Cavannagh">

Pro stránku*:* odkazy-souvisejici-s-Anathemou.html

<meta name="keywords" content="Anathema, odkazy, oficiální stránky">

6.2.6 Optimalizace nadpisů

Všechny hlavní nadpisy na stránce byly uzavřeny do pro optimalizaci významného tagu <H1> a také do nich bylo přidáno více klíčových slov, protože jim vyhledávače přikládají poměrně velkou váhu.

<h1 align="center">Diskografie kapely Anathema</h1>

*<h1 align="center">Fotogalerie Anathemy</h1>
*

*<h1 align="center">Historie a sestava Anathemy</h1>
*

*<h1 align="center">Odkazy související s Anathemou</h1>
*

<h1 align="center">Rozhovor s Anathemou</h1>

<h1 align="center">Texty písní Alternative 4</h1>

6.2.7 Optimalizace obrázků

Všechny obrázky na stránkách byly opatřeny atributy ALT a TITLE.

Příklad:

Původně: *<img src="Alba/Crestfallen-cover.jpg">*

Po optimalizaci: *<img src="Alba/Crestfallen-cover.jpg" ALT="Anathema - album Crestfallen" TITLE="Anathema: Obal desky Crestfallen">*

Původně: *<img src="Photos/konc1 s.jpeg" border="1">*

Po optimalizaci: *<img src="Photos/konc1 s.jpeg" ALT="Fotografie z koncertů Anathemy" TITLE="Anathema: Živé fotografie z vystoupení Anathemy" border="1">*

6.3 Optimalizace off-page faktorů

V rámci optimalizace off-page faktorů byly stránky zaregistrovány do následujících katalogových vyhledávačů:

http://jyxo.cz/

http://morfeo.centrum.cz/

http://www.bezvaportal.cz

http://www.aaakatalog.wz.cz

http://www.hledejte.cz

http://www.lep.cz

http://www.infotip.cz

http://www.infoweb.cz

http://www.linkovnik.cz

http://najduvse.cz

http://www.kormidlo.cz

http://www.cent.cz/

http://alfa.elchron.cz

http://www.czprima.cz

http://www.zacatek.cz

Další internetové odkazy na testované stránky byly vloženy do diskuzí na webech s hudební tématikou ke článkům o kapele Anathema. Tyto zpětné odkazy byly umístěny pouze na stránky, jejichž PageRank měl hodnotu alespoň 3. Jedná se o následující weby:

http://www.musicserver.cz

http://www.musiczone.cz

http://www.metalopolis.net

http://www.muzikus.cz

http://www.fobiazine.net

http://www.net-1.cz

http://www.freemusic.cz

http://www.abradio.cz

http://erhyho.blog.cz

http://magazin.musicweb.cz

http://teenage.cz

http://www.abysszine.com

http://www.showpark.cz

http://hudebniskupiny.cz

http://anathema.czweb.org/

http://www.sbernice.net

http://www.xichty.cz

http://www.musicer.net

Protože získávání zpětných odkazů je práce velmi náročná a zdlouhavá, byly některé odkazy umístěny i na servery, které s hudbou příliš nesouvisejí. Opět platí, že PageRank těchto stránek byl minimálně 3 a vyšší.

http://www.xchat.cz

http://www.idnes.cz

http://cislo5zije.bloger.cz/

http://www.xhokej.wz.cz

http://www.CSFD.cz

http://www.hrajeme.cz

http://www.last.fm

http://www.vltava.cz

http://kf.hocke.eu/

http://jam.hocke.eu/

http://dorrit.blog.cz

http://www.nabidky.cz

http://www.letmo.net

Tímto způsobem bylo získáno přes 45 zpětných odkazů, což by měl být již dostatečně vysoký počet, který by se na off-page optimalizaci měl výrazněji projevit. K tomu je potřeba připočítat dalších přibližně 25 odkazů, kterými jsou vzájemně provázány optimalizované stránky. Dohromady se tedy povedlo získat více než 60 zpětných odkazů.

6.4 Představení testovacích kritérií

Všechny verze stránek s rozdílným stupněm optimalizace byly testovány ve dvou českých vyhledávačích Jyxo a Morfeo (jeho výsledky přebírá druhý nejpoužívanější vyhledávací portál Centrum) a dvou nejvíce používaných zahraničních vyhledávačích Google a Yahoo!. Do posouzení vlivu optimalizace měly být zahrnuty i výsledky z nejpoužívanějšího českého vyhledávacího portálu Seznam.cz, ale protože doba, která byla potřebná pro zařazení stránek do výsledků vyhledávání, byla neúměrně dlouhá, výsledky z tohoto vyhledávače nakonec použity nebyly.

Samotné vyhledávání probíhalo tak, že do zvolených vyhledávačů byla zadávána taková klíčová slova a spojení, která by do nich nejpravděpodobněji zadávali uživatelé, jenž by chtěli získat nejen obecné informace o skupině Anathema, ale i uživatelé, kteří by hledali konkrétní informace (informace o historii skupiny, textech písní, fotografie skupiny a další), jímž se věnují jednotlivé stránky této internetové prezentace. Do tabulky byla zaznamenána jednak pozice, na které se stránky ve výsledcích nacházely, a také celkový počet stránek, které pro dané klíčové slovo či spojení vyhledávač nalezl. Zaznamenané hodnoty jsou uvedeny ve formě – *pozice stránky/celkový počet nalezených stránek vyhledávačem.*

Vyhledávání probíhalo pouze mezi českými stránkami, protože vytvořené stránky nebyly konstruovány pro vyhledávání zahraničními uživateli internetu. Výsledky nalézání jednotlivými vyhledávači jsou uvedeny v tabulkách.

6.5 Výsledky vyhledávání neoptimalizovaných stránek

Jak je uvedeno v tabulce 3, zcela neoptimalizované stránky žádný z vyhledávačů vůbec nezařadil ani po měsíci čekání (což by měla být dostatečně dlouhá doba) do výsledků vyhledávání, a to ani při pokusu o vyhledání jejich přesné internetové adresy anathema.webz.cz. Protože na ně nevedl jediný odkaz, stránky jakoby pro vyhledávací stroje vůbec neexistovaly.

Tab. 3 Výsledky stránek s neoptimalizovanými faktory; zdroj: vlastní

Vyhledávače / Klíčová slova	*Google*	*Yahoo!*	*Jyxo*	*Morfeo (Centrum)*
Anathema	**-/121 000**	**-/96300**	**472155**	**-/5598**
Anathema, historie	-/4110	-/3100	-/388	-/1390
Anathema, fotky	-/24230	-/25500	-/11990	-/375
Anathema, fotogalerie	-/4470	-/5710	-/53	-/980
Anathema, texty	-/10820	-/3100	-/634	-/1077
Anathema, rozhovor	-/3420	-/2750	-/764	-/1037

6.6 Výsledky vyhledávání stránek s optimalizovanými on-page faktory

Výsledky stránek umístěných na adrese http://anathema.unas.cz s optimalizovanými on-page faktory a neoptimalizovanými off-page faktory dopadly úplně stejně jako kompletně neoptimalizované stránky. Ani v tomto případě nebyly po dalším měsíci čekání stránky zařazeny do výsledků a vyhledávání podle klíčových slov i jejich přesné adresy vůbec nevedlo k jejich nalezení. Jak ukazuje tabulka 4, změnil se pouze počet nalezených stránek na zadané hledané slovo pro každý vyhledávač, což je způsobeno vlivem času. Došlo totiž ke změně množství zaindexovaných stránek jednotlivými vyhledávači.

Tab. 4 Výsledky stránek s optimalizovanými on-page faktory, zdroj: vlastní

Vyhledávače / počet výskytů klíčového slova / Klíčová slova	*Google*	*Yahoo!*	*Jyxo*	*Morfeo (Centrum)*
Anathema /	**-/123 000**	**-/97400**	**-/49478**	-/5714
Anathema, historie	-/3890	-/3000	-/395	-/1510
Anathema, fotky	-/26600	-/24900	-/12324	-/402
Anathema, fotogalerie	-/4650	-/5730	-/55	-/1150
Anathema, texty	-/11700	-/3600	-/640	-/1093
Anathema, rozhovor	-/3200	-/2910	-/776	-/1056

6.7 Výsledky vyhledávání stránek s optimalizovanými off-page faktory

Následující tabulka 5 uvádí, jak se stránky umístěné na adrese http://anathema.cz s optimalizovanými off-page faktory umístily zhruba po dvou týdnech.

Tab. 5 Výsledky stránek s optimalizovanými off-page faktory; zdroj: vlastní

Vyhledávače / počet výskytů klíčového slova / Klíčová slova	*Google*	*Yahoo!*	*Jyxo*	*Morfeo (Centrum)*
Anathema /	**118/115 000**	**41/93300**	**24/48125**	**19/5637**
Anathema, historie	13/4230	12/3020	8/401	9/598
Anathema, fotky	26/25120	22/26150	21/12147	6/1436
Anathema, fotogalerie	-/4310	-/5730	-/52	-/417
Anathema, texty	34/10770	30/2040	23/631	18/1094
Anathema, rozhovor	26/3350	17/2730	15/740	22/1025

Nejlepšího výsledku dosáhly stránky u vyhledávače Morfeo (Centrum), který však našel pouhých 5637 výskytů pro toto slovo. Naopak vyhledávač Google zobrazil stránky až na 118. místě, celkový počet nalezených stránek byl však přes 100 000. Špatně si stránky nevedly ani u vyhledávače Yahoo!, který, přestože našel také téměř 100 000 stránek, zařadil stránky na 44. místo. Opět lze pozorovat změnu počtu indexovaných stránek oproti předchozímu stavu.

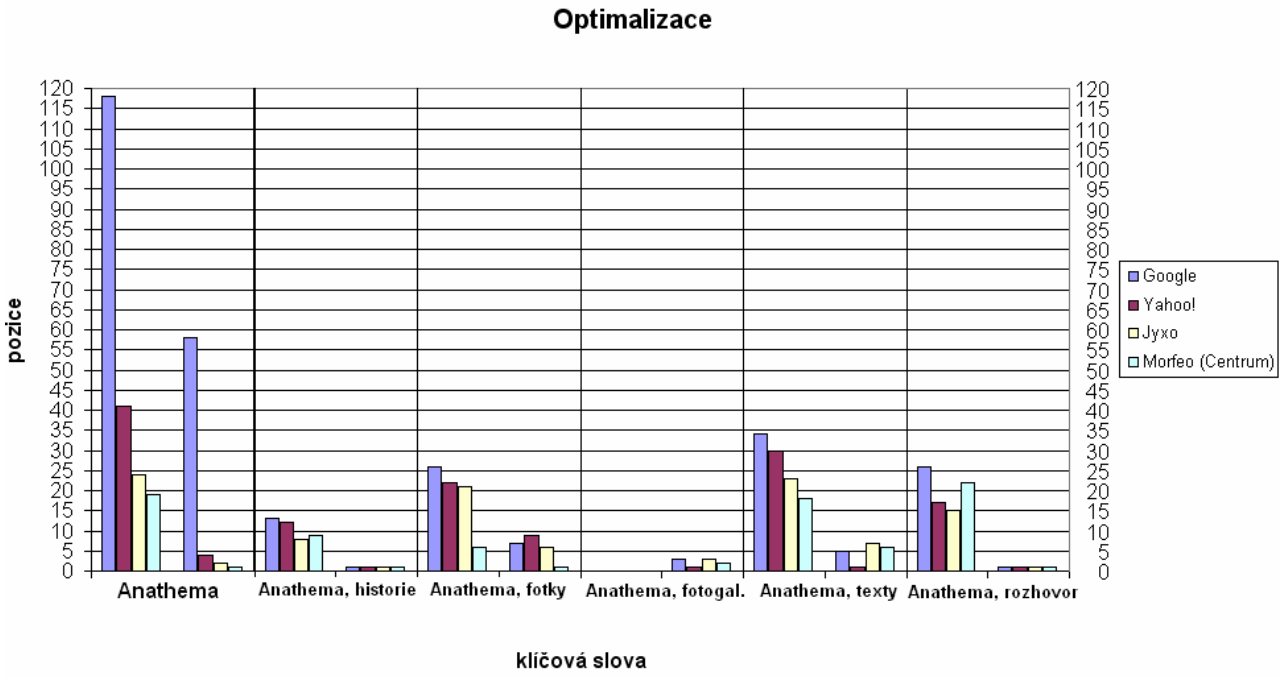
6.8 Výsledky vyhledávání plně optimalizovaných stránek

Stránky, které měly optimalizované on-page i off-page faktory, a tedy již byly kompletně zoptimalizovány, se ve výsledcích vyhledávání následujících vyhledávačů umístily tak, jak je uvedeno v následující tabulce 6. Doba potřebná k získání těchto výsledků byla opět další zhruba dva týdny.

Tab. 6 Výsledky stránek s plně optimalizovanými faktory; zdroj: vlastní

Vyhledávače / počet výskytů klíčového slova	*Google*	*Yahoo!*	*Jyxo*	*Morfeo (Centrum)*
Klíčová slova				
Anathema	**58/118 000**	**4/97300**	**2/46782**	**1/5694**
Anathema, historie	1/3900	1/3020	1/379	1/573
Anathema, fotky	7/23500	9/23500	6/12235	1/1398
Anathema, fotogalerie	3/4580	1/5630	3/49	2/385
Anathema, texty	5/10700	1/3500	7/652	6/1067
Anathema, rozhovor	1/3100	1/2710	1/720	1/669

Na následujícím grafu 1 je možno vidět vliv optimalizace on-page faktorů na posun stránek ve vyhledávačích s již zoptimalizovanými off-page faktory. Každý sloupec obsahuje pozice stránek v jednotlivých vyhledávačích před a po optimalizaci pro dané klíčové slovo.

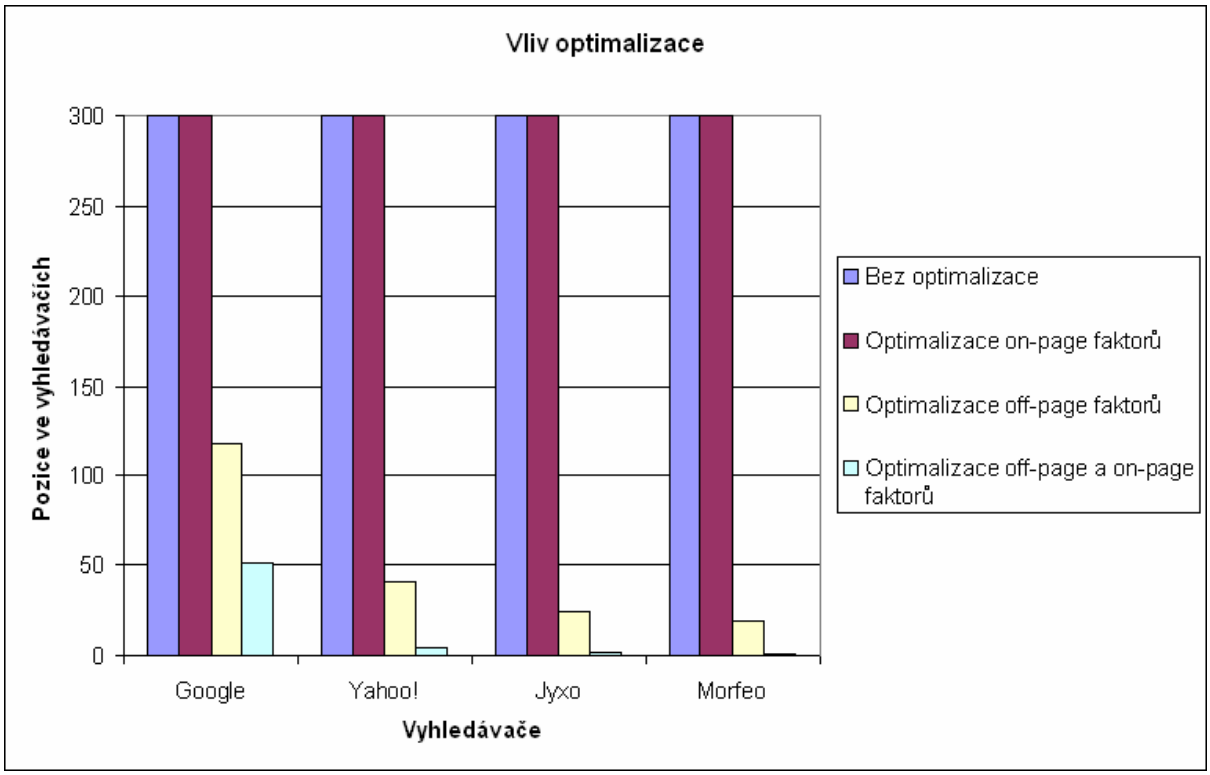


Graf 1: Vliv optimalizace on-page faktorů; zdroj: vlastní

6.9 Zhodnocení vlivu optimalizace

Z předchozích tabulek vyplývá, že optimalizace jednoznačně ovlivnila pozici stránek po jejich vyhledání. V jediném vyhledávači – Googlu, se stránky nepovedlo dostat na první stranu s výsledky. Přesto je posun vpřed o 60 pozic ze 118. na 58. místo díky optimalizaci on-page faktorů velmi významným skokem. U zbylých vyhledávačů Morfeo, Jyxo a Yahoo! se stránky díky optimalizaci posunuly na 1., 2. a 4. místo, které stránkám zaručují vynikající viditelnost ve výsledcích. I vyhledávání specifičtějších klíčových slov vedlo k výborným výsledkům. Stránky, které se před optimalizací on-page faktorů vyskytovaly na 2. až 4. stranách mezi výsledky, se po ní všechny dostaly na stranu první. Pouze ve 3 případech z 20 byly stránky nalezeny na horším než 6. místě. V 11 případech byly umístěny na místě prvním.

Na následujícím grafu 2 je možné vidět, jaký vliv měly jednotlivé fáze optimalizace na posun stránek ve vyhledávačích při hledání klíčového slova Anathema. Svislá osa má z důvodu zachování přehlednosti maximální hodnotu 300 a stránky, které této hodnoty dosahují, nebyly ve výsledcích vůbec nalezeny.



Graf 2: Vliv optimalizace; zdroj: vlastní

7 Závěr

Porovnání výsledků vyhledávání stránek před a po optimalizaci potvrdilo, že optimalizace má zásadní vliv na pořadí umístění stránek ve výsledcích vyhledávání. Ukázalo se, že nejdůležitějším faktorem potřebným pro to, aby se stránky ve výsledcích vůbec zobrazily, jsou off-page faktory. Bez zpětných odkazů, které vyhledávací roboty na naše stránky upozorní, se tyto stránky do výsledků vůbec nedostanou. Jejich optimalizace je tedy základním předpokladem pro to, aby byly stránky na Internetu k nalezení. Čím více bylo těchto zpětných odkazů získáno, tím bylo postavení stránek lepší. Optimalizace on-page faktorů také přispěla ke zlepšení postavení stránek, ale smysl měla až po optimalizaci off-page faktorů. Stránky, které byly ve vyhledávačích teprve díky optimalizaci off-page faktorů k nalezení, se díky ní významným způsobem posunuly ještě výš.

Dá se tedy říct, že správná celková optimalizace stránek je nejen velmi důležitá pro postavení stránek ve vyhledávačích, ale i nezbytná pro samotné jejich zobrazení v nich. Povedlo se tak dosáhnout cíle, který byl stanoven v úvodu bakalářské práce – tedy ověřit význam optimalizace webových stránek z hlediska jejich umístění ve vyhledávačích.

Použitá literatura

[1] *Základní techniky optimalizace stránek pro vyhledavače* [online]. 18. 1. 2002 [cit. 2008-03-02]. Dostupný z WWW: <http://www.lupa.cz>.

[2] *SEO - Search Engine Optimization* [online]. 12. 8. 2003 [cit. 2008-03-10]. Dostupný z WWW: <http://interval.cz>.

[3] SMIČKA, Radim. *Optimalizace pro vyhledávače - SEO : Jak zvýšit návštěvnost webu*. [s.l.] : [s.n.], 2004. 120 s.

[4] *Jak pracuje tematický PageRank* [online]. 24. 3. 2004 [cit. 2008-03-15]. Dostupný z WWW: <http://www.lupa.cz>.

[5] *Google PageRank* [online]. 5. 6. 2003 [cit. 2008-03-17]. Dostupný z WWW: <http://www.lupa.cz>.

[6] *Optimalizace pro vyhledávače – SEO II* [online]. 28. 03. 2006 [cit. 2008-03-02]. Dostupný z WWW: <http://programujte.com>.

[7] *Letní škola SEO* [online]. 01.10.2003 [cit. 2008-02-11]. Dostupný z WWW: <http://seo-sem.robertnemec.com/>.

[8] *Analyzujeme klíčová slova* [online]. 5. 12. 2003 [cit. 2008-02-18]. Dostupný z WWW: <http://www.zive.cz>

[9] *Zpravodaj ÚVT MU* [online]. 01.10.2003 [cit. 2008-03-18]. Dostupný z WWW: <http://www.ics.muni.cz>.

[10] POKORNÝ, Jaroslav, VÁCLAV, Snášel, MICHAL, Kopecký. *Dokumentaristické informační systémy*. [s.l.] : [s.n.], 2005. 184 s.

[11] *Science World* [online]. 01.10.2003 [cit. 2008-03-20]. Dostupný z WWW: <http://www.scienceworld.cz>.

[12] KENT, Peter. *Search Engine Optimization For Dummies*. [s.l.] : [s.n.], 2004. 354 s.

[13] WALL, Aaron Matthew. *Search Engine Optimization Book*. [s.l.] : [s.n.], 2005. 156 s.

[14] *Comscore* [online]. 24.01.2008 [cit. 2008-03-20]. Dostupný z WWW: <http://www.comscore.com/press/release.asp?press=2018>

[15] *Wikipedia* [online]. 23.04.2008 [cit. 2008-04-24]. Dostupný z WWW: <http://cs.wikipedia.org/wiki/PageRank>

[16] *Seo Master* [online]. 20.03.2008 [cit. 2008-04-21]. Dostupný z WWW: <http://www.seomaster.cz/seznam-vyhledavacu-a-popis>

[17] *WebPrezent* [online]. 04.08.2007 [cit. 2008-02-16]. Dostupný z WWW: < http://www.webprezent.cz/webdesign/>

[18] *Mapsys* [online]. 23.03.2003 [cit. 2008-02-17]. Dostupný z WWW: <http://www.mapsys.com/ma3look.htm>