

ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

Technická fakulta

Katedra technologických zařízení staveb



**Způsoby optimalizace webových prezentací pro
internetové vyhledávače**

Bakalářská práce

Vedoucí práce: Ing. Michal Hruška

Autor práce: Martin Němec

Praha 2.3.2010

Prohlášení

*Prohlašuji, že jsem bakalářskou práci na téma „**Způsoby optimalizace webových prezentací po internetové vyhledávače**“ zpracoval samostatně a použil jen literaturu uvedenou v seznamu literatury.*

Ve Velimi dne 25. dubna 2010

.....

Martin Němec

Abstrakt

Tato bakalářské práce se zaměřuje na způsoby, jakým můžeme optimalizovat webové prezentace tak, aby dosahovaly lepších pozic ve výsledcích vyhledávání. Jsou zde uvedeny hlavní On-Page i Off-Page faktory a další vhodné postupy. Práce také obsahuje praktické příklady těchto metod a zápisu kódu při samotné optimalizaci stránek.

Klíčová slova

Internet, web, stránky, klíčová slova, HTML, on page, off page, optimalizace

Summary

This Bachelor thesis is focused on ways how to optimize web presentation to get better ranking in search results. It includes main On-page and Off-page factors and other suitable techniques. This thesis also contains practical examples of these methods and examples of entries of the code in optimization of the websites themselves.

Key words

Internet, web, pages, key words, HTML, on page, off page, optimization

Obsah

1. ÚVOD	6
2. ZÁKLADNÍ POJMY	7
2.1 Historie vyhledávačů	7
2.1.1 Vývoj ve světě.....	7
2.1.2 Vývoj v tuzemsku.....	7
2.2 Druhy vyhledávání:.....	8
2.3 Webová stránka.....	8
2.4 Typy webových stránek.....	11
2.4.1 Statické stránky	11
2.4.2 Dynamické stránky	12
2.5 Druhy vyhledávačů	13
2.5.1 Katalogové vyhledávače	13
2.5.2 Fulltextové vyhledávače	14
2.5.3 META vyhledávače	15
2.6 Spider.....	16
2.7 Indexer.....	17
2.8 Vyhledávač	18
2.8.1 Podíl vyhledávačů na českém trhu	19
3. OPTIMALIZACE A JEJÍ METODY	20
3.1 Optimalizace obecně.....	20
3.1.1 Klíčové faktory pomáhající ve výsledcích vyhledávání.....	20
3.1.2 Nežádoucí faktory	20
3.2 Optimalizace rychlosti	21
3.2.1 Optimalizace na straně web serveru.....	21
3.2.2 Optimalizace kódu stránek a rychlosti zobrazení obsahu	22
3.2.3 Optimalizace prohlížeče	23
3.3 On-page faktory.....	23
3.3.1 URL	23
3.3.2 Klíčová slova (keywords)	23

3.3.3 Stop slova	25
3.3.4 Title	25
3.3.5 Nadpisy	26
3.3.6 Meta tagy	26
3.3.7 Robots.txt	28
3.4 Off-page faktory	29
3.4.1 PageRank	29
3.4.2 Anchor text	32
3.4.3 Zpětné odkazy	32
3.4 Optimalizace URL adresy	34
3.4.3 Změna dynamických adres na statické	34
4. ZAKÁZANÉ TECHNIKY V SEO	36
4.1 Skrytý a malý text	36
4.2 CLOAKING	37
4.3 „Google bomby“	37
4.4 Odkazové farmy	37
4.5 Klamné přesměrování	38
4.6 DOORWAY PAGES	38
4.7 Keyword stuffing	38
5. ZÁVĚR	39
SEZNAM POUŽITÝCH ZKRATEK	40
SEZNAM POUŽITÝCH TABULEK	41
SEZNAM POUŽITÝCH OBRÁZKŮ	41
SEZNAM POUŽITÉ LITERATURY	42

1. Úvod

V posledních letech docházelo a stále dochází k velkému nárůstu v oblasti internetu a informačních technologiích. Internet je zde brán jako celosvětová síť vzájemně propojených počítačů, který tímto umožňuje vzájemné sdílení dat a informací. Mezi nejpoužívanější služby poskytované v rámci internetu patří World Wide Web (WWW). Jedná se stránky vzájemně propojené sítí hypertextových odkazů umožňující uživateli nalézt to co potřebuje. Jednotlivé stránky jsou psané v jazyce HTML nebo XHTML. V dnešní době se můžeme setkat s internetovými stránkami sloužícími k nejrůznějším účelům. Může jít o firemní stránky určené k prezentaci produktů a služeb, blogy (soukromé stránky), e-shopy, multimediální či zpravodajské weby a nebo v poslední době hojně se rozšiřující tzv. sociální weby.

Abychom se v té spoustě internetových stránek, dat a informací vyznali, slouží nám k tomuto účelu internetové vyhledávače. Snad každý uživatel se již s tímto druhem prohledávání obsahu setkal a použil ho. S rostoucí návštěvností vyhledávačů vzrostl také požadavek na přední umístění ve výsledcích hledání, protože přední umístění zaručuje i větší návštěvnost stránek a tím pádem i větší počet potenciálních zákazníků a vyšší příjmy z reklamy. Proto, aby se stránka umístila ve výsledcích vyhledávání na prvních místech musí být dobře optimalizována. Jedná se o tzv. Search Engine Optimization (SEO).

Ve své práci na téma „způsoby optimalizace webových prezentací pro internetové vyhledávače“ bych chtěl co nejlépe popsat a osvětlit danou problematiku, uvést základní metody optimalizace a hlavně vysvětlit jakým způsobem samotná optimalizace pro internetové vyhledávače probíhá.

2. Základní pojmy

2.1 *Historie vyhledávačů*

2.1.1 *Vývoj ve světě*

Historicky první vyhledávač existoval již v prosinci 1990, neprohledával však weby, ale soubory uložené na ftp. Tento vyhledávač se jmenoval Archie a vyvinul ho Alan Emtage z univerzity McGill v Montrealu. Archie dokázal prohledávat názvy souborů na několika ftp serverech najednou.

První robot se objevil až v roce 1993, kdy se začíná rozvíjet World Wide Web. Jmenoval se World Wide Web Wanderer a naprogramoval ho Matthew Gray. Původní zaměření tohoto robota bylo počítat stránky a monitorovat tak rozvoj internetu, později byl rozšířen a ukládal i URL stránek.

První robot WebCrawler, který indexoval celý obsah stránky, vznikl 20. dubna 1994, vymyslel ho Brian Pinkerton z University of Washington. Postupně vznikaly další vyhledávače. Počet internetových stránek velmi rychle rostl a vyhledávače měly problém udržet vysokou relevanci vyhledávání.

Tento problém se v roce 1998 rozhodli vyřešit Larry Page a Sergey Brin a založili vyhledávač Google. Google jako první začal používat Page Rank, což je hodnocení webu na základě počtu a kvality zpětných odkazů. Díky započítání těchto off page faktorů se výrazně zvýšila relevance vyhledávání a popularita Google rychle rostla. V současnosti je Google celosvětově nejpoužívanější vyhledávač. Nejvíce se mu snaží konkurovat Yahoo! a Microsoft.

2.1.2 *Vývoj v tuzemsku*

Historie nejznámějšího českého vyhledávače seznam.cz začíná na jaře roku 1996, kdy ho založil Ivo Lukačovič. Seznam byl ze začátku jen katalog internetových stránek, který se pravděpodobně inspiroval u katalogu Yahoo!. Umožňoval vyhledávat v názvech stránek, jež se nacházely v katalogu, a zájemce o fulltextové hledání přeposílal na fulltext Altavista. První český fulltext Seznam.cz spustil až v roce 1997, kdy ho naprogramoval Ivo Lukačovič a Štěpán Škrob. Tento historicky první český fulltextový vyhledávač se jmenoval Kompas.

Kompas indexoval všechny české weby, které našel, a uměl vyhledávat správně i s diakritikou. [7]

2.2 Druhy vyhledávání

Vyhledávání funguje jako samostatný celek. V dnešním světě je to už jakási součást každodenní práce s počítačem a rychlé orientaci na internetu. S příchodem webových prezentací nastal problém jak jednotlivé stránky řadit, jak se mezi nimi pohybovat.

Způsob vyhledávání stránek na internetu může probíhat ve dvou podobách. Nazývají se fulltext a katalog. Každá z variant má své výhody a nevýhody proto je na většině vyhledávacích portálů možnost zvolit vhodný způsob. I když každý portál je prioritně zaměřen buď na první, nebo druhou variantu. Méně známým způsobem je pak tzv. meta vyhledávač.

2.3 Webová stránka

Ať už se zde bavíme o internetové optimalizaci či vyhledávání, prvořadým cílem každého webového tvůrce by mělo být vytvořit obsahově kvalitní a přínosnou internetovou prezentaci. To znamená, že by se měl nejprve zaměřit na obsah stránek a až později na samotnou optimalizaci. A jak by tedy měla vypadat správná webová stránka? Webových prezentací (stránek) je na internetu nespočet. Každým dnem přitom neustále přibývají nové a nové. Aby se v té záplavě právě ta naše stránka neztratila, musí něčím zaujmout. Mělo by to být samozřejmě v tom dobrém slova smyslu. Platí zde pár doporučení, kterých by se měl člověk při tvorbě držet a brát je v úvahu. Cílem by mělo být vytvořit každou stránku natolik intuitivní, aby průměrný uživatel pouhým pohledem pochopil, co je stránka zač a jak ji má používat.

To znamená, že by měla mimo jiné obsahovat:

- **Logo a poslání serveru.** Hned na začátku by mělo být zřejmé, o jaký server se jedná a k čemu slouží. To znamená, že zcela evidentní by mělo být zaměření a tematický obsah stránek.
- **Aktuální obsah.** Závisí-li úspěch serveru na tom aby se uživatel často vracel, měla by domovská stránka obsahovat oblast, která bude často aktualizována. I server, který nepotřebuje pravidelné návštěvníky, by měl projevovat nějaké známky života.
- **Hierarchie serveru.** Z domovské stránky by mělo být zřejmé, co na serveru mohu „najít“, co na serveru mohu „dělat“. O to by se měla postarat vhodně tvořená navigace.
- **Struktura stránek.** Pomocí hypertextu jsme schopni vytvořit stránky nejrůznější struktury. Vše by ale mělo být jednoduché a na první pohled pro návštěvníka intuitivní. Vhodné je například použít členění podle hierarchie. To znamená vytvoření hlavní stránky s obsahem odkazující na jednotlivé kapitoly a sekce.
- **Validnost moderní technologie** Webové stránky by měly splňovat požadavky a doporučení konsorcia W3C. Zejména se jedná o validní a sémanticky správný kód. Při tvorbě by měly být použité moderní a osvědčené technologie tak, aby výsledná webová stránka odpovídala pravidlům v oblasti prezentace na internetu.
Kontrolu, zda je kód dané webové stránky validní, lze snadno ověřit pomocí online validátoru na adrese <http://validator.w3.org/>.
- **Pěkný grafický návrh (design).** Zajímavý, pěkný a přehledný design je základem pro každou webovou stránku a doplňuje tak její hodnotný obsah. Design zapůsobí na návštěvníka prvním dojmem již při vstupu na stránku. Design celé stránky musí být shodný s firemním grafickým stylem (corporate design) a tvořit tak spolu s ostatními propagačními materiály jednotný celek.
- **Hodnotný obsah.** Web, který nenabídne dostatek informací ve vhodné míře a kvalitě, ztrácí smysl. Všechny informace by měly být srozumitelné a věrohodné. Webová stránka by měla obsahovat zejména kontakt na firmu (včetně všech náležitostí), nabídku služeb či produktů a ceník.

- **Přístupnost a použitelnost** Kvalitní webové stránky by měly být použitelné a přístupné. To znamená, že budou bez omezení k dispozici všem návštěvníkům, potenciálním zákazníkům. Nezáleží tedy na tom, jaký počítač (zařízení) uživatel používá, jestli má nějaká omezení, vždy dostane shodné a úplné informace. Návštěvník by měl stránce plně rozumět, orientovat se v navigaci a celkově se na webu cítit dobře.
- **Podpora různých prohlížečů.** I když se podpora prohlížečů řadí mezi prvky použitelnosti, bylo by vhodné se nad tímto tématem pozastavit. V dnešní době mnohem více uživatelů používá alternativní prohlížeče k prohlížení webových stránek. Za alternativní považujeme všechny ostatní prohlížeče kromě Internet Exploreru. Mezi nejznámější pak patří například Mozilla Firefox nebo Opera. Alternativním prohlížečem může být i prohlížeč textový (např. Lynx). I přes zavedené standardy se webové stránky v různých prohlížečích zobrazují různě. Nejčastěji používaný prohlížeč Internet Explorer mnohé standardy nepodporuje vůbec a některé chybně. Webdesigner je tak velmi omezen a je mnohdy velkým oříškem naprogramovat webovou stránku tak, aby byla korektně zobrazovaná ve všech prohlížečích.
- **Tisknutelná verze.** I v dnešní digitální době je stále mnoho lidí, kteří upřednostňují papír a tužku. Uživatelé si informace z internetu rádi vytisknou a přečtou raději v „papírové formě“, než na obrazovce monitoru. Kvalitní webové stránky umožňují veškerý obsah vytisknout ve vhodné formě nebo odkazují na tisknutelnou verzi, popř. na stránku ve formátu PDF.
- **Funkčnost bez stylů, obrázků a skriptů.** Dnešní webové stránky obsahují kromě vlastního textu i mnoho doplňků: styly, obrázky, skripty a další multimedia. Kvalitní webové stránky musí být použitelné i bez těchto doplňků. Styly určují, jak má být stránka zobrazena.

Styly nemusí být vždy žádoucí (např. hlasová čtečka) a stránka tak musí být bez problému použitelná i bez stylů.

Obrázky nejčastěji slouží jako doplňující grafika, mnohdy však zastupují i tak důležité prvky jako je menu či nadpis. Pokud se obrázek z nějakého důvodu nenačte (uživatel může mít vypnuté načítání obrázků,

obrázek je poškozen, textový prohlížeč), musí být uveden alternativní text, který je zobrazen namísto obrázku.

Velmi oblíbenou součástí webových stránek jsou *flashové animace*. Pokud je použita technologie flash, vždy by měla být k dispozici její textová nebo grafická alternativa. Další nevýhodou je, že pro zobrazení této animace musí být nainstalovaný příslušný plugin – zásuvný model pro flash animace. Ne všechny počítače tímto pluginem disponují a někdy jsou flash animace dokonce zakázány.

Mezi skripty se řadí například *JavaScript*. Skripty by měly uživateli pobyt na webových stránkách zpříjemnit. Používají se především pro kontrolu formuláře před odesláním, počítání zadaných znaků, nebo pro práci s časem. Skriptování bývá mnohdy z bezpečnostních důvodů zakázáno a proto by stránky měly být použitelné i bez skriptů.

- **Rychlé načítání.** Doba, za kterou se požadovaná webová stránka otevře a zobrazí, je velmi důležitá. Stránky, zejména použitá grafika, musí být optimalizována tak, aby načítání bylo co možná nejrychlejší. Doba, za kterou se návštěvníkovi stránka zobrazí, záleží na rychlosti připojení k internetu. Nezapomínejme však na to, že i dnes je stále spousta uživatelů připojená přes telefonní linku či moderněji přes mobilní telefon a otevření velké, graficky náročné, stránky může být záležitostí i několika minut. [3]

2.4 Typy webových stránek

2.4.1 Statické stránky

Podle typu rozdělujeme webové stránky na dva druhy- statické a dynamické. Statické stránky obsahují stále stejný obsah, který se pro návštěvníka nijak nemění. Tento způsob zveřejňování informací má své výhody i nevýhody.

Výhody:

- Stránky mají jednodušší strukturu
- Menší náklady na vytvoření a údržbu webových prezentací
- Jsou vhodné pro malé firmy a jako osobní stránky
- Stačí ovládnutí základního kódovacího jazyka HTML

Nevýhody:

- Omezená interaktivita webové prezentace
- Nemožnost zjišťování podrobných statistik o uživateli

2.4.2 Dynamické stránky

Na rozdíl od statických stránek se dynamické stránky mění svůj obsah v čase. Základem je programovací technologie, a data uložená v databázi. Program umístěný na webovém serveru, pak na základě požadavku na zobrazení konkrétní stránky webovou stránku dynamicky sestavuje a odesílá do prohlížeče uživatele. Zjednodušeně se dá říci, že stránka je sestavena na základě "šablony" určující vzhled a dat získaných z databáze (data jsou například seznam zboží, článků nebo konkrétní článek). Sestavení zajišťuje program běžící například na technologii ASP nebo ASP.NET. Dynamické stránky tedy fyzicky neexistují v uloženém stavu, na rozdíl od stránek statických. [4] Dynamické stránky jsou v dnešní době velice používané. Každý se s nimi na internetu již setkal. Může se jednat například o ankety, webové diskuze a dalších různých dynamických prvků. Nevýhodou pro vytvoření takovýchto stránek může být nutnost znalosti některých z pokročilých programovacích jazyků jako je java script, xml, dhtml, css, php atd.

2.5 Druhy vyhledávačů

2.5.1 Katalogové vyhledávače

V katalogovém vyhledávači jsou stránky řazeny do jednotlivých kategorií podle svého zaměření, např. volný čas, sport, zábava, kultura atd.(Obr. 1.). Do těchto skupin můžeme vytvořenou stránku zaregistrovat, avšak musí projít procesem schválení příslušným editorem. Až poté je stránka do databáze zařazena. V katalogu je možné listovat podle kategorií a nebo vyhledávat pod příslušným klíčovým slovem. Výhodou katalogového vyhledávání je že probíhá ručním zařizováním, stránka je tedy správně zařazena. Znamená to že, při procházení kategorií nalezneme stránky, které do nich skutečně patří. Nevýhodou je naopak to, že při ručním zařazování je zde možné umístit jen celé „weby“, tedy nikoli jednotlivé stránky a podstránky.

Jak je z výše popsaného způsobu zařazování stránek patrné, je celý proces schvalování časově náročnější. Proto se doporučuje do katalogu zařazovat webové stránky u kterých není požadavek na jejich okamžitý přístup ke čtenáři jako je například zpravodajství, aktuální události a pod. Naopak vhodné je přidávat informace které jsou platné z delšího časového hlediska. To může být například katalog firem, e-shopů, knih, zájmových kroužků, veřejných institucí atd.



Obr. 1. Setřizovaný katalog. Zdroj: <http://www.firmy.cz/>

České katalogové vyhledávače	Zahraniční katalogové vyhledávače
http://www.firmy.cz/	http://www.allesklar.de/
http://najisto.centrum.cz/	http://www.zoznam.sk/
http://www.cent.cz/	http://www.yahoo.com/
http://www.caramba.cz/	http://directory.google.com/
http://www.hledat.cz/	http://web.de/

Tab 1. Příklady českých a zahraničních katalogových vyhledávačů

2.5.2 Fulltextové vyhledávače

Jak už název sám napovídá, jedná se o vyhledávání se zaměřením na samotný text (z anglického full – plný, celý a text). Cílem fulltextového vyhledávače je najít na internetu nebo konkrétních stránkách určité slovo nebo slovní spojení. Když si ale představíme to obrovské množství informací a textu který by vyhledávač musel v aktuálním čase projít, vyhledávání by trvalo neúměrně dlouho. Proto se textová data z webových stránek předem indexují a umísťují do databází, ve kterých už je pro vyhledávací roboty snadné najít během několika desetin sekundy požadovaný výsledek na vyhledávaný výraz.

České fulltextové vyhledávače	Zahraniční fulltextové vyhledávače
http://seznam.cz/	http://www.google.com/
http://www.centrum.cz/	http://www.altavista.com/
http://www.zoofoo.cz/	http://search.yahoo.com/
http://morfeo.centrum.cz/	http://www.alltheweb.com/
http://jyxo.cz/	http://www.wp.pl/

Tab 2. Příklady českých a zahraničních fulltextových vyhledávačů

Jak už ale bývá pravidlem většinou nic není takové, jak se na první pohled zdá. Proto některé větší vyhledávací portály nabízí vyhledávání oběma možnými způsoby. A je jednomu uživateli jestli použije fulltextové vyhledávání k prohledání internetu a nebo raději zabrousí do seřazeného katalogu a vybere si z nabízeného seznamu.

2.5.3 META vyhledávače

Posední skupinou jsou méně známé meta vyhledávače. Ty si neudržují vlastní databázi webových stránek, ale spoléhají se na jiné vyhledávače. Myšlenka metavyhledávače (metasearch, metasercher) je jednoduchá: když uživatel zadá dotaz do vyhledávacího políčka, metavyhledávač jej mnohonásobně „zkopíruje“ a rozešle na jiné vyhledávače – na několik málo nebo i třeba na několik desítek nebo stovek, podle toho jak je naprogramován; položí daný dotaz za vás. Zmíněné vyhledávače odpoví, zobrazí své stránky s výsledky. Metavyhledávač tyto stránky s výsledky sebere a předloží vám je. Pokud je chytrý, tak ještě dokáže utřídit, seřadit, vyházet z nich duplicitu a předložit vám je v příjemné podobě. Na některých portálech si uživatel dokonce může vybrat (nebo omezit okruh) mezi jaké vyhledávače bude jeho dotaz rozeslán.

Metavyhledávačů je na světě hodně a i u nás je jich několik, průběžně nové vznikají a staré zanikají. Mezi provozovateli velkých portálů a fulltextů nebývají příliš oblíbené a některé je dokonce i blokují (tj. blokují dotazy od metavyhledávačů), protože se vlastně chlubí cizím peřím. Velké portály je skoro nikdy neobsahují a proto jsou metavyhledávače poměrně málo známe; určitě ale stojí za to o nich vědět a občas je využít, už proto, že vám mohou ukázat na další kvalitní vyhledávací zdroje, třeba i úzce specializované, o kterých nevíte a jež vám mohou být ku prospěchu. Hlavní nevýhodou metavyhledávačů ale je, že je vlastní vyhledávací servery dokáží snadno a spolehlivě blokovat, a tak také činí. Dobrým příkladem je třeba Google, který blokuje skoro všechny metavyhledávače. Vhodnou alternativou může být instalace speciálního programu který se bude tvářit jako když požadavek serveru odešle uživatel a poslušně vrátí všechny výsledky. Takto například funguje program copernik. [1]

Příklady metavyhledávačů:

<http://www.metacrawler.com/>

<http://www.mamma.com/>

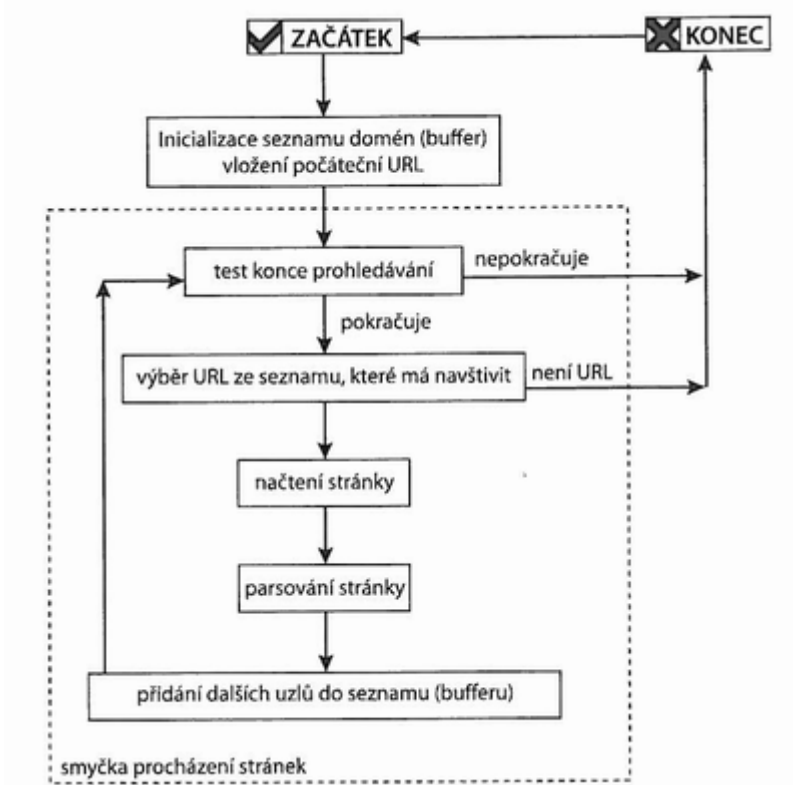
<http://dogpile.com/>

<http://www.search.com/>

<http://www.polymeta.com/>

2.6 Spider

Technologie fulltextového vyhledávání je založena na třech částech. První je program, který prochází webové stránky – říká se mu spider (pavouk) nebo trawler (slídlil). Pavoukova cesta začíná ve vlastní databázi. Zde si vybere odkaz na stránky, kterou následně navštíví. Přijde na hlavní stránku, tu si přečte a její obsah si stáhne do skladiště na vlastním serveru. Ve zdrojovém kódu hledá odkazy ukryté v párových značkách `<a>` a ``. Jakmile na tento odkaz přijde, načte novou stránku a toto několikrát opakuje. Jak je vidět na schématu, pokud odkazy na další stránky nenajde, vrátí se o úroveň zpět a celý postup pokračuje. Protože obsahuje rekurzivní algoritmus, a tak může volat sám sebe, reaguje na změny webu okamžitě a nikdy není jeho cesta stejná. Cestu pavouka znázorňuje schematický obrázek. (Obr 2.)



Obr 2. Schematická cesta pavouka

Jsou ovšem různé druhy pavouků a každý může mít jiný úkol a činnost:

- Stahovači, kteří vytvářejí výtahy z dokumentů, které se ukládají do databáze. Vytvářejí indexy, které stanoví vztah slov a frází nalezených v dokumentech HTML k adrese URL.
- Prohlížeči si prohlížejí okolí a názvy obrázků, ty potom stahují opět v náhledech do databáze společně s odkazem na stránku, kde jsou umístěny. Jejich funkce je podobná jako u stahovačů, jen s tím rozdílem že se soustředí na obrázky. Podobně webem procházejí pavouci specializující se na další typy dokumentů, jako jsou textové formáty DOC, XLS, PDF a další.
- Kontroloři zase mohou hledat odkazy na již neexistující stránky. Informaci o neexistujících stránkách pak předávají do databáze a většinou, pokud robot stránku nenalezne opakovaně, dává pokyn k jejímu vyřazení z hlavní databáze.
- Statisticy shromažďují údaje o počtech odkazů a využívají jednotlivých stránek, podle čehož poté určují, které stránky jsou jak oblíbené.
- Počtáři počítají množství stránek, čímž zajišťují jejich nárůst. [2]

2.7 Indexer

Druhá část vyhledávače se nazývá indexér (indexer) a je to právě ten program jež zpracovává postupně všechny stránky, jež mu snese slídlil. Jakmile se informace o stránce dostane do hlavní databáze, na místo vyrazí robot, který má na starosti načítání stránek. Denně tak robot zpracuje až dva miliony stránek. Robot odešle dotaz HTTP a přečte a zpracuje odpověď. Zajímá ho:

- Hlavička dokumentu, kde zpracuje informace v metaznačkách
- Redirekce, přesměrování
- Stavové kódy
- Informace o poslední změně na stránce
- Čas za který stránku zpracuje

Index je již utříděná, uspořádaná kniha stránek. Je to vlastně jakýsi automaticky vytvořený katalog. Cílem třídění je aby následně vyhledávač nemusel při vyhledávání výsledku již znovu a znovu pracně procházet tisíce stránek a miliony slov jedno podruhé. Indexování je velice časově a výpočetně náročná činnost; proto se stává, že stránku kterou sebere slídil, zpracuje indexér i se značným zpožděním. [1]

2.8 Vyhledávač

Poslední částí je již samotný vyhledávač (search engine). Je to vlastně jakýsi prostředník mezi uživatelem a nepřehledným množstvím internetových stránek který se snaží podat uživateli odpovídající výsledky na jeho dotaz.

Je to program, který přebírá od uživatele dotaz na vyhledávání. Je to vlastně jediný program ze zde popsaných pracujících v reálném čase, jehož odezva a vrácení výsledků musí být co nejrychlejší, neboť uživatel čeká na výsledek. Jak je ale možné že vyhledávač nám vrátí výsledek na námi zadaný dotaz během několika milisekund? Je možné, aby během této doby prohledal „celý internet“ Jak již určitě správně tušíte, není to reálné ani to není v možnostech vyhledávacího algoritmu. Celý proces je mnohem jednodušší než se na první pohled může zdát. Vyhledávač sáhne do databáze zaindexovaných slov a nalezne stránky, které odpovídají dotazu. Tím ale úkol nekončí, neboť je musí seřadit podle jejich relevantnosti a pak teprve zobrazí výčet stránek uživateli.

Portály v jednotlivých státech většinou prohledávají pouze internetové stránky v dané státní (národní) doméně, jen několik největších světových portálů usiluje o prohledání v celosvětovém měřítku. Není však pravda, že by nám k prohledávání celosvětového internetu vždy vystačil jen například hodně univerzální Google.com. Lokální vyhledávače se soustředí na menší území a často používají některé speciální, místně vhodné a „výnosnější“ techniky. Některé jsou zase přizpůsobené na místní podmínky tím, že lépe vyhledávají v místní znakové řeči. Příkladem může být čínský <http://www.baidu.com/>.

2.8.1 Podíl vyhledávačů na českém trhu

Na většině zahraničních trhů opanovali internetové vyhledávání giganti jako je Google Yahoo případně Bing. U nás si však Seznam stále drží své prvenství v počtu příchozích návštěvníků. (Údaje z roku 2009). Jak je vidět z Tab. 3 měly jednotlivé vyhledávače v průběhu roku vyrovnanou tendenci. Do přehledu je zařazen i vyhledávač Bing.com společnosti Microsoft, který v červnu nahradil její dřívější produkt Live Search.

Vyhledávač	Únor [%]	Červenec [%]	Prosinec [%]
Seznam.cz	55,88	50,92	53,91
Google (.cz,.com)	31,54	32,22	32,85
Zboží.cz	3,43	3,63	4,95
Firmy.cz	4,01	3,33	3,52
Centrum.cz	2,57	2,35	2,50
Bing.com	0,49	5,79	0,48

Tab. 3. Podíl vyhledávačů na českém trhu. Zdroj: NAVRCHOLU.cz, 2009

3. Optimalizace a její metody

3.1 Optimalizace obecně

Snem každého webového vývojáře je, aby jeho webová prezentace vydělávala. Náklady na vytvoření a samotnou údržbu nejsou zanedbatelné a proto je zde snaha o přilákání co největšího počtu lidí na naši internetovou stránku. Čím více je naše stránka navštěvována, tím se i zvyšuje samotný zisk z internetové reklamy umístěných na stránkách nebo produktů které nabízíte. Prvořadým cílem je tedy vytvořit internetovou prezentaci s vysokou návštěvností ať už se jedná o pravidelné nebo unikátní návštěvníky. Jak ale takové návštěvnosti dosáhnout? Když pomineme reklamu a jiné marketingové propagační metody je nejvhodnějším způsobem jak takové návštěvníky oslovit dobrá pozice naší stránky ve vyhledávačích. Tohoto výsledku se dá dosáhnout jedině dobrou a kvalitní optimalizací webové stránky (SEO). O tom a o možnostech optimalizace si povíme v následcích odstavcích.

3.1.1 Klíčové faktory pomáhající ve výsledcích vyhledávání

- Klíčové slovo použité v metaznačce Title
- Celková popularita stránky vyjádřená vysokým počtem stránek odkazujících na web
- Text odkazů, které na web odkazují (anchor text)
- Váha stránky podle vyhledávače (rank)
- Stáří stránek – starší stránky jsou považovány za relevantnější než nové
- Klíčová slova na stránce
- Hodnota odkazujících stránek – čím je stránka, která na náš web odkazuje, lépe hodnocena vyhledávači, tím je odkaz pro web hodnotnější.
- Tempo nárůstu nových příchozích odkazů

3.1.2 Nežádoucí faktory

- Podobnost nebo totožnost s obsahem jiných webových stránek
- Stránky jsou často nepřístupné pro roboty (a návštěvníky)

- Příchozí odkazy nemají velkou hodnotu, přichází ze spamových stránek

3.2 *Optimalizace rychlosti*

V dnešní době, hlavně z důvodů rozšíření vysokorychlostního připojení a konektivity není tento údaj až tolik klíčovým parametrem. Na druhou stranu ovšem narůstá množství přenášených dat z důvodů velkého množství multimédií, ale mění se také chování uživatelů.

Rychlost načítání webových stránek uživateli je čím dál tím víc důležitá i pro vyhledávače. Jednou z novinek které Google v posledních dnech zveřejnil (19.4.2010) je, že již zmíněnou rychlost načítání stránek promítl jako další kritérium řazení výsledů vyhledání. Průzkumy totiž prokázaly, že lidé tráví podstatně méně času na stránkách s delší reakční dobou než na těch, které fungují rychle. Zatím však kritérium rychlosti používá Google jen zhruba v jednom procentu ze všech vyhledávání.

Velmi pomalé stránky nemusí být vyhledávačem indexovány i přesto, že jste provedli náročnou SEO optimalizaci. Roboti vyhledávačů totiž mají stanoveny limity pro načtení jedné stránky. Po překročení tohoto limitu není www stránka vyhledávačem indexována a tudíž se ani nemůže ve výsledcích vyhledávání objevit.

3.2.1 *Optimalizace na straně web serveru*

Prvním optimalizačním krokem je povolit kompresi obsahu odesílaného webserverem. Tím se několikanásobně zmenší objem přenášených dat. Následně jsou veškeré textové soubory (HTML, CSS, JavaScript) serverem komprimovány (mod_deflate) čímž se rapidně snižuje množství přenášených dat.

Nastavení/aktivaci komprese lze provést v konfiguračním souboru webserveru nebo pomocí souboru .htaccess (Apache).

```
// compress all text & html:  
AddOutputFilterByType DEFLATE text/html text/plain text/xml  
  
// Or, compress certain file types by extension:  
<Files *.html>  
SetOutputFilter DEFLATE  
</Files>
```

Pokud nemůžete měnit konfiguraci webserveru přímo (pravděpodobně na hostingu) a ani měnit obsah souboru `.htaccess`, ale používáte PHP, můžete kompresi zapnout přidáním jednoho řádku na začátek PHP skriptu

```
<?php
  if (substr_count($_SERVER['HTTP_ACCEPT_ENCODING'], 'gzip'))
    ob_start("ob_gzhandler"); else ob_start();
?>
```

3.2.2 Optimalizace kódu stránek a rychlosti zobrazení obsahu

Základem optimalizace na straně kódu stránek je jednak minimalizace kódu samotného a dále pak minimalizace požadavků na objekty jakými jsou obrázky a různé externí skripty (JavaScript, CSS apod.).

Pro obrázky platí:

- Používat vždy obrázky s korektní velikostí, tedy takových rozměrů, jaké se na stránce mají zobrazit.
- Nepoužívat zmenšování pomocí atributů `width` a `height`
- Vhodné je také přesunutí veškerých obrázků na samostatnou doménu, ze které se neposílají cookies.

Pro externí skripty platí:

- Odstranění zbytečných definic z CSS, odstranění zbytečných mezer, odřádkování a odsazení z JavaScript (Minify JS), HTML a CSS souborů. Mezery, tabulátory a jiné formátovací znaky usnadňují tvůrcům stránek orientaci v kódu, nicméně pro prohlížeč jsou k ničemu. Pouze zpomalují načítání stránky a zvyšují nároky na procesor při parsování dokumentu
- Omezit nebo zrušit zobrazování neefektivních reklam (jakékoliv PPC, Affiliate apod.) [6]

3.2.3 Optimalizace prohlížeče

Další možností zrychlení načítání stránek je optimalizace rychlosti zobrazení obsahu na straně klienta, tedy prohlížeče. Touto částí se nebudeme zabývat, jelikož nemá na SEO optimalizaci vliv. Doporučuji však mít svůj prohlížeč pravidelně aktualizovaný.

3.3 On-page faktory

Mezi tzv. On-page faktory můžeme zařadit všechno, co je umístěno na příslušné stránce u které provádíme optimalizaci. Vhodné vyladění on-page faktorů můžeme velice pozitivně ovlivnit hodnocení stránky vyhledávači a tím i jeho výhodnější umístění ve výsledcích vyhledávání. Řadíme sem tedy vše, počínaje textem na stránce, přes nadpisy, titulku stránky až k optimalizovaným obrázkům a hypertextovým odkazům. Je nutné připomenout, že optimalizace se týká každé stránky, ne jenom té úvodní.

3.3.1 URL

Obecně se traduje, že většina vyhledávačů přikládá klíčovému slovu v URL velký význam. Je tedy dobré ho (při On Page optimalizaci) v URL mít. To však záleží případ od případu, podle toho jaké je ono klíčové slovo. Pokud se slova v URL shodují s hledanou frází, tak jsou v SERP zvýrazněna tučně a zvyšují tak míru prokliku. Klíčové slovo obsažené přímo v názvu domény má vyšší váhu než klíčové slovo v URL, které se nachází za doménovým jménem. Jednotlivá slova se v URL oddělují pomocí pomlčky.

3.3.2 Klíčová slova (keywords)

Je více než nutné mít na stránce (On Page) dostatek klíčových slov. Pokud tomu tak není, nemůže dojít k optimalizaci (ani indexaci) klíčového slova na stránce, čímž okamžitě ztrácí optimalizace svůj význam. Robot vyhledávače sice stránku zaindexuje, ale prakticky nikdy ji vyhledávač nezobrazí ve svých výsledcích vyhledávání, jelikož neexistuje slovo ani slovní spojení, pro které je stránka charakteristická. Volba vhodných klíčových slov je velice důležitá, neboť to je půlka cesty k úspěchu. Vezměme si malý příklad. Budeme-li například

optimalizovat stránky pro Autobazar. Klíčové slovo „autobazar“ je sice velice populární, ale je zde i velká konkurence, proto je takové klíčové slovo velice neefektivní. Aby uživatel hledající náš autobazar, se skutečně dostal na naše stránky, musíme klíčové slovo – slovní spojení konkretizovat. V tomto případě například Autobazar Praha Vinohrady. Zúží se tím i okruh nalezených stránek. Viz Tab. 4. Klíčem k úspěchu je tedy nalezení správné rovnováhy mezi přesností a popularitou. S výběrem vhodného klíčového slova nám také může pomoci webová služba „Nástroj pro návrh klíčových slov“ který se nachází na adrese <https://adwords.google.com/select/KeywordToolExternal>.

	Google.cz	Seznam.cz	Atlas.cz
Autobazar	5 890 000	5 817 201	732 005
Autobazar Praha	871 000	1 038 823	125 006
Autobazar Praha Vinohrady	14 600	8 803	4 036

Tab. 4. Počet nalezených stránek na klíčová slova a slovní spojení

Co by nemělo mezi klíčovými slovy chybět:

- Název organizace
- Název stránek (serveru)
- Název produktu (čím se zabývají stránky)
- Odborné označení nebo slangový výraz
- Slova nejlépe vystihující naše podnikání/služby/výrobky
- Víceslovné výrazy, varianty bez háčků a čárek, slova podobná a synonyma, hovorové výrazy, užívané zkratky.

Umístění klíčivých slova musí být na stránce tam, aby byl jejich výskyt co nejefektivnější. To znamená že by měly být na stránce v přirozeném množství, poměru a na přirozených místech.

Hustota klíčových slov

Klíčová slova by se měla vyskytovat v textu stránky v míře větší než žádné a menší než nadměrné. Výskyt klíčových slov v textu by měl být přirozený a neměl by být na úkor

dobré čitelnosti textu. Optimální hustota výskytu těchto slov v textu stránky není přesně stanovena. Doporučuji řídit se rozumným odhadem, několikrát klíčové slovo či frázi v textu hezky použít a nepřehánět to. Nadměrný výskyt klíčových slov v textu stránky může mít totiž negativní dopad na výsledky v SERP. [9]

Prominence klíčových slov

Prominence znamená v tomto ohledu jak vysoko se klíčová slova zobrazují na stránce - přeneseně řečeno, je to výsledné "těžiště" těchto klíčových slov. Důležitost prominence má opět své opodstatnění. Když uživatel přijde na stránku, čte text od shora dolů. Čili co je napsáno dole, k tomu už se ve většině případů ani nedostane. Čím je klíčové slovo níže, s tím menší pravděpodobností si ho uživatel všimne a tedy i vyhledávač mu přisoudí nižší váhu. Prominence se uvádí v %.

3.3.3 Stop slova

Stop slova (v angličtině stop words) jsou pojmy, které nenesou samy o sobě žádný význam. V češtině jde především o předložky, spojky a některá další slova (a, i, nebo, ale, na atd.). V optimalizaci pro vyhledávače získala stop slova zvláštní postavení díky výpočtu hustoty klíčových slov. Vzhledem k tomu, že sama žádný význam nenesou, působí jako výplň, kterou je dobré vypustit z titulku a URL stránky.

3.3.4 Title

Title, patří mezi párové tagy a je umístěn v části <HEAD>. Čiji jde o jakýsi nadpis stránky. Z pohledu SEO se jedná u jeden z nejdůležitějších prvků na stránce a musíme věnovat velkou pozornost jeho vyplňování. Mnohé vyhledávače, v čele s Google, zobrazují jako titulek vyhledaného spojení právě slova z titulku stránky. Značka title zůstává z hlediska vyhledávání nejvýznamnějším on-page faktorem. Dobře zvolený titulek stránky korespondující jak s obsahem, tak s tím, co lidé hledají, může stránce ve výsledcích vyhledávání velmi pomoci ale stejně tak při nevhodném využití mlže stránku poškodit. [2] Musí být výstižný. "Další stránka" je špatně. Měl by být výstižný i z globálního hlediska. Čtenář může titulek používat i pro pochopení struktury serveru. Naopak

když budou mít všechny stránky serveru titulek stejný (jako hlavní stránka), tak čtenář o tuhle možnost orientace přijde.

Příklad jak může vypadat titulek stránky:

```
<head>
<title>SEO služby </title>
</head>
```

V titulku se zobrazí název stránky „SEO služby“

```
<head>
<title>SEO služby - konzultant SEO, SEO Expert, SEO
konzultant </title>
</head>
```

Nebo můžeme použít dokonce seznam vět. Doporučuje se však do titulku psát jen stručné a výstižné názvy. Počet znaků by neměl přesáhnout 70.

3.3.5 Nadpisy

Nadpis, je párový tag vyskytující se kdekoliv v textu. Rozeznáváme šest úrovní <h1> až <h6> To, co je vyznačeno nadpisem, je nějakým způsobem nadhodnoceno oproti normálnímu textu. Nejvyšší váhu má tag h1, ostatní potom postupně váhu nižší. Při On Page optimalizaci, mohou mít nadpisy celkem velký význam. Platí pravidlo, že by se nadpisy měly na stránce vyskytovat v „přiměřeném" množství. V opačném případě by to stránce spíše mohlo uškodit.

3.3.6 Meta tagy

Meta Description

Jedná se o tag který je opět umístěn v části <HEAD>. Meta tag Description slouží pro krátký popis obsahu stránky. Podstatné je, že některé vyhledávače tento tag používají pro zobrazení popisku stránky v SERP. Díky tomu může mít Description poměrně velký vliv na míru prokliku. Uživatelé raději kliknou na odkaz, který má u sebe zobrazený autorem pečlivě napsaný text, než na odkaz s útržky nekompletních vět.

Z vyhledávačů, které meta Description podporují, je nejdůležitější Google. Už jen kvůli němu se vyplatí tímto tagem zabývat. Ze zahraničních vyhledávačů dále stojí za zmínku ještě MSN, české vyhledávače Description nepoužívají. [10]

Příklad popisku:

```
<meta name="description" content="Popis SEO optimalizace stránek pro začátečníky i profesionály.">
```

Meta Keywords

Dalším příkladem meta značky je Keywords. Keywords má význam pro interní vyhledávače (např. [Atomz](#)), které obsah zahrnou do svého indexu. Naopak většina globálních vyhledávačů tuto metaznačku pravděpodobně ignoruje. Není ale na škodu ho uvést.

Celý zápis pak může vypadat například takto:

```
<HEAD>
<TITLE>Elektro Dvorak, s.r.o.</TITLE>
<META NAME="description"
      CONTENT="Elektro dvorak - firma zabývající
              se prodejem domácích spotřebičů
              a elektroniky">
<META NAME="keywords"
      CONTENT="Elektro Dvorak, prodej, elektrické
              spotřebiče, elektronika, leasing">
</HEAD>
```

Meta robots

Opačný problém nastává, když obsah stránky nechceme indexovat. Vhodným řešením může být použitím meta tagu robots. Meta tag v hlavičce HTML stránky umožňuje zakázat robotům jednak indexování obsahu, jednak sledování odkazů. Důvody pro takové zákazy jsou dosti řídké, leda snad má smysl zakázat indexování vnitřku některých rámců (typicky stránka s menu: noindex, follow).

Meta robots může nabývat hodnot:

noindex	Obsah stránky nebude indexován
index	Obsah stránky bude indexován (normální hodnota)
nofollow	Odkazy nebudou sledovány
follow	Odkazy budou sledovány (normál)
all	Vše povoleno, tedy jako index, follow

Zápis může vypadat takto:

```
<meta name="robots" content="noindex, nofollow">
```

Neuveďte-li se na začátku dokumentu meta tag robots, je celá stránka standardně indexována. Meta robots se musí psát do hlavičky každé stránky znovu. Pro hromadné zakázky, tj. zakázání indexace u více stránek najednou, se využívá robots.txt.

3.3.7 Robots.txt

Jak už bylo psáno v předchozím postu, můžeme zakázat indexování jednotlivých stránek. Pokud však, ať už z jakéhokoliv hlediska, chceme náš web skrýt před vyhledávači a indexováním, lze to tak učinit pomocí jednoduchého textového souboru robots.txt. Roboti, kteří přijdou prohledávat náš server, se nejprve podívají do souboru robots.txt, jestli není jejich pohyb omezen nebo zakázán. Až teprve pak se může vydat indexovat všechny stránky. Soubor robots.txt musí být umístěn v kořenovém adresáři webu a jeho název musí být jen malými písmeny. [11]

- <http://www.dvorak.cz/robots.txt>
- <http://elektro.dvorak.cz/robots.txt>

Příklady užití robots.txt

```
User-agent: *  
Disallow:
```

Tento zápis znamená, že všichni roboti mohou indexovat na celém webu

```
User-agent: *  
Disallow: /php/priklady/
```

*User-agent: * = platí pro všechny roboty*

Disallow: /php/priklady/ = znamená, že se nebude indexovat pouze tento adresář

```
User-agent: Googlebot, Seznambot  
Disallow: /
```

User-agent: Googlebot, Seznambot = příkaz platí jen pro Seznambot a Googlebot

Disallow: / = znamená, že se nic nebude indexovat.

3.4 Off-page faktory

Tzv. off-page faktory jsou ty , které se na stránce přímo nevyskytují a autor stránky není většinou schopen tyto faktory přímo ovlivnit. Mají však přímý vliv na relevanci stránky při hledání určitého dotazu. Mezi Off Page faktory patří vše, co není přímo obsaženo na daném webu. Patří sem zejména odkazy směřující na náš web z jiných internetových stránek.

3.4.1 PageRank

PageRank je algoritmus pro ohodnocení důležitosti webových stránek, navržený Larry Pagem a Sergeyem Brinem, tvořící základ vyhledávače Google. Je to vlastně číslo, které si Google přiřazuje ke každé stránce (přesněji řečeno ke každé URL). Vyjadřuje něco jako věrohodnost nebo důležitost stránky. Je to jeden z důležitých faktorů hodnocení dané stránky a jeho následné vyšší či nižší pozici při vyhledávání. Celý algoritmus je založen na zpětných odkazech. To znamená ty, které vedou na naši stránku a ty které z ní odchází. Důležitý ale není jenom počet těchto zpětných odkazů, v úvahu se také bere z jakých stránek tyto odkazy přicházejí a jak hodnotné jsou tyto stránky z hlediska PageRanku.

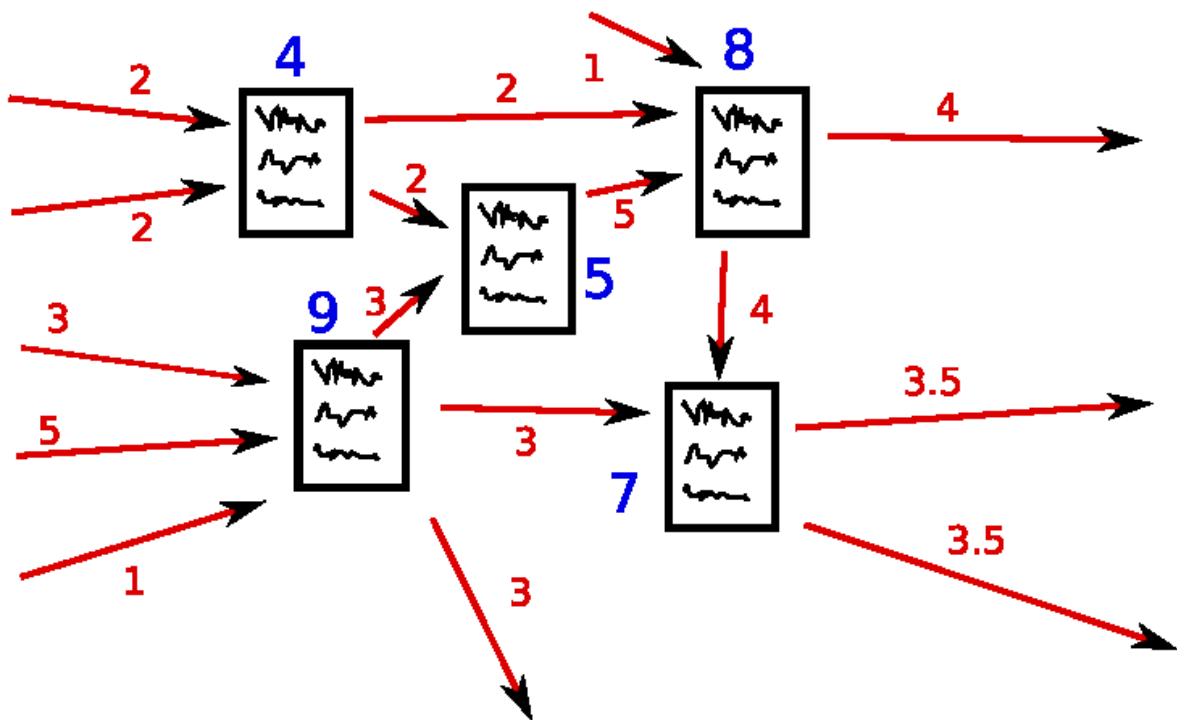
Výpočet page ranku

PageRank stránky A označíme jako $PR(A)$. Vypočítá se z PageRanků stránek, které na ni odkazují. To jsou stránky T_1 až T_n .

$$PR(A) = (1-d)/m + d * (PR(T_1)/C(T_1) + \dots + PR(T_n)/C(T_n))$$

Kde d je damping faktor (nastavený pravděpodobně na 0,85), m je celkový počet zaindexovaných stránek a $C(T)$ je počet odkazů vedoucích ze stránky T . Jako vstupní hodnoty $PR(T_i)$ se berou hodnoty PageRanku stránek z minulé iterace výpočtu. Vzoreček po několika iteracích dobře konverguje (tím lépe, čím je nižší d). Hodnoty PageRanku všech stránek se pohybují těsně nad nulou.

Vzorec se dá zjednodušeně přetlumočit tak, že stránka předává část svého PageRanku stránkám, na které odkazuje. Čím víc obsahuje odkazů (hodnota C), tím méně každé stránce předá. Čím méně má stránka odkazů, tím víc PageRanku se každým odkazem přeposílá. Viz. Obr. 3. Čím má stránka vyšší PageRank, tím bude pravděpodobně výše ve výsledcích. PageRank zdaleka není jediné kritérium pro nalezení stránky v Google. [12]



Obr. 3. Distribuce PageRanku mezi provázanými stránkami.

Zdroj: <http://cs.wikipedia.org/wiki/PageRank>

Zvýšení PageRanku svým stránkám

Zjednodušený PageRank stránky si můžeme zjistit sami a to pomocí jednoduchého Google Toolbaru. Ten nám ukáže jaká je přibližná hodnota našich stránek na stupnici od jedné do deseti (deset je nejlepší výsledek). Skutečný PageRank si však Google nechává pro sebe a nezveřejňuje ho. Hlavní snahou, jak si na svých stránkách zvýšit PageRank, by mělo být to, aby na náš web odkazovalo co nejvíce stránek který mají vysoký svůj vlastní PageRank. Zpětné odkazy ale není lehké získat, zvláště když naše stránky jsou poměrně nové. Nejsnáze se základní zpětné odkazy získávají v katalogích. Stačí si vyhlédnout tematickou stránku nějakého katalogu a zaregistrovat do ní ručně URL a popis stránky. Čím více registrací v různých katalogích, tím lépe. Jakmile ale nemá vyhlédnutá stránka katalogu na toolbaru PageRanku alespoň 3, je otázka, zda má cenu mít na takové stránce odkaz.

Dobrý nápad je oslovovat spřízněné weby se žádostí o výměnu odkazů. Čím méně je na zdrojové stránce odkazů, tím je odkaz hodnotnější.

S-Rank

PageRank ale není jediný hodnotící algoritmus zpětných odkazů. Existují samozřejmě i další zahraniční ukazatele. U nás má však již mnohokrát zmiňovaný PageRank obdobu a tím je S-Rank od domácího vyhledávače Seznam.cz. Pracuje na podobném způsobu hodnocení stránek podle zpětných odkazů. Jeho zjednodušená verze hodnotí internetové stránky známkami na stupnici od nuly do sta (sto je nejlepší výsledek). Pro názornost si zde ukážeme hodnocení těchto dvou algoritmů na některých stránkách. Tab. 5

	PageRank	S-Rank
http://www.idnes.cz/	7/10	100/100
http://www.czechcomputer.cz/	5/10	90/100
http://www.czu.cz	7/10	80/100
http://tf.czu.cz	5/10	50/100
http://www.youtube.com/	9/10	100/100
http://www.kolin.cz/	4/10	70/100
http://www.lupa.cz/	7/10	90/100

Tab. 5 Porovnání hodnocení stránek z pohledu PageRanku a S-Ranku

3.4.2 Anchor text

Termínem anchor text rozumíme viditelnou část odkazu, který vede na jinou webovou stránku. Je to vlastně část textu, kterou vidíme z hypertextového odkazu. I když se to na první pohled třeba nezdá, z pohledu Off-Page faktorů je to poměrně důležité. Je tedy podstatné, jaké slovo se v tomto odkazu použije. Je vhodné používat slova která mají co dočinění s obsahem webu na který se odkazují. Zvýší se tím pádem i váha klíčových slov na odkazované stránce.

Nevhodné použití:

```
<a href="www.jak-na-web.cz">Zde</a>
```

Vhodné použití

```
<a href="www.jak-na-web.cz">tvorba webu</a>
```

Klíčové slovo „tvorba webu“ bude určitě vyhledávanější a i pro optimalizaci vhodnější. Kdežto klíčové slovo „zde“ nikomu nic neřekne a nezabodujete s ním ani u vyhledávače.

3.4.3 Zpětné odkazy

Zpětné odkazy patří mezi jedny z klíčových vlastností úspěšného webu. Ve chvíli kdy máme vytvořené stránky v souladu s pravidly přístupnosti, naplněné hodnotným obsahem a informacemi, jsem se dostali k další důležité části v off-page optimalizaci. Při budování zpětných odkazů je důležitá kvalita stránky, z níž je odkazováno. Když na stránky bude odkazovat 10 stránek s PageRankem 6/10, bude to lepší než 20 stránek s PageRankem 2/10. Odkazy z lépe hodnocených stránek zvýší v očích vyhledávačů důležitost stránek.

Přirozené budování zpětných odkazů

Nejvhodnější je získávat odkazy zpětné odkazy přirozenou cestou (word of mouth), což v praxi znamená, že vytvoříme tak kvalitní projekt se zajímavým a hodnotným obsahem, že na něj budou uživatelé odkazovat sami. Tato metoda však nemusí být vždy úspěšná, zvláště tomu může tak být v prostředí s vysokou konkurencí.

Registrace do katalogů

Registrace do katalogů by měla být první věc, kterou po vytvoření nového webu uděláme. Jedná se o snadné a jednoduché získání zpětného odkazu, který k nám může přivést vyhledávače tak samotné uživatele hledající webové stránky v dané kategorii. Registrace do více katalogů nám může určitě pomoci, ale neměl by to být jediný zdroj zpětných odkazů. Některé základní katalogy, do kterých je vhodné se registrovat:

- <http://odkazy.seznam.cz/>
- <http://katalog.centrum.cz>
- <http://katalog.atlas.cz/>
- <http://www.dmoz.org/World/Česky/>

Výměna zpětných odkazů

Prostá výměna odkazů je vedle registrace do katalogů nejsnadnějším způsobem získávání zpětných odkazů. Nemusí to být pravidlo ale dobré je dbát na to, aby stránka se kterou si chceme odkaz vyměnit, byla tématiky shodná nebo podobná, neměla moc externích odkazů a aby měla stejný nebo vyšší rank než vaše stránka. Web, s kterým odkaz vyměňujete, by měl být tematicky stejný, protože vyhledávače mají raději tematicky relevantní odkazy. Moc externích odkazů nebo nízký rank na webu, který na vás odkazuje, zase znamená, že vám předá málo ranku. Dobré umístění odkazu je také důležité, vyhledávače přikládají odkazům, které jsou umístěné třeba v zápatí webu, malý význam. Nejlépe umístěný odkaz by měl být ten, na který se skutečně kliká.

Výměnu zpětného odkazu můžeme provést tak, že kontaktujete webmastera vybraného webu, a nebo můžeme umístit nabídku na výměnu na internetové tržiště <http://webtrh.cz/f33>.

Nákup zpětných odkazů

Při nákupu zpětných odkazů platí podobné zásady jako při jejich výměně. Cenu odkazu ovlivňuje téma webu, počet externích odkazů, umístění odkazu, ranky a návštěvnost. Odkazy je třeba kupovat postupně, necílit je všechny na titulní stránku, ale i na podstránky. Vhodné je například vytvořit poptávku na koupi zpětných odkazů na internetovém tržišti <http://webtrh.cz/f14>, kde si pak lze vybrat z nejlepších nabídek. [13]

3.5 Optimalizace URL adresy

Jak už zde bylo zmíněno, URL adresy jsou jednou z důležitých součástí stránky a mají nezanedbatelný vliv pro vyhledávače, jelikož mohou obsahovat i klíčová slova pomáhající zvýšit hodnocení dané stránky. Tento případ můžeme pozorovat na všech statických webech. Každá URL adresa zde směřuje na fyzicky vytvořenou stránku. Naopak větší problém může nastat, používáme-li dynamické stránky. Zde se každému požadavku na zobrazení stránky generuje vlastní URL. adresa obsahující několik proměnných a parametrů. Výsledek může vypadat například takto:

Příklad dynamické URL adresy:

```
http://seo.optimalizace.cz/index.php?p=1579372&c=3
```

Celá adresa pak obsahuje několik proměnných a parametrů. Její nevýhody jsou zcela patrné. Může to být například těžká zapamatovatelnost pro běžné uživatele, v některých případech příliš dlouhá adresa a v neposlední řadě i žádné klíčové slovo v URL.

3.5.1 Změna dynamických adres na statické

Jak ale dosáhnout toho aby dynamické URL měli stejné výhody jako ty statické? Řešení může být změna dynamických adres na statické. To se provede pomocí modulátoru `mod_rewrite`. Díky němu můžeme na stránkách vidět adresy tvářící se jako statické.

Požadavky nutné k použití `mod_rewrite`:

- na serveru musí běžet Apache v min verzi 1.2
- `mod_rewrite` zde musí být nainstalován a povolen
- musí zde být možnost konfigurovat server - buďto souborem `httpd.conf`, nebo svým souborem `.htaccess`.

Následující použití ukazuje jednoduchý příklad přechodu na prepisované adresy. Původní adresy byly typu `/index.php?page=neco` - rádi bychom `/neco.html` skrytě

přepsané na fungující skript. Pro zachování odkazů z jiných webů a vyhledávačů (ale se snahou o propagaci nové podoby adres) vytvoříme přesměrování na nové adresy. [16]

```
RewriteEngine On
RewriteCond %{HTTP_HOST} ^muj\.blog\.cz
RewriteRule ^ivt(.*)\.html$
http://muj.blog.cz/index.php?p=$1 [L,QSA]
```

První řádek je nevyhnutelný, syntaxe jím začíná vždy protože RewriteEngine On zapíná mod_rewrite. Druhý řádek, tedy RewriteCond, určuje, za jakých podmínek se bude pravidlo provádět. V opačné případě se syntaxe neprovede. Na třetím řádku se můžeme setkat s tím, že pokud uživatel požaduje stránku, která začíná řetězcem ivt s příponou html, tak se mu podstrčí URL, která vede na skutečné umístění článku. Díky tomuto kódu stránka s příponou .html nemusí existovat. Pokud do příznaků v hranatých závorkách přidáme R, bude se jednat o přesměrování, ne o podstrčení. [15]

4. Zakázané techniky v SEO

Nebo-li také jinými slovy black hat SEO. K dobrým výsledkům ve vyhledávacích vedou dvě cesty. Na jedné straně je možnost použití etického SEO přístupu, a na druhé použití spamu . Obě cesty něco stojí. Náklady na spam by se mohly zdát na první pohled menší, ale není tomu tak. Při použití spam taktiky vždy hrozí propad z předních pozic někam úplně dolů z důvodu, že to někdo odhalí (konkurence nebo přímo vyhledávač).

Cílená a nedovolená manipulace s výsledky vyhledávání ve vyhledávacích může mít za následek penalizaci stránek (většinou ztrátou PageRanku). V horším případě ji může natrvalo vyřadit z indexu. Proto je dobré se těchto pochybných metod vyvarovat.

4.1 *Skrytý a malý text*

Skrytý text patří asi mezi nejrozšířenější praktiky aplikované v různých variantách. Celá podstata věci se skrývá v tom, že autor webu umístí na stránku text, který má shodnou barvu jako pozadí stránky. Výsledný text poté samozřejmě pro běžného uživatele je neviditelný ovšem internetový vyhledávač v něm nevidí oproti normálnímu textu žádný rozdíl. Do „neviditelného“ textu může autor stránky například zapsat text který se odlišuje od tématu stránky a nebo přidat klíčová slova která jsou často vyhledávána. Neblahý vliv to může mít pro samotného návštěvníka webu, kdy se dostane na jinou stránku než původně zamýšlel.

Obdobnou používanou metodou je takzvaný malý text. Jedná se c podstatě o to samé ovšem text není nijak skrytý je jenom tak malý že si ho internetový návštěvník vůbec nemusí zaregistrovat. Přitom text se může nacházet kdekoli na stránce.

Na druhou stranu skrytým text nejsou nejrůznější dynamické akce webu. Může se jednat například o najetí myši na odkaz, rozbalovací menu atd. Takový text uživatel vidí a nemá problém s jeho přečtením.

4.2 Cloaking

Cloaking v podstatě znamená vytváření webových stránek které jsou speciálně optimalizované pro webové roboty . Malý příklad za všechny. Pokud se uživateli podstrčí jiná stránka než ta, která se podstrčí indexovacímu robotovi, tak je to přesně cloaking. Autor takové stránky zoptimalizuje web pro robota, ten dá stránku na první místo ve vyhledávání, ale pokud na něj uživatel klikne, zobrazí se mu úplně jiná stránka.

Za cloaking se nemůže považovat to, že pokud se uživateli, který například používá k procházení internetu Internet Explorer dává trochu jiné HTML než tomu, co používá Firefox. V tomto případě jde totiž o jiné formátování obsahu pro různé webové prohlížeče.

4.3 „Google bomby“

O síle zpětných odkazů vypovídají tzv. Google bomby. Jedná se v podstatě o soustředěnou snahu odkázat na danou stránku na konkrétní fráze nebo klíčové slovo, a to především o takové, které se na dané stránce vůbec nenachází. V naprosté většině k tomu dochází bez vědomí majitele nebo provozovatele stránek. Mezi nejznámější případy patří ty, kdy se skupina lidí dohodne, že bude na svých stránkách odkazovat třeba na stránku neoblíbeného politika a do textu odkazu umístí nějaký nelichotivý výraz. V poslední době se však snaží Google podobným bombám bránit. [2]

4.4 Odkazové farmy

Odkazové farmy vznikly krátce po vzniku Google. Jde o několik stovek až tisíců navzájem provázaných stránek . Vyhledávače jim díky velkému počtu zpětných odkazů, přiřadí relativně vysoké hodnocení. Tyto stránky většinou nemají žádný informativní obsah. Jejich návštěvníci jsou pomocí scriptů nebo výrazných odkazů, přesměrování jinam.

Odkazové farmy obsahují velký počet stránek, který je často generuje i automaticky na základě slovníků a vzájemně prolink mezi sebou. Odkazové farmy se nedoporučuje využívat, protože v současnosti existuje něco jako špatné okolí stránky, kvůli kterému by mohly být naše stránky zbytečně penalizovány. Proto je dobré zvážit, s kým si odkazy vyměňujete. Google totiž postihuje většinu účastníků odkazových farem nulovým PageRank.

4.5 Klamné přesměrování

Odkazy, kde je použito klamné přesměrování přesměrují uživatele na jinou stránku než uživatel předpokládal. Uživatel si například zadá do vyhledávače klíčové slovo autobazar. Po kliknutí na konkrétní stránku je ale přesměrován na úplně jinou stránku s docela jiným obsahem.

4.6 Doorway Pages

Za doorway pages se považují vysoce optimalizované stránky, které však nejsou naplněny kvalitním a relevantním obsahem, ale pouze množstvím odkazů na domovskou stránku, a tím se snaží zvýšit její hodnocení.

Naopak za doorway pages není požadována registrace domény druhého stupně pod více národních domén (například .cz, .com, .eu...), To se dělá z toho důvodu, že si doménu zakoupil obchodník pod vidinou snadného zisku z obchodu s doménou a nyní jí chce výhodně prodat. Je to čistě ochrana značky (názvu stránky).

4.7 Keyword stuffing

Keyword stuffing spadá do black hat SEO. Klíčová slova se v tomto případě použijí ve vysokém počtu a hustotě na stránce bez toho aby měly význam pro návštěvníka stránky. Výsledkem bývá jako ve většině případů penalizace od vyhledávačů. Někdy může dojít k penalizaci i "preoptimalizovaním" stránky. [14]

Veškeré zde popsané techniky black hat SEO se v praxi nedoporučuje využívat. Jednak přichází v úvahu možnost penalizace nebo až vyřazení stránek z indexu a jednak také z morálního hlediska pro samotného uživatele.

5. Závěr

Webové stránky jsou a budou z jednou možností prezentování se na internetu široké veřejnosti. Tito lidé se zde orientují pomocí moderních vyhledávacích nástrojů, jaké dnešní vyhledávače nabízejí. I proto je zde snaha tvůrců webu optimalizovat své stránky tak, aby se umístily na předních pozicích a zvýšili si jak návštěvnost, tak i potenciální zisk.

Budoucnost a možný vývoj vidím ve značném rozšíření sociálních webů a internetových stránek kde se obsah tvoří v reálném čase samotnými uživateli. I tyto informace musí uživatelé zpětně najít a tudíž je i jejich provozovatelé vhodně optimalizovat. V poslední době začal například Google poskytovat službu výpisu nalezených informací v reálném čase. Jedná se však zatím jen o testovací provoz a do budoucna se máme určitě na co těšit.

Ve své práci jsem se pokusil shrnout možnosti a metody optimalizace. Věnoval jsem se hlavním on-page i off-page faktorům které jsou nejdůležitější věci samotné optimalizace. I v dnešní době se však najde spousta internetových stránek, které se jimi neřídí a dobrovolně se tím připravují o nemalé množství návštěvníků. V neposlední řadě jsem se věnoval i zakázaným a nedovoleným praktikám, se kterými se můžeme v SEO optimalizaci setkat ač bychom neměli. Mějme přitom ale neustále na paměti, že stránky které vytváříme by měly sloužit a být prospěšné hlavně uživateli samotnému a optimalizace se zde bere až jako nadstavba kvalitní a dobře zpracované stránky.

Seznam použitých zkratek

WWW	World Wide Web
HTML	Hypertext Markup Language
XHTML	eXtensible Hypertext Markup Language
SEO	Search Engine Optimization
SEM	Search Engine Marketing
URL	Uniform Resource Locator
PDF	Portable Document Format
SERP	Search engine results page
PPC	Pay per Click
CSS	Cascading Style Sheets
ASP	Active Server Pages
PHP	Hypertext Preprocessor <i>původně</i> (Personal Home Page)
XML	Extensible Markup Language
DHTML	dynamické HTML

Seznam použitých tabulek

Tab. 1.	<i>Příklady českých a zahraničních katalogových vyhledávačů</i>	13
Tab. 2.	<i>Příklady českých a zahraničních fulltextových vyhledávačů</i>	13
Tab. 3.	<i>Podíl vyhledávačů na českém trhu</i>	18
Tab. 4.	<i>Počet nalezených stránek na klíčová slova a slovní spojení</i>	23
Tab. 5.	<i>Porovnání hodnocení stránek z pohledu PageRanku a S-Ranku</i>	30

Seznam použitých obrázků

Obr. 1.	<i>Setříděný katalog</i>	12
Obr. 2.	<i>Schematická cesta pavouka</i>	15
Obr. 3.	<i>Distribuce PageRanku mezi provázanými stránkami</i>	29

Seznam použité literatury

- [1] Hlavenka, Jiří. *Mistrovství ve vyhledávání na Internetu*. 2. Vydání. Brno: Computer Press, 2004. 196 s. ISBN 80-7226-759-0
- [2] Kubiček, Michal. *Velký průvodce SEO*. 1. Vydání. Brno: Computer Press, 2008 318 s. ISBN 978-80-251-2195-5
- [3] *Jak by měla vypadat kvalitní webová stránka* [online]. [cit. 2010-03-20]. Dostupné z: <http://www.koldasoft.cz/pozadavky-kvalitni-stranky/#design>.
- [4] *Dynamické webové stránky* [online]. [cit. 2010-03-23] Dostupné z: <http://www.mediacentrik.cz/slovník/dynamicke-webove-stranky.aspx>
- [5] *Google ve vyhledávacích zohledňuje rychlost stránek* [online] [cit. 2010-04-19] Dostupné z: <http://digitalne.centrum.cz/google-ve-vyhledavani-zohlednuje-rychlost-stranek/>
- [6] *Optimalizace webu* [online] [cit. 2010-04-05] Dostupné z: <http://www.klosko.net/blog/=clanek=optimalizace-webu>
- [7] *Charakteristika a vývoj internetových vyhledávačů – SEO – díl 1* [online] [cit. 2010-04-10] Dostupné z: <http://www.swmag.cz/264/charakteristika-a-vyvoj-internetovych-vyhledavacu-seo-dil-1/>
- [8] *Česká stop slova (stop words)* [online] [cit. 2010-04-05] Dostupné z: <http://www.michaljanik.cz/stop-slova>
- [9] *Optimalizace ON Page faktorů – SEO – díl 2*. [online] [cit. 2010-04-10] Dostupné z: <http://www.swmag.cz/264/charakteristika-a-vyvoj-internetovych-vyhledavacu-seo-dil-1/>

- [10] *Web-Design BLOG* [online] [cit. 2010-04-20] Dostupné z: <http://blog.maxell.cz.cz/item/k-cemu-dnes-slouzi-meta-description>
- [11] *Seznam – Komunikace s vyhledávači* [online] [cit. 2010-04-10] Dostupné z : <http://help.seznam.cz/cz/hledani-fulltext-komunikace-s-vyhledavaci-robots-txt.html>
- [12] *Google PageRank* [online] [cit. 2010-04-20] Dostupné z : <http://www.jakpsatweb.cz/seo/pagerank.html>
- [13] *Optimalizace Off Page faktorů a podvodné techniky – SEO – díl 3.* [online] [cit. 2010-04-20] Dostupné z <http://www.swmag.cz/371/optimalizace-off-page-faktoru-a-podvodne-techniky-seo-dil-3/>
- [14] *Seo – Optimalizace pro vyhledávače* [online] [cit. 2010-04-20] Dostupné z: <http://seoprojekt.szm.com/zakaz.html>
- [15] *Nelíbí se vám dynamické URL?* [online] [cit 2010-04-26] Dostupné z: <http://www.iceboy.cz/rewrite/predelani>
- [16] *Mod_rewrite – příklady* [online] [cit 2010-04-26] Dostupné z: http://www.jakpsatweb.cz/server/mod_rewrite/mod-rewrite-priklady.html