ČESKÁ ZEMĚDĚLSKÁ UNIVERZITA V PRAZE

PROVOZNĚ EKONOMICKÁ FAKULTA

KATEDRA INFORMAČNÍCH TECHNOLOGIÍ

ICT jako významný faktor konkurenceschopnosti

disertační práce

Autor: Ing. Pavel Šimek

Školitel: Doc. PhDr. Ivana Švarcová, CSc.

© 2007

Prohlášení

Prohlašuji, že disertační práci na téma „**ICT jako významný faktor konkurenceschopnosti“** jsem vypracoval samostatně a použil jsem pramenů, které jsou uvedeny v přiloženém seznamu literatury.

V Praze dne 21. září 2007

Pavel Šimek

Poděkování

Rád bych při této příležitosti poděkoval své školitelce **Doc. PhDr. Ivaně Švarcové, CSc.,** za ochotu a odborné vedení, nejen při psaní této práce, ale během celého studia.

Dále děkuji Ing. Karlu Jiránkovi a Bc. Ireně Krupičkové za pomoc při praktické implementaci metodického postupu optimalizace dokumentu na několika reálných projektech.

Souhrn

Předkládaná disertační práce se zabývá problematikou optimalizace dokumentu a celých website pro fulltextové vyhledávače a je rozdělena do dvou základních částí. První část se zabývá teoretickou základnou obsahující principy a možnosti služby World Wide Web, principy a možnosti fulltextových vyhledávačů, přehled již známých technik Search Engine Optimization, poslední vývojové trendy v oblasti Search Engine Marketingu a analýzu vlivů různých faktorů na hodnocení relevance WWW stránky vyhledávacím strojem.

Ve druhé části disertační práce je splněn její hlavní cíl, tedy navrhnutí metodického a ověřeného postupu pro optimalizaci dokumentu na určitá klíčová slova pro nejpoužívanější fulltextové vyhledávače od úplného začátku, čili od předoptimalizačních příprav, až do úplného konce, tedy po postoptimalizační analýzu efektivity. Podkladem pro navrhnutí metodického postupu optimalizace dokumentu pro fulltextové vyhledávače je mimo teoretické základny disertační práce hlavně zhodnocení vlivu faktorů na hodnocení relevance dokumentu vyhledávacím strojem.

Klíčová slova

World Wide Web, Search Engine Optimization, dokument, WWW stránka, website, vyhledávač, fulltext, katalog, PageRank, klíčové slovo, výsledek vyhledávání, relevance, Google

Summary

Submitted dissertation thesis on the topic of documents and Website optimalization for full-text search engines comprises of two core sections. The first section deals with the theoretical principles and possibilities of the World Wide Web and full-text search engines. Further an overview of known Search Engines Optimization techniques and current trends in the development of Search Engine Marketing is offered and analysis of various factors influence on relevancy of Web page determined by search engine is presented.

The second section deals with the construction and testing of a methodical tool for document optimalization, thus the main goal of the dissertation thesis is fulfilled. Optimalization is designed for keywords used in the most favoured full-text search engines - from its beginning, pre-optimalization adjustments, till its end, post-optimalization analysis of effectiveness. Basis for construction of the methodical tool for documents optimalization is, besides the theoretical principles of the dissertation thesis, built mainly on analysis of various factors influence on relevancy of document determined by search engine.

Key words

World Wide Web, Search Engine Optimization, document, WWW page, website, search engine, fulltext, catalog, PageRank, key word, result of searching, relevantion, Google

Obsah

1 Úvod ... 3
2 Cíle disertační práce ... 5
3 Metodika ... 6
4 ICT a konkurenceschopnost ... 8
5 Internet, prostředí World Wide Web a vyhledávače ... 10
5.1 World Wide Web ... 11
5.1.1 Hypertext ... 13
5.1.2 Změny ve filosofii World Wide Web ... 15
5.1.3 HyperText Markup Language ... 17
5.1.4 Domény a Unique Resource Locator ... 20
5.1.5 Diakritika v doménách ... 22
5.2 Vyhledávače ... 25
5.2.1 Katalogy a vyhledávače ... 28
5.2.2 Nejpoužívanější vyhledávače a katalogy ... 31
5.2.3 Google ... 35
5.2.3.1 Základy práce uživatele s Googlem ... 36
5.2.3.2 PageRank a Hilltop ... 46
6 Search Engine Optimization ... 51
6.1 Marketing ... 51
6.2 Analýza klíčových slov ... 54
6.2.1 Profesionální nástroje ... 57
6.2.2 Váhy jednotlivých slov ... 58
6.3 Úprava serveru ... 60
6.4 Kód WWW stránky ... 62
6.5 SEO copywriting ... 64
6.6 Zpětné odkazy ... 67
6.7 Kontinuální SEO ... 69
7 Search Engine Marketing ... 72
7.1 Matice FCB a faktory ovlivňující SEM ... 73
7.2 SEM na Google AdWords ... 76
7.3 Agentura DoubleClick a Performics 50 ... 79
8 Faktory ovlivňující algoritmus Google ... 86
8.1 Faktory spojené s používáním klíčových slov ... 87
8.2 Faktory spojené se stránkou ... 91
8.3 Faktory spojené s website (doménou) ... 95
8.4 Faktory spojené s vlastnostmi zpětných odkazů ... 98
8.5 Faktory s negativním vlivem na crawling nebo ranking ... 100
9 Metodický postup optimalizace dokumentu pro WWW vyhledávače ... 103
9.1 Analýza klíčových slov ... 104
9.2 Analýza konkurence ... 112
9.3 Analýza aktuálního stavu website ... 118
9.4 MPOD copywriting ... 120
9.5 Optimalizace dokumentu a serveru ... 125
9.5.1 Klíčová slova v dokumentu ... 125
9.5.2 Vlastnosti dokumentu ... 128
9.5.3 Doména dokumentu ... 130
9.6 Tvorba zpětných odkazů a registrace do katalogů ... 133
9.7 Analýza efektivity ... 137

10 Ověření metodického postupu optimalizace dokumentu pro WWW vyhledávače ... 141
10.1 Aplikace MPOD na WWW prezentaci VCC Dvůr Králové nad Labem .. 141
10.1.1 Analýza klíčových slov ... 142
10.1.2 Analýza konkurence ... 143
10.1.3 Analýza stávajícího stavu ... 148
10.1.4 MPOD copywriting ... 148
10.1.5 Optimalizace dokumentu a serveru ... 149
10.1.6 Registrace do katalogů a tvorba zpětných odkazů ... 149
10.1.7 Analýza efektivity ... 150
10.2 Stěhovací firma Čtvrtečka ... 152
11 Závěrečné shrnutí ... 158
12 Přehled literatury ... 162
13 Seznamy ... 165
13.1 Seznam publikační činnosti autora (výběr) ... 165
13.2 Seznam grafů ... 167
13.3 Seznam obrázků ... 167
13.4 Seznam příkladů ... 167
13.5 Seznam schémat ... 167
13.6 Seznam tabulek ... 168
14 Přílohy ... I
14.1 Použité zkratky ... I
14.2 Pojmy World Wide Web ... III
14.3 Internetové organizace ... VI
14.4 Standardizační instituce ... IX
14.5 České fulltextové vyhledávače a katalogy ... XII
14.6 Zahraniční fulltextové vyhledávače a katalogy ... XIV

1 Úvod

V poslední době se informační a komunikační technologie rozvíjejí nebývalou rychlostí. Komerční subjekty, subjekty státní správy a samosprávy, ale i jednotlivci jsou nuceni ICT[1] stále více používat, mají stále větší a větší nároky, což vede ke zdokonalování, zlevňování a masovému rozšiřování těchto technologií. Neustálé zkracování vývojových, realizačních i produkčních cyklů je současně důsledkem explozivního nárůstu IT[2] aplikací a jejich prorůstání do všech složek lidských aktivit [30]. Zároveň jsou ICT jedním z rozhodujících faktorů růstu a produktivity vyspělých ekonomik [43]. V době počínající „computerizace“ firem bylo ICT vnímáno jako jeden ze způsobů, jak získat výhodu před konkurencí [1].

Elektronická forma prezentace firmy vůči svému okolí se tak stává jedním ze základních zdrojů konkurenceschopnosti a samotné webové prezentace s nezpomalujícím rozvojem Internetu se dostávají stále více do popředí zájmu především z hlediska získávání informací. Ať už vznikly, aby prezentovaly osobní názory, znalosti a dovednosti, více či méně umělecké výtvory jednotlivců i zájmových skupin, či jako zdroje informací, novinek a vědeckého poznání, představují mnoho dat, ve kterých je velmi obtížné vyhledávat.

Vytvořit kvalitní internetový projekt, ať už je to webová prezentace firmy, elektronický magazín nebo webová služba, je bezesporu náročná záležitost. Bohužel, mnoho kvalitních projektů zůstává koncovým uživatelům skryto hluboko v útrobách Internetu a ztrácí tak obě zainteresované strany – provozovateli website uchází zisk a uživatel (potencionální zákazník) přichází o kvalitní obsah nebo produkt.

Rozhraním, zprostředkujícím návštěvníkům Internetu odkaz vedoucí k požadovaným informacím, jsou vyhledávací servery. Ve většině z nich je návštěvník vyzván k zadání hledaného slova nebo fráze a vyhledávač mu nabídne seznam odkazů (doplněných stručným popisem), které s hledaným slovem nebo frází souvisí. Pokud dotazu vyhovuje více odkazů, jsou seřazeny podle domnělé relevance. Tato relevance je plně v režii vyhledávacího stroje.

Právě zde má vysoký význam optimalizace WWW stránek pro fulltextové vyhledávače (SEO[3]) a marketing založený na vyhledávačích (SEM[4]), neboť běžní

[1] Information and Communication Technologies
[2] Information Technologies
[3] Search Engine Optimization
[4] Search Engine Marketing

uživatelé obvykle zkusí odkazy jen z prvních několika stránek výsledků fulltextových vyhledávačů na určitá klíčová slova a v katalozích využívají především hierarchicky výše položené odkazy v jednotlivých kategoriích.

Klíčem k úspěchu je aplikace optimalizačních metod, které se zabývají problematikou klíčových slov, kvalitou a strukturou obsahu, názvy domén i jednotlivých stránek a množstvím a věrohodností zpětných odkazů. Proces je to náročný, dlouhodobý a bez zaručeného výsledku. Mnoho provozovatelů webů se proto uchyluje k alternativě, kterou představuje celá řada variant placeného zobrazení odkazů.

Současní největší hráči na trhu fulltextových vyhledávačů jsou Google, Yahoo!, Microsoft a ASK. O AOL, které používá technologii Google, se spekuluje, že bude koupeno některým „velkým hráčem“. Podíl vyhledávačů je samozřejmě obtížně měřitelný a existují různé přístupy. O tom, že ale globálně vede Google však žádné spory nejsou. Od dubna 2007 do července 2007 sice Google necelá 2 procenta z celkového počtu vyhledávání ztratil, absolutní počet vyhledávání však stoupl o více než 370 mil [7] [8].

České vyhledávače se vyvíjely se zpožděním za těmi světovými. V roce 1997 naprogramoval Ivo Lukačovič a Štěpán Škrob (spolužáci na ČVUT[5]) fulltextový vyhledávač Kompas, který používal katalog Seznam jako vyhledávací technologii. V určité podobě fungoval Kompas až do roku 2001, kdy Seznam začal používat Google. Později zakoupil technologii Jyxo, kterou na začátku roku 2005 nahradil svou vlastní – Seznam Fulltextem. Seznam, tak již primárně nevyhledává v katalogu, ale ve fulltextu.

Zatímco tradiční Search Engine Optimization je orientováno spíše na dílčí úpravy webových stránek, které mají za cíl co nejlepší umístění odkazů ve fulltextových vyhledavačích, Search Engine Marketing na základě pečlivé analýzy nejprve formuluje účinnou strategii a tu pak aplikuje nejen v oblasti typických fulltextových vyhledavačů, ale i na katalogy stránek a vyhledavače typu pay-per-click.

[5] České vysoké učení technické

2 Cíle disertační práce

Výchozí myšlenkou předkládané disertační práce na téma „ICT jako významný faktor konkurenceschopnosti“ je, že v současné době neexistuje žádný, veřejně známý a ověřený, metodický postup optimalizace dokumentu pro nejrozšířenější fulltextové vyhledávače. Přitom dle statistik renomovaných agentur, např. DoubleClick, využití fulltextových vyhledávačů při vyhledávání informací v celosvětové síti Internet stále roste a samotní provozovatelé website investují nemalé finanční částky do tvorby a úprav vlastních dokumentů a celých website jen z důvodu, aby se jejich odkazy objevily na prvních místech ve výsledcích vyhledávání na určitá klíčová slova, nebo si kupují textové reklamní kampaně na přesně definovaná klíčová slova.

Největším problémem různých postupů a metod samotné optimalizace WWW stránek pro fulltextové vyhledávače jsou jednak nepřesně propracované jednotlivé kroky optimalizace, popř. absence určitých kroků, dále tzv. „zamrznutí“ optimalizace na určitém bodě (kroku) a dynamický vývoj algoritmů vyhledávacích strojů.

Hlavním cílem disertační práce je navrhnutí metodického postupu optimalizace dokumentu na určitá klíčová slova pro nejrozšířenější fulltextové vyhledávače od úplného začátku, čili od předoptimalizačních příprav, až do úplného konce, tedy po analýzu efektivity. Jelikož v současné době většina předních fulltextových vyhledávačů neindexuje jen XHTML dokumenty, ale i řadu jiných formátů (např. `.pdf`, `.doc`, `.ppt`, `.txt`, apod.), bude metodický postup optimalizace dokumentu navržen obecně pro nejširší možné použití.

Hlavní cíl disertační práce, tedy navrhnutí metodického postupu optimalizace dokumentu pro fulltextové WWW vyhledávače, částečné vychází z vypracovaného přehledu vlivu jednotlivých faktorů na vyhledávací algoritmus Google.

Dalším cílem disertační práce je navrhnutý metodický postup optimalizace dokumentu pro fulltextové WWW vyhledávače ověřit na několika reálných projektech, výsledky změřit a kvantifikovat přínos.

3 Metodika

V předkládané disertační práci je nejdříve vymezen pojem „konkurenceschopnost“ ve vztahu k moderním informačním a komunikačním technologiím, zpracovány základní charakteristiky a principy služby World Wide Web a s ní souvisejících technologií s důrazem na doménová jména a základní rozdělení internetových vyhledávačů a jejich principů. Důraz je kladen na, z uživatelského hlediska, nejrozšířenější fulltextový vyhledávač Google.

V další části jsou popsány již známé a částečně ověřené postupy optimalizace samotných WWW stránek, popř. celých website, pro fulltextové vyhledávače. Popsány jsou jak vybrané on-page faktory, tak několik off-page faktorů a podkladem je několik českých a zahraničních zdrojů, zejména firem a odborníků, kteří se optimalizací zabývají prakticky.

Pro vypracování stručné vývojové analýzy v oblasti marketingu fulltextových vyhledávacích strojů jsou použity zejména čtvrtletní zprávy agentury DoubleClick. Jelikož samotný marketing na fulltextových vyhledávačích bývá v určitých případech doplňkem optimalizace WWW stránek pro vyhledávací roboty a na druhou stranu v určitých případech konkurencí, je nutné, aby disertační práce takovouto analýzu obsahovala.

Poslední kapitolou první části metodického postupu je vliv jednotlivých faktorů (on-page i off-page) na hodnocení relevance dokumentu vyhledávacím strojem na určitá klíčová slova. Výsledná míra vlivu jednotlivých faktorů byla stanovena jednak ze cca čtyřletého pozorování různého průběžného umisťování 10 webových prezentací z různých oblastí a dokumentu „Google Search Engine Ranking Factors V2“, na kterém se podílelo 37 předních světových odborníků v oblasti vyhledávacích strojů.

Jako základ pro navržení metodického postupu optimalizace dokumentu pro fulltextové WWW vyhledávače je použita celá rešeršní část disertační práce, zejména však již známe principy a kroky optimalizace WWW stránek a website a míra vlivu jednotlivých faktorů na ohodnocení relevance dokumentu vyhledávacím strojem. Požadavkem metodického postupu optimalizace dokumentu pro fulltextové WWW vyhledávače je zahrnutí i předoptimalizační přípravy a postoptimalizační analýzy efektivity.

V předoptimalizační analýze jsou využity metody vícekriteriální analýzy variant, zejména metoda TOPSIS, metoda váženého součtu a Saatyho metoda. Současně je použit volně dostupný aplikační SW, bez kterého by nebylo možné předoptimalizační analýzy realizovat.

Součástí předoptimalizačních analýz je i analýza konkurence, kde je navržen výpočet konkurenční hodnoty dokumentu, konkurenční hodnoty website v rámci klíčového slova a konkurenční hodnoty website. Výpočty konkurenčních hodnot jsou zdrojovou informací pro samotný metodický postup optimalizace dokumentu pro fulltextové WWW vyhledávače.

Druhým zdrojovým podkladem metodického postupu optimalizace dokumentu pro fulltextové WWW vyhledávače je míra vlivu jednotlivých on-page a off-page faktorů na hodnocení relevance dokumentu vyhledávacím strojem. Součástí metodického postupu jsou i podpůrné kroky pro vyšší úspěšnost dokumentu ve výsledcích fulltextového vyhledávání a postoptimalizační analýzy, zejména analýza efektivity.

4 ICT a konkurenceschopnost

Ve výkladovém slovníku lze pod heslem „konkurenceschopnost“ nalézt schopnost prosadit se v určitém oboru v porovnání s ostatními, s odkazem na definici konkurence. Význam pojmu „konkurenceschopnost“ se v průběhu let vyvíjel. V počátcích byla konkurenceschopnost v získání disponibilních zdrojů, dnes ji lze chápat jako širší soubor konkurenčních výhod [10].

Historie ICT není nijak dlouhá, ale již dnes lze pozorovat, že vývoj ICT do jeho současné pozice probíhal v několika cyklech o periodě 10-15 let. Každá z těchto etap představovala i model pozice ICT v podnikání. Od automatizace agend v centrálách firem přes produktivitu oddělení a poboček až po dnešní stav, kdy ICT podporuje integraci všech klíčových obchodních procesů [28].

Při pohledu na roli ICT v libovolné organizaci lze dojít k jednoznačnému závěru, že hlavním úkolem je automatizovat efektivně hlavní, vedlejší a podpůrné podnikové procesy, vytvořit informační základnu, která slouží pro realizaci těchto procesů a pro podporu rozhodování a poskytnout služby pro interní i externí komunikaci (elektronickou, hlasovou či obrazovou). Komplexnost ICT systémů v dnešních podnicích vyplývá z rozsáhlosti nasazení a vzájemné provázanosti jednotlivých systémů. Samozřejmě, že velkou roli hrají také globalizační procesy, které vytvářejí stále mohutnější organizace, jejichž chod je podporován systémy s odpovídající složitostí.

Podoba jednotlivých částí ICT systémů může být odlišná – samostatný systém, komponenta, balíková aplikace, modul integrovaného systému v závislosti na zvoleném přístupu té či oné organizace. Také názvy těchto aplikací a jejich funkcionalita se mohou mírně odlišovat podle pojetí konkrétního dodavatele [9]. V dnešní době je stále patrnější, ze podmínky na trhu IT se od prasknutí internetové bubliny v roce 2001 radikálně změnili. Nyní si každá společnost začíná více zajímat o to, kde je hlavní hodnota jejích IT oddělení, jak toto oddělení přispívá k hlavnímu zaměření společnosti, tzn. dosažení obchodních a strategických cílů, jak co nejpřesněji sledovat a měřit náklady, efektivitu a kvalitu poskytovaných služeb IT, jak vybrat nejefektivnější investici do IT a jak je každá taková investice svázána s obchodními výsledky [23].

Současný potenciál a stále se rozšiřující možnosti informačních a komunikačních technologií představují pro naprostou většinu vyspělých zemí jeden

z klíčových zdrojů ekonomického růstu a na úrovni jednotlivých ekonomických subjektů i zdroj jejich konkurenceschopnosti. Předpokladem pro dosažení tohoto stavu je nejen adekvátní míra a struktura investic do ICT, ale zejména úroveň jejich využití směrem k dosažení cílových efektů a konkurenceschopnosti firmy [32].

Konkurenceschopnost moderní organizace významně ovlivňuje mnoho faktorů. Negativním následkem nemusí být vždy přímo ztráta zákazníka, ale i propad cash-flow, nižší zisk díky vysokým nákladům na IS[6]/ICT apod. Hlavními faktory tedy jsou:

- dostupnost systémů,
- vlastní řešení IS/ICT, nákup IS/ICT, pronájem IS/ICT,
- předávání dat,
- náklady,
- bezpečnost a zranitelnost informačních systémů,
- multiutility,
- provázanost IS/ICT a obchodních procesů a
- kompetence personálu IS/ICT [39].

Ještě důležitějším faktorem, ovlivňující úspěšné fungování moderní organizace, je potřeba dostatečného množství informací. Dnes již ovšem nestačí získat informace jako takové, ale tyto musí splňovat určité parametry. Hodnotná a smysluplná informace je poskytnuta v pravou chvíli a na pravém místě, je relevantní, korektní, úplná a jistě by se našlo ještě mnoho dalších atributů [37]. Sdílení informací a sdílení znalostí jsou rovněž témata současné postindustriální doby. Je to disciplína, která vyžaduje náležitý trénink, jasné cíle a odpovídající prostředky [26].

Nedílnou součástí konkurenceschopnosti je i technologicky kvalitní způsob poskytování svých informací, v současné době prostřednictvím Internetu, za účelem získání co největšího počtu přínosných zákazníků. Bohužel, mnoho kvalitních projektů zůstává koncovým uživatelům skryto hluboko v útrobách Internetu a ztrácí tak obě zainteresované strany – provozovateli website uchází zisk a uživatel (potencionální zákazník) přichází o kvalitní obsah nebo produkt.

[6] Information Systems – informační systémy

5 Internet, prostředí World Wide Web a vyhledávače

Informace je základním pojmem teorie poznání. Lze ji neformálně popsat jako veličinu, která určuje stupeň jistoty, že nastane nějaká událost. Poznání je třeba zaznamenávat a vzájemně sdělovat [40]. Informace představují obrovský potenciál ekonomického růstu a rozvoje společnosti. Tvoří vlastně základ informační společnosti [17].

Architektury internetových aplikací se nejvíce blíží centralizovanému výpočetnímu modelu s mnoha distribuovanými „tenkými“ klienty, kteří většinou vykonávají o trochu více než jen prezentaci dat a připojují se k centrálnímu „tlustému“ serveru, na kterém dochází k podstatné části zpracování. Co odlišuje webové architektury od centralizovaných výpočetních modelů je to, že jsou založeny na využití jazyka HTML[7] a jeho primárního transportního média, kterým je protokol HTTP[8] [35].

Elektronická forma prezentace firmy vůči svému okolí se stává jedním ze základních zdrojů konkurenceschopnosti a samotné webové prezentace s nezpomalujícím rozvojem Internetu se dostávají stále více do popředí zájmu především z hlediska získávání informací. Ať už vznikly, aby prezentovaly osobní názory, znalosti a dovednosti, více či méně umělecké výtvory jednotlivců i zájmových skupin, či jako zdroje informací, novinek a vědeckého poznání, představují mnoho dat, ve kterém je velmi obtížné něco vyhledat.

S intenzivním rozvojem nejrůznějších webových prezentací a aplikací je webhosting atraktivní formou jejich provozu, zejména pro menší zákazníky. Webhosting je vlastně poskytování všech služeb sloužících k provozu WWW prezentací i aplikací a zahrnuje:

- poskytování diskového prostoru na serverech
- jejich připojení do sítě a zajištění neustálého provozu
- e-mailové služby a podporu různých vývojových prostředí a databází nutných pro rozvoj a provoz webových aplikací
- technickou podporu při provozu webových aplikací.

Další variantou hostingu je poskytování technických kapacit serverů, tzv. serverhosting. Serverhosting je služba umístění vlastního webového (i jiného)

[7] HyperText Markup Language
[8] HyperText Transfer Protocol

serveru do prostor, který se pronajímá od poskytovatele, a to včetně připojení k Internetu.

Rozhraním, zprostředkujícím návštěvníkům Internetu odkaz vedoucí k požadovaným informacím, jsou vyhledávací servery. Ve většině z nich je návštěvník vyzván k zadání hledaného slova nebo fráze a vyhledávač mu nabídne seznam odkazů (doplněných stručným popisem), které s hledaným slovem nebo frází souvisí. Pokud dotazu vyhovuje více odkazů, jsou seřazeny podle domnělé relevance. Tato relevance je plně v režii vyhledávacího stroje.

5.1 World Wide Web

O službě World Wide Web, zkratkou WWW či pouze Web, se dnes hovoří jako o nosné aplikaci Internetu, a je na ni soustředěna opravdu velmi velká pozornost – jak ze strany širší uživatelské veřejnosti, tak i ze strany poskytovatelů nejrůznějších informačních služeb. World Wide Web prošel za dobu své existence určitým vývojem, během kterého se poněkud měnily představy obou zúčastněných stran o tom, k čemu a jak je vhodné Web využívat [29].

Všeobecná popularita i vhodnost služby World Wide Web se v posledních patnácti letech projevuje jedním velmi zajímavým trendem, který úzce souvisí i se snahou zpřístupnit možnosti Internetu co možná nejširšímu okruhu uživatelů. Jde o to, že až dosud se pro různé činnosti v rámci Internetu musely používat různé více či méně specializované služby. Ty museli uživatelé nejen znát a umět ovládat, ale museli mít k dispozici i specifické programy, nutné pro práci s těmito službami. To bylo nevýhodné jak z pohledu samotných uživatelů a nároků na jejich znalosti a dovednosti, tak i z pohledu správců sítí, kteří museli udržovat v řádně nakonfigurovaném a provozuschopném stavu mnoho různých síťových aplikací.

Jako výhodné řešení se ukazuje možnost přenést tyto „další“ služby na platformu služby World Wide Web – neboli umožnit uživatelům práci s těmito službami prostřednictvím Webu a jeho klientských programů, označovaných jako WWW prohlížeče, alias browsery. U starších, již existujících služeb, se to řeší prostřednictvím brány (fungující na straně WWW serveru), která umožňuje komunikaci ze světa Webu do specifického světa původní služby. Novější služby, které teprve vznikají, se již od začátku budují jako nadstavby nad Webem. Jde

například o nejrůznější vyhledávací služby, adresářové služby, informační služby zajišťující přístup do databází apod.

Výsledkem je pak zajímavý všeobecný trend, v rámci kterého se služba World Wide Web stává jednotnou a univerzální klientskou platformou: její prohlížeče (browsery) jsou jednotným nástrojem, prostřednictvím kterého uživatel může pracovat s mnoha různými službami, navíc jednotným a velmi intuitivním způsobem, charakteristickým pro celou službu WWW jako takovou [1].

Trend směřující k využití Webu v roli jednotné klientské platformy souvisí i s dalším vývojem výpočetního modelu klient/server. Ten standardně předpokládá dvoučlennou dělbu práce, přičemž každá z obou těchto složek (tj. klient i server) jsou pro danou aplikaci specifické. V praxi to znamená nejen potřebu provozovat specifické serverové části jednotlivých aplikací, ale zejména také potřebu používat na straně klienta specifické klientské části aplikací, které jsou odlišné od klientských částí jiných aplikací, vyžadují jiný styl práce, po uživateli chtějí specifické znalosti a dovednosti apod. To je samozřejmě ve sporu s tendencí k jednotné a univerzální klientské platformě, která by dokázala nabídnout uživateli jednotný nástroj a jednotné rozhraní pro přístup k více různým aplikacím (a která se postupně prosazuje i mimo Internet, byť poněkud pomaleji a méně výrazně).

Z právě naznačených důvodů proto došlo i ve výpočetním modelu klient/server k zajímavému vývoji, který ze dvoučlenné dělby práce nakonec vytvořil trojčlennou dělbu práce: prostředním článkem je zde část, zajišťující tzv. aplikační logiku, neboli maximum činností, které jsou specifické pro danou aplikaci. Z jedné strany pak na tuto prostřední část navazuje složka zajišťující běžné „uskladnění“ dat (typicky standardní databázový server), a z druhé strany pak složka zajišťují poměrně jednoduché „prezentační“ funkce – přebírá od uživatele jeho dotazy resp. povely, a posléze mu zase předkládá výsledky jeho dotazů, povelů, požadavků atd. Právě tato role je přitom doslova jak ušitá pro službu World Wide Web: ta dokáže velmi snadno nabídnout uživateli jednoduchý formulář, jehož vyplněním uživatel vyspecifikuje svůj požadavek, a pak mu stejně tak snadno může předložit výsledek ve formě dynamicky vygenerované WWW stránky, obsahující například přehledný graf, tabulku apod. [26]

Službu World Wide Web si pro svou potřebu původně vyvinula komunita fyziků, zabývajících se vysokými energiemi. Jejich cílem bylo co možná

nejefektivněji sdílet mezi sebou množství dostupných informací, ponejvíce textové povahy. Tomuto účelu pak vznikající Web také uzpůsobili – od začátku jej „postavili“ na myšlence tzv. hypertextu. Co do vznikajícího Webu naopak zpočátku nezabudovali, a to záměrně, byla podpora grafiky. Původní autoři se totiž obávali lidské lenosti, konkrétně toho, že by se lidé nenamáhali převádět textové informace do takového tvaru, jaký byl službě Web vrozen (tj. do tvaru jazyka HTML), a místo toho by například nasnímali stránku psaného textu do podoby rastrového obrázku, a tento obrázek pak na Internetu prostřednictvím služby WWW vystavili. Rozdíl by byl dokonce několikařádový: stejná textová informace ve formátu HTML by měla nejspíše několik málo kilobytů, zatímco ve formě obrázku by stejná informace měla objem několika megabytů. Autoři zde tedy měli na zřeteli skromnost Webu a jeho šetrnost k dostupným přenosovým kapacitám. Jak se později ukázalo, jejich předpoklad nebyl úplně správný – jedním z faktorů, které přispěly k dnes tak velkému prosazení Webu, byla zřejmě i jeho pozdější neskromnost a dosti vysoké nároky na přenosovou kapacitu a počítač, na kterém běží [29].

5.1.1 Hypertext

Základní vlastností, kterou Web dostal do vínku okamžitě při svém vzniku, byl koncept hypertextu. Ten úzce souvisí se způsobem lidského myšlení: ačkoli tato oblast není vědou ještě dostatečně prozkoumána, již dnes se tuší, že lidé nemyslí „lineárně“ – tedy ne tak, že by vždy vzali určitou myšlenku, tu domysleli do konce, a pak obdobným způsobem pokračovali další myšlenkou. Lidé myslí spíše „na přeskáčku“ – tedy tak, že uprostřed určité myšlenky přeskočí na jinou myšlenku, na základě podnětu vyvolaného nějakou asociací. Mechanismus vzniku těchto podnětů, resp. asociací také není dostatečně znám, ale nejspíše bude mít poněkud nedeterministický charakter, závislý na osobnosti, zkušenostech a mnoha dalších subjektivních faktorech (možná že právě tomu lidé vděčí za svou schopnost tvůrčího myšlení).

Hypertext je pak stejná možnost „přeskakování z myšlenky na myšlenku“, aplikovaná na textové informace – tyto se rozdělí na vhodně velké (či spíše malé) části, označované jako stránky, a mezi těmito stránkami se zavede možnost přeskoků. Cílem je umožnit čtenáři přijímat textové informace nelineárně, ve snáze „vstřebatelných“ celcích a s takovým postupem, který si díky možnosti přeskoků čtenář volí sám podle svého uvážení, znalostí, zájmu či jiných motivací [29].

Konkrétním mechanismem, který umožňuje přeskakování z jedné části textu na druhou, jsou tzv. aktivní odkazy. Nejlépe je si je představit jako ukazatele či směrníky, vedoucí z určitého výchozího místa k jinému, cílovému místu. Obě tato místa se přitom nejčastěji nachází v jiných textech, přesněji na různých stránkách (označovaných také jako WWW dokumenty či hypertextové dokumenty). Stejně tak se ale mohou nacházet i v rámci jedné a téže stránky, tj. odkazy mohou začínat i končit na téže WWW stránce.

Pokud odkazy směřují z jedné stránky „ven“ do jiné stránky, pak je velmi podstatné, že vůbec nezáleží na konkrétním umístění dotyčných stránek – obě se mohou nacházet „hned vedle sebe“, ve stejném adresáři na stejném WWW serveru, ale stejně tak se může každá z nich nacházet na opačné straně země, na různých WWW serverech. Pokud je tato možnost extrapolována do měřítka dnešního celoplanetárního Internetu, lze dojít k představě skutečně celosvětové soustavy WWW stránek, provázaných hustým předivem vzájemných odkazů. V této souvislosti se tedy hovoří o „celosvětové pavučině", a služba World Wide Web má pavučinu dokonce ve svém názvu.

Zajímavé a důležité je i to, v jakém smyslu jsou hypertextové odkazy aktivní. Podstata jejich aktivnosti souvisí s tím, že jednotlivé stránky si uživatelé prohlíží prostřednictvím vhodných prohlížecích programů (prohlížečů, klientů služby WWW), které se starají jak o správné zobrazení vlastního obsahu stránek, tak i o to, aby uživateli nabízely všechny odkazy vedoucí z dané stránky.

Výsledkem je posloupnost stránek, kterými se uživatel podle svého vlastního uvážení a zájmu postupně „proklikává“. Takovéto postupné procházení jednotlivými stránkami, označované jako tzv. brouzdání (anglicky: browsing), je přitom pro službu WWW natolik charakteristické, že dalo vzniknout i výstižnému neformální označení prohlížecích programů, neboli klientských programů služby WWW: jelikož slouží k brouzdání, jsou označovány obecně jako „browsery“ [29].

Jestliže World Wide Web původně vzniknul jako služba pro sdílení informací textové povahy, pak s postupem času se proměnil na obecnější zpřístupňující službu, která umožňuje „dát uživatelům přístup“ k informacím v mnoha různých podobách, nejen v podobě textové (psané). Důležitá je přitom role jednotlivých WWW stránek – ty jsou jednak nositeli vlastní informace, zpřístupňované uživateli, a jednak jsou

i nabídkami takovýchto informací. Rozdíl mezi „cílovými“ informacemi a jejich nabídkami však díky hypertextu nemusí být prakticky žádný – například když určitý text výkladového charakteru obsahuje hypertextové odkazy na jiné stránky věnované podrobnějšímu rozboru určitých dílčích otázek, je současně „cílovou informací“ i nabídkou [29].

5.1.2 Změny ve filosofii World Wide Web

Ruku v ruce s postupným zvyšováním uživatelské atraktivnosti služby WWW docházelo i k dosti výrazné změně celkové filosofie Webu obecně, a konkrétně pak jazyka HTML, ve kterém jsou zapisovány jednotlivé WW stránky. Původní filosofie jazyka HTML totiž byla taková, že o grafické podobě každé WWW stránky bude rozhodovat až ten, kdo ji bude skutečně zobrazovat, neboli až WWW prohlížeč (browser) na straně uživatele. Zajímavým vedlejším efektem pak byla úspornost celé této koncepce – po síti se přenášely pouze relativně malé objemy informací o tom, co je která část WWW stránky zač a jaký má význam (zatímco kdyby šlo o informaci vyjadřující přesné grafické ztvárnění stránky, byl by jejich objem velmi výrazně větší).

S postupem času se ale ukázalo, že právě popsaný a značně univerzální přístup neumožňuje dosáhnout takových efektů, jaké tvůrci WWW stránek stále více požadovali. Proto se s postupem času do jazyka HTML prosadily takové mechanismy, které umožňují již předem definovat jak mají jednotlivé prvky na WWW stránce přesně vypadat. Výsledkem jsou sice hezčí stránky, ale na druhé straně je jejich „zdrojový tvar“ čím dál tím více vázán na konkrétní platformu, a jejich zobrazení na jiných platformách způsobuje problémy (konkrétní předpis o grafické podobě nemusí být možné dodržet). Stejně tak se zvyšuje i objem dat, které jsou přenášeny v rámci přenosu jednotlivých WWW stránek.

Celkový trend je tedy takový, že služba WWWW sice zvyšuje své prezentační možnosti, ale na druhé straně se stává méně univerzální (na jiných platformách mohou stránky vypadat i dosti odlišně), a také výrazně náročnější na nejrůznější zdroje – jak na přenosovou kapacitu sítě, tak i na systémové zdroje počítače, na kterém běží uživatelův browser (zejména na paměť a výkonnost procesoru tohoto počítače).

Skutečně přelomovým momentem ve vývoji služby WWW však bylo něco jiného, než postupné „kvantitativní“ zvyšování uživatelské atraktivnosti. Šlo

o zavedení formulářů, a s tím související zpětné vazby od uživatele k poskytovateli informací. Do jejich příchodu měla služba WWW prakticky výhradně jednosměrný charakter. Byla stavěna tak, aby umožnila přísun informací od jejich poskytovatele směrem k uživateli (byť na zpřístupňujícím principu, tj. aktivitu musel projevit uživatel a „stáhnout“ si požadované informace). Neexistovala zde však obecnější zpětná vazba, neboli možnost kterou by uživatel mohl využít pro vznesení nějakého specifičtějšího požadavku na poskytovatele informací (například vyhledávacího dotazu). Jedinou, byť velmi rudimentární zpětnou vazbou, byla možnost odeslat zpět přesnou specifikaci stránky, kterou si uživatel přeje zobrazit jako následující.

Později se ale takováto obecná zpětná vazba zavedla. Pro její konkrétní využití však bylo nutné vytvořit i vhodný mechanismus, umožňují „převzít“ data od uživatele, a ty pak následně odeslat zpět k poskytovateli (přesněji: k WWW serveru, který je dále zpracuje). Tímto mechanismem se staly formuláře, vkládané přímo do jednotlivých stránek – lze je přirovnat k WWW obdobě běžných papírových formulářů, s různými vyplňovacími a zaškrtávajícími položkami. Uživatel tyto položky vyplní (resp. zaškrtne či ponechá nezaškrtnuté), a poté zmáčknutím na příslušnou ikonku (tlačítko) formulář odešle – což ve skutečnosti znamená, že jeho WWW browser vezme vyplněné údaje, a ty přenese zpět k WWW serveru, který stránku s formulářem původně poskytnul, a který takto získaná data může dále zpracovat.

Možnosti využití formulářů a zpětné vazby jsou doslova nepřeberné – data, poskytnutá uživatelem skrz vyplnění formuláře mohou být například vyhledávacím dotazem, který zpracuje vyhledávací mechanismus běžící na WWW serveru a následně poskytne uživateli svou odpověď. Bez takovéto možnosti by dnes těžko fungovaly nejrůznější vyhledávací služby, téměř nezbytné pro orientaci v Internetu. Dalšími příklady využití formulářů jsou nejrůznější ankety, dotazníky, průzkumy apod. [29].

Původní přístup k poskytování různých služeb na Internetu byl spíše takový, že každá služba byla koncipována jako samostatná a nezávislá na ostatních službách – zejména v tom smyslu, že uživatelé pro její využití potřebovali specifické klientské programy, a každý z nich se ovládal jiným, specifickým způsobem, který se uživatelé museli naučit. V akademickém Internetu to ještě nebylo tolik na závadu, ale ve světě komerčního Internetu už ano.

Trend ke sjednocování služeb na platformě Webu sleduje velmi důležitý cíl, tedy, aby uživatelé mohli používat jeden nástroj (WWW prohlížeč) a jeden styl práce (brouzdání), i když chtějí využívat různé služby a mít přístup k různým informačním zdrojům a informacím v nejrůznějších podobách. Způsob, jak tohoto cíle dosáhnout, však není omezen jen na použití „přizpůsobujících“ bran. Může mít více různých podob – například tu, která dává možnost vkládat do WWW stránek objekty nejrůznějších typů, a WWW browserům dává schopnost s těmito objekty potřebným způsobem pracovat [29].

Skutečně koncepční řešení problému, přináší až myšlenka tzv. proveditelného obsahu. Je založena na možnosti, aby si jednotlivé objekty doslova „nesly sebou“ vše, co je zapotřebí k jejich zpracování – tedy například k jejich zobrazení, přehrání, vypočítání, spuštění apod.

Až do nástupu „proveditelného obsahu“ se předpokládalo, že po síti budou v rámci jednotlivých stránek přenášena pouze data, s tím že vše potřebné pro jejich zpracování bude již dopředu připraveno na místech, kam data směřují – tj. na koncových počítačích. Ty ale kvůli tomu značně mohutněly, stávaly se komplikovanějšími a čím dál tím složitějšími na své udržování a správu. Naproti tomu možnost „přibalit“ k datům i vše potřebné pro jejich zpracování dovoluje dosáhnout pravého opaku – používat v roli pracovních stanic jednotlivých uživatelů relativně jednoduché počítače, které si vše potřebné ke svému fungování „stáhnou“ ze sítě [29].

S odstupem času se servery staly postupně stále výkonnějšími tahouny chodu většiny obchodních organizací, neboť na ně bylo nakládáno stále více a více úkolů. Byly to zpravidla nejmodernější systémy provozované špičkovým hardwarem. Jejich správci je navíc stále dolaďovali, aby bylo dosahováno stále lepších a lepších výkonů [22].

5.1.3 HyperText Markup Language

Filosofii jazyka HTML nejlépe vystihuje přívlastek „Markup“ v jeho názvu, a který je možné přeložit jako „značkovací“. Jazyk HTML skutečně funguje tak, že „značkuje“ části textu, neboli opatřuje konkrétní části textu značkami, které vypovídají o jeho významu. Příkladem může být uvození a zakončení určité části textu značkou označující, že daná část textu je nadpisem (dokonce v jedné ze šesti možných hierarchických úrovní), nebo že má být zobrazena zvýrazněně apod.

Stejným způsobem jsou pak řešeny i odkazy – má-li například určitá část textu být odkazem na jinou stránku, je také „obalena“ příslušnými značkami (tagy), které přesně specifikují kam odkaz vede. Analogicky, tedy pomocí značek, je pak řešeno i vkládání obrázků a dalších objektů, které mohou být součástí stránky. Důležité přitom je, že samotný jazyk HTML specifikuje to, co jednotlivé části WWW stránky znamenají, a nikoli to, jak mají být zobrazeny – o jejich konkrétním zobrazení tedy rozhoduje až konkrétní browser, který má stránku zobrazit. Ten se pak také může sám rozhodnout, jak zobrazí nadpis nejvyšší úrovně, jak bude zobrazovat zvýrazněný text apod. Proces vytváření grafické podoby WWW stránky (na základě jejího HTML tvaru) je označován jako tzv. rendering[9]. Tato základní filosofie jazyka HTML usnadňuje zobrazování WWW stránek na různých platformách, s různými grafickými a prezentačními možnostmi, ale na druhé straně dosti omezuje autory WWW stránek v tom, aby uskutečnili svůj autorský záměr, např. pokud jde o celkové grafické ztvárnění [29]. HTML jazyk se tedy skládá ze značek, atributů a hodnot atributů[10]. Značky lze z hlediska významu rozdělit na tři základní skupiny:

- strukturální značky – rozvrhují strukturu dokumentu
- popisné (sémantické) značky – popisují „povahu“ obsahu elementu
- stylistické značky – určují vzhled elementu při zobrazení

XHTML[11] je jiná, novější norma jazyka HTML. HTML jako takové už se nevyvíjí a zůstalo ve verzi HTML 4.01. Konsorcium W3C[12] specifikovalo značkovací jazyk HTML v poslední verzi 4.01 24. října 1999[13] a značkovací jazyk XHTML[14] ve verzi 1.0 SE[15] 1. srpna 2002. Podstatné je, že podpora jazyka XHTML je v současných prohlížečích naprosto stejná jako podpora HTML. Ačkoli se usuzuje, že v budoucnu bude podpora XHTML lepší než podpora HTML, na základě zkušeností s historickým vývojem prohlížečů není důvod se domnívat, že tomu tak vskutku bude [20].

[9] česky: vizualizace
[10] např.: párová značka pro tabulku <table width=”100%”> , table – značka, width – atribut, 100% – hodnota atributu
[11] eXtensible HyperText Markup Language
[12] World Wide Web Consorcium – mezinárodní konsorcium, jehož členové společně s veřejností vyvíjejí webové standardy pro WWW
[13] http://www.w3.org/TR/html4/
[14] http://www.w3.org/TR/xhtml1/
[15] Second Edition – druhé vydání – původní specifikace XHTML byla vydána 26. ledna 2000, revize právě 1. srpna 2002

HTML dokument lze napsat tak, aby vyhovoval pravidlům pro XML[16] dokumenty a také ho lze napsat tak, aby těmto pravidlům nevyhovoval. Tento postup však vůbec neovlivní vzhled webové stránky. Prakticky každou webovou stránku lze v zásadě opravit tak, aby vyhovovala pravidlům pro XML, aniž by to vůbec kdo na vzhledu poznal. Což je velmi příznivé. Prakticky tedy napsání HTML stránky takovým způsobem, aby vyhovovala i standardu XML nijak neomezí webové designery ve své činnosti. Stačí pouze dodržovat o něco striktnější pravidla při psaní HTML. Proto vznikl standard XHTML, který – zjednodušeně řečeno – popisuje, jaká pravidla musí být dodržována při psaní HTML, aby bylo učiněno zadost i standardu XML. Bude-li tvůrce při psaní webových stránek brát v úvahu XHTML, má zaručeno, že výsledný dokument vyhovuje jak standardu HTML, tak i XML [30].

XHTML se v praxi vyskytuje ve třech verzích:

- XHTML 1.0 transitional (přechodové)
- XHTML 1.0 strict (striktní)
- XHTML 1.0 frameset (podpora rámů)

Zatímco XHTML 1.0 transitional by neměl autorům WWW prezentací dělat větší problémy a lze jej vcelku doporučit k používání, XHTML 1.0 striktní je velmi náročné.

Zásadní rozdíly mezi HTML a XHTML jsou:

- všechny atributy mají hodnoty v uvozovkách (platilo již v HTML, XHTML však striktně vyžaduje)
- zákaz křížení značek (platilo již v HTML, XHTML však striktně vyžaduje)
- značky a atributy jsou psány malými písmeny
- nepárové značky končí lomítkem
- párové značky jsou párové povinně
- všechny atributy musejí mít hodnotu
- interní JavaScript[17] a CSS[18] se zapisují rozdílným způsobem

[16] eXtensible Markup Language – značkovací jazyk pro uchovávání a předávání dat ve strukturované a semistrukturované formě
[17] klientský skriptovací jazyk zapisovaný přímo do (X)HTML kódu WWW stránky

- dokument má mít XML prolog
- dokument požaduje správný DOCTYPE[19]

(X)HTML dokument se skládá ze dvou základních částí, z hlavičky, která se nachází mezi značkami `<head>` a `</head>` a vlastního těla dokumentu, nacházejícího se mezi značkami `<body>` a `</body>`. Hlavička (X)HTML dokumentu sděluje prohlížeči přehled základních informací, např. informace o použití znakové sady, popis obsahu dokumentu, charakterizování klíčových slov, apod. V těle (X)HTML dokumentu je vlastní obsah WWW stránky [2].

```
<?xml version="1.0" encoding="utf-8"?>

<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">

<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" >

<head>

          <title>ICT jako významný faktor konkurenceschopnosti</title>
          <meta http-equiv="Content-Type" content="text/xhtml; charset=utf-8" />
          <meta http-equiv="Content-Language" content="CZ" />
          <meta name="Description" content="Disertační práce" />
          <meta name="Keywords" content="ICT, WWW, SEO, SEM, vyhledávač, Google"
lang="CZ" />
          <meta name="Author" content="Pavel Šimek – KIT PEF ČZU v Praze,
http://pef.czu.cz/~simek" />
          <meta name="ICT jako významný faktor konkurenceschopnosti" content="(C) 2007
Pavel Šimek, All rights reserved." />

</head>

<body>

        ...

        ...

        ...

        ...

</body>

</html>
```

Příklad 1 – základní struktura XHTML dokumentu.

5.1.4 Domény a Unique Resource Locator

Doména nejvyššího řádu (TLD[20]) je internetová doména[21] na nejvyšší úrovni stromu internetových domén. V doménovém jméně je doména nejvyšší úrovně na konci (např. u `www.pef.czu.cz` je doména nejvyššího řádu `.cz`). TLD popisuje základní skupinu doménových jmen, např. všechna doménová jména daného

[18] Cascading Style Sheets – kaskádové styly – kolekce metod pro grafickou úpravu webových stránek
[19] specifikace DTD – Document Type Definition – definice typu dokumentu
[20] Top Level Domain
[21] doménové jméno – jednoznačné jméno (identifikátor) počítače nebo počítačové sítě, které jsou připojené do internetu

státu. Domény nejvyššího řádu jsou pevně stanoveny internetovou standardizační organizací IANA[22]. TLD existují následujících tří typů:

- Národní TLD (country-code TLD, ccTLD), sdružující doménová jména jednoho státu. Jejich název je dvoupísmenný (pro Českou republiku je to .cz). V každém státu určuje příslušný správce národní domény (v ČR je to CZ.NIC, z. s. p. o.) pravidla, podle kterých se doménová jména v dané TLD přidělují. Některé státy umožňují registraci jména v národní TLD libovolnému zájemci, bez ohledu na to, jestli má s příslušným státem něco společného. Mezi takové státy a území patří i Česká republika se svou doménou `.cz`, dále také např. Arménie (`.am`), Itálie (`.it`), Jersey (`.je`), Německo (`.de`), Rakousko (`.at`), Západní Samoa (`.ws`), Tonga (`.to`), apod. To umožňuje existenci „zajímavých“ doménových jmen, např. `coje.to`, `I.am` (anglicky jsem), `start.at` (anglicky začít u), nebo `go.to` (anglicky jít na). Jiné státy a závislá území povolují registraci domény v příslušné TLD pouze občanům či podnikům té které země. Příkladem je Kanada (`.ca`) nebo Slovensko (`.sk`).
- Generické TLD (generic TLD, gTLD), sdružující obecné domény (např. `.org` pro neziskové organizace), nespojené s jedním konkrétním státem (až na výjimky `.mil` a `.gov`, které jsou z historických důvodů vyhrazeny pro vojenské, resp. vládní počítačové sítě USA). Její název je nejméně třípísmenný.
- Infrastrukturní TLD využívané pro vnitřní mechanismy Internetu. V současné době existuje jediná taková doména nejvyšší úrovně, doména `.arpa`.

Doména druhé úrovně je doménové jméno v určité doméně první úrovně (např. doménové jméno České zemědělské univerzity v Praze v národní TLD `.cz` je

[22] Internet Assigned Numbers Authority - Pro praktické fungování celosvětového Internetu jsou nutné i nejrůznější parametry, adresy a jiné číselné identifikátory, které musí mít stejný význam v rámci celého internetu, nebo musí být v celém internetu unikátní (použity jen jednou). V Internetu proto musí existovat centrální autorita (centrální subjekt), který bude používání a přidělování takovýchto číselných hodnot zastřešovat a koordinovat. Touto organizací je v současné době organizace jménem IANA, a ve skutečnosti jde o malé středisko při jedné z univerzit v USA, které stále ještě funguje z grantů vlády USA.

czu.cz). Doména třetí úrovně je již subdoména v doménovém jménu, např. pef.czu.cz nebo slovnik.seznam.cz).

5.1.5 Diakritika v doménách

Systém DNS[23], díky kterému lze v Internetu používat symbolická doménová jména (jako např. www.czu.cz) místo číselných adres (např. 193.84.33.251), vznikl již relativně dávno, v roce 1984. Počítal přitom s tím, že dílčí složky doménových jmen (tzv. labely, jako např. „czu“ ve www.czu.cz) bude možné vytvářet pouze z tzv. čistých ASCII[24] znaků, a to ještě ne všech – používat se smí pouze písmena, číslice a pomlčka, která ještě navíc nesmí být na začátku ani na konci, a nesmí také sousedit s jinou pomlčkou.

Dlouhou dobu takovéto řešení plně postačovalo, ale s rozvojem Internetu se situace změnila. Především v zemích, jejichž národní abecedy jsou úplně jiné, či alespoň mají řadu specifických znaků, vzrůstalo volání po rozšíření celého systému tvorby doménových jmen. Například v České republice by se mohla hodit možnost používat ve jménech domén háčky a čárky. Symbolická doménová jména by se tak mohla lépe přizpůsobit pravidlům češtiny, byla by o něco lépe zapamatovatelná, a například v reklamách by se URL[25] odkazy mohly psát i s diakritikou. V současné době je tento přístup používán, ale takovéto odkazy nefungují, a ne každý, koho reklamy oslovují, si to musí hned uvědomit. V určitých případech však může být situace mnohem horší, zejména u národů, které vůbec nepíší latinkou a nemají tedy „jen některé“ znaky, které ve jménech svých domén použít nemohou.

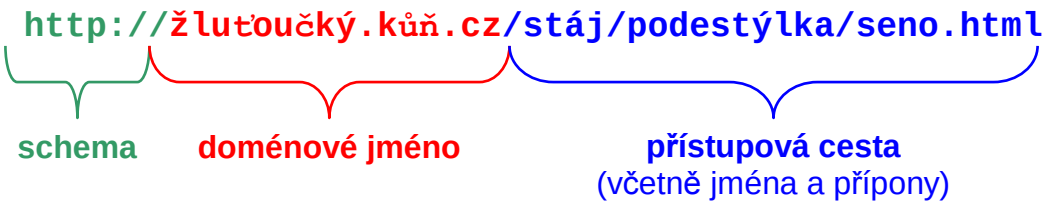
Zavedení podpory pro diakritiku (obecně pro celé národní abecedy) by také významně rozšířilo prostor pro vytváření nových domén národních znaků. Pro doménové registrátory by to jistě bylo vítané rozšíření segmentu jejich podnikání. Pomoci by to mohlo i některým zájemcům o domény, kteří až dosud si nemohli registrovat doménové jméno, protože odpovídající doménu bez diakritiky vlastní již někdo jiný [29].

Způsobů, jak do doménových jmen zabudovat nejen českou diakritiku, ale i podporu jiných národních znaků a celých „cizokrajných“ abeced, může existovat

[23] Domain Name System
[24] American Standard Code for Information Interchange – americký standardní kód pro výměnu informací – kódová tabulka definující znaky anglické abecedy a jiné znaky používané v informatice
[25] Unique Resource Locator – jednoznačné určení zdroje – způsob, jak jednoznačně zapsat umístění souboru na Internetu nebo na intranetu

celá řada. Za zmínku jistě stojí i důležitý fakt, že pro podporu diakritiky (národních abeced) ve jménech domén dnes existuje řádný standard (IDN[26]), který přesně říká co a jak je třeba udělat [29]. Důležité je, že již od března 2004 má podobu platného internetového standardu (RFC[27] 3490, 3491, 3492 a 3454). Pokud se tedy nějaká instituce (správce centrálního registru doménových jmen) rozhodne podporu diakritiky (národních abeced) implementovat, ví jak to má udělat. Samozřejmě to ještě neznamená, že IDN vůbec musí implementovat.

Celý koncept IDN se týká pouze symbolických doménových jmen. Nepokrývá tedy úplně vše, co by v souvislosti s národními abecedami mělo být vyřešeno a standardizováno, aby bylo možné bezproblémově používat například diakritiku v celých internetových adresách. V tzv. URI[28], které např. identifikují jednotlivé WWW stránky, se totiž vedle doménového jména (a tzv. schématu, které říká zda jde o WWW stránky či jiný typ objektu) vyskytuje mj. i přístupová cesta.



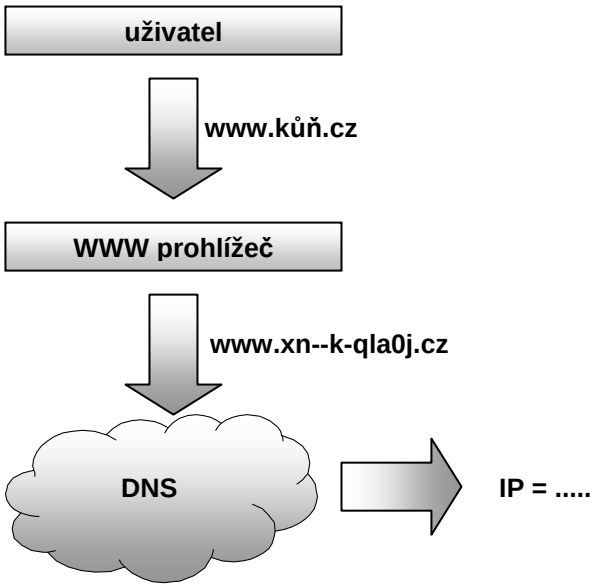
Obrázek 1: Schéma URI [29].

Princip IDN spočívá v překladu doménových jmen na číselné IP adresy. Tento překlad zajišťuje systém DNS, který si lze představit jako rozsáhlou distribuovanou databázi, obsahující informace typu *stroj se jménem "žluťoučký" v doméně "kůň.cz" (tj. "žluťoučký.kůň.cz") má IP adresu 123.321.111.222.* Lidé, kteří vymýšleli IDN, mohli navrhnout změnu fungování celé této obrovské distribuované databáze, resp. všech tzv. name serverů, které ji tvoří, tak aby uměly pracovat s diakritikou i s národními abecedami. Ovšem aktualizace takého množství individuálních prvků obrovského systémy by byla velmi obtížná.

[26] Internationalized Domain Names – standard, podle kterého mohou doménová jména obsahovat i znaky národních abeced

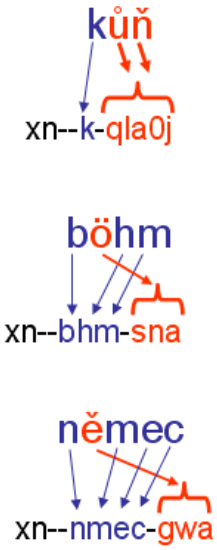
[27] Request For Comments – žádost o komentáře - se používá pro označení řady standardů a dalších dokumentů popisujících internetové protokoly, systémy apod. Jak už název napovídá, RFC jsou oficiálně považovány spíše za doporučení než normy v tradičním smyslu, přesto se podle nich řídí drtivá většina Internetu.

[28] Uniform Resource Indicator – jedinečná identifikace zdroje – obecně použitelná množina na všechny druhy adres, URL je zkratka pro Uniform Resource Locator a je podmnožinou URI



Obrázek 2: překlad národního tvaru domény na IP adresu [29].

Národní tvar doménového jména, který může obsahovat nejrůznější ne-ASCII znaky, je přeložen do takového tvaru, který již obsahuje jen samé ASCII znaky. Proto se také označuje jako ACE[29]. Tento „ACE-tvar“ doménového jména již je stávající systém DNS schopen zpracovat, tj. překládat doménová jména v tomto tvaru na číselné IP adresy. Vzhledem k tomu, že celé IDN je řešeno jako nadstavba nad stávajícím DNS, je překlad do ACE tvaru řešen plně v klientském programu, se kterým pracuje uživatel. Tedy například v jeho browseru. Do DNS pak vstupuje (k překladu) již přímo ACE tvar doménového jména.



Obrázek 3: Princip překladu do ACE tvaru [29].

[29] od: ASCII Compatible Encoding

Příslušný způsob kódování je označován jako „punycode“, a jeho princip spočívá v:

- přeložený (ACE) tvar vzniká připojením pevně daného prefixu ("`xn--`"), který signalizuje že jde o IDN jméno
- všechny čistě ASCII znaky z původního tvaru se ponechají beze změny
- jiné než ASCII znaky se zakódují pomocí čistě ASCII znaků

Kromě právě naznačeného překladu „punycode“ je součástí IDN ještě jedna úprava „národního tvaru“ (tzv. nameprep). Existuje kvůli tomu, že „národní tvar“ může mít více různých podob, a nemusí být úplně triviální je mezi sebou převádět, resp. najít jeden kanonický tvar, který je pak převáděn do ACE tvaru. U jmen s českou diakritikou to ještě jde, takže se „nameprep“ redukuje prakticky jen na převod z velkých písmen na malá (např. jméno „NĚMec“ převede na „němec“). V jiných jazycích, třeba čínštině či japonštině, ale může být „nameprep“ velmi obtížné.

Nejjednodušší přístup registrace domény s národními znaky by mohl znít, že je nutné zaregistrovat doménu přímo v ACE tvaru. Například zájemce o doménu „`böhm.de`“ by si fakticky zaregistroval doménu „`xn--bhm-sna.de`“. Zájemce o doménu „`böhm.ch`“ zase doménu „`xn--bhm-sna.ch`“, zájemce o doménu „`němec.cz`“ doménu „`xn--nmec-gwa.cz`“ atd.

Ovšem tam, kde správce národní domény IDN nepodporuje, by registrace domény v „punycode“ neměla být možná. Klasická doménová jména by totiž neměla obsahovat dvě pomlčky těsně za sebou a například v České republice CZ.NIC z. s. p. o.[30], stejně jako většina ostatních správců národních domén, má toto (respektive do zavedení IDN měla) explicitně zakázáno ve svých pravidlech pro tvorbu doménových jmen. Díky tomu by se nemělo stát, aby jakékoli nové jméno domény (podle IDN) kolidovalo s nějakou již existující (ne-IDN) doménou. Stejně tak by se tímto opatřením (zavedením prefixu "`xn--`" mělo zabránit tomu, aby do zavedení IDN někdo dopředu obsadil budoucí IDN domény [29].

5.2 Vyhledávače

Dokud byl Internet ještě zcela akademický, mohl si dovolit požadovat po svých uživatelích poměrně velké znalosti a dovednosti. Proto také v tehdejší době

[30] zájmové sdružení právnických osob – správce centrálního registru doménových jmen v TLD .cz

existovala poměrně široká paleta samostatných služeb s vlastním mechanismem fungování, a od uživatelů se očekávalo že se je všechny naučí používat. Když pak Internet přešel do rukou komerční sféry a mezi jeho uživateli převážili lidé stojící mimo akademickou sféru – tedy například podnikatelé, úředníci, ale i ženy v domácnosti – bylo nutné značně zjednodušit způsob práce s Internetem. To znamenalo jak redukci příliš velkého počtu různých služeb, tak i prosazení takových služeb, které jsou velmi intuitivní a pro uživatele snadno použitelné. Díky tomu se prosadil dobře známý World Wide Web, který v dnešním Internetu jednoznačně dominuje.

Časem se ale ukázalo, že ani takovéto zjednodušení Internetu a práce s ním není dostatečné – důvodem je stále ještě značná roztříštěnost nejrůznějších zdrojů a jejich malá přehlednost, uspořádanost a utříděnost. Dnešní typický uživatel se nedokáže sám dostatečně orientovat v celé obrovské šíři toho, co Internet nabízí: někdy proto, že neví jak a kde hledat, jindy proto že v doslovné záplavě informací nedokáže ani rozlišit to, co ho skutečně zajímá. Nebo jednoduše nemá čas něco pracně vyhledávat [29].

Webové prezentace se s nezpomalujícím rozvojem Internetu dostávají stále více do popředí zájmu právě z hlediska získávání informací. Zatímco na Zemi žije přes 6,7 miliardy lidí, jenom na serverech vyhledávací služby Google[31] je zaindexováno více než 16 miliard[32] internetových stránek a každý den vznikají milióny nových. Některé vznikly, aby prezentovaly osobní názory, znalosti a dovednosti, více či méně umělecké výtvory jednotlivců i zájmových skupin, další slouží jako zdroje informací, novinek a vědeckého poznání. Některé existují pouze pro zábavu, jiné mají i ekonomický význam, když slouží jako prostředek komunikace a v neposlední řadě jako nabídka nebo přímo odbytiště zboží a služeb firem.

Enormně vysoký počet existujících webových stránek však ztěžuje orientaci a omezuje možnosti nalezení relevantních informací. Na jakékoliv webové stránky – k jakékoliv informaci – lze zavítat dvojím způsobem: buď zadáním jejich přesné adresy do adresního řádku prohlížeče nebo následováním odkazu z jiné stránky [42].

[31] největší světový internetový vyhledávač

[32] Zatímco do poloviny roku 2006 byl počet stránek indexovaných vyhledávačem Google uváděn na titulní stránce (1.3.2003 to bylo 3 083 324 000 stránek, 15.3.2005 to bylo již 8 392 627 000 stránek), v současnosti (červen 2007) není přesný počet stránek zveřejňován.

První způsob používá malé procento návštěvníků Internetu, ale je nedílnou součástí vyhledávání. Spočívá v přímém zadání hledaného slova nebo fráze bez diakritiky doplněné o národní doménu. V zájmu provozovatelů webů je proto vhodné mít zaregistrovanou snadno zapamatovatelnou doménu, která přímo odpovídá jejich názvu (např. `vzp.cz` pro největší zdravotní pojišťovnu v České republice) nebo mít zaregistrovaný obecný výraz (např. `dovolena.cz` pro cestovní kancelář). Ačkoliv náhodný pokus o zadání tvoří pouze neodhadnutelný zlomek (ze statistik ani serverových logů nelze poznat, zda se uživatel adresu někde dozvěděl a pamatoval nebo právě vymyslel).

Většina uživatelů však volí druhý způsob, který je jistější z hlediska dosažení výsledku. Rozhraním, zprostředkujícím návštěvníkům Internetu možnost následovat odkaz vedoucí k požadovaným informacím, jsou potom vyhledávací servery, vyhledávače (anglicky search engines).

První vyhledávač existoval ještě před vznikem samotného World Wide Web,. Jednalo se o službu Archie, naprogramovanou Alanem Emtagem z Montrealské univerzity v roce 1990 a sloužil k prohledávání FTP serverů. Archie vytvářela pouze databázi názvů a umístění souborů, ale s celosvětově prudce rostoucím počtem serverů se těšila velké popularitě.

V následujících letech došlo k rozvoji WWW a světlo světa spatřil první prohlížeč Mosaic. Krátce poté vznikl i první robot s názvem World Wide Web Wanderer, který vytvořil Matthew Gray. Jeho úkolem bylo počítat WWW stránky a měřit tak rychlost růstu Internetu. Do konce roku 1993 vzniklo několik dalších vyhledávačů s vlastními roboty – World Wide Web Worm, JumpStation a RBSE[33], ale ty indexovaly pouze hlavičky a URL adresy stránek.

V roce 1994 vytvořili Jerry Yang a David Filo ze standfordské univerzity katalog Yahoo!. Původní myšlenkou bylo sestavit přehledný seznam jejich oblíbených odkazů. Ve stejném roce pak vznikl i WebCrawler, robot Briana Pinkertona z University of Washington, který indexoval kompletní obsah stránek. Následovaly vyhledávače Excite, Altavista, AskJeeves a jiné. Jak vznikaly další a další internetové stránky, stávaly se vyhledávače používanější a populárnější [35].

[33] Repositury Based Software Engineering

Spolu s rostoucím objemem stránek však klesala relevance výsledků vyhledávání. Tento problém vyřešili až ke konci roku 1998 Larry Page a Sergey Brin fulltextových vyhledávačem Google, který kromě souladu s hledaným slovem nebo frází řadil výsledky podle toho, kolik odkazů na vyhledané stránky vedlo.

Google „neusnul na vavřínech“, svůj řadící algoritmus neustále vylepšoval a jeho popularita rostla, až se stal celosvětově nejrozšířenější. V posledních letech se kromě zdokonalování algoritmů zjišťujících relevanci na hledaná klíčová slova[34] věnuje především tzv. lokálnímu vyhledávání (roli hraje umístění sídla nebo působiště provozovatele webu) a personalizovanému vyhledávání, které se snaží odhadnout pořadí specificky podle přednastaveného a vyvíjejícího se profilu návštěvníka.

Dohnat náskok tržního podílu Google zkouší se svým MSN společnost Microsoft, která si pomáhá jeho integrací do OS Windows a prohlížeče Internet Explorer, a skupováním konkurence se chce na špičku vrátit i Yahoo!.

5.2.1 Katalogy a vyhledávače

Katalog, popř. portál, je web, který obsahuje celou řadu odkazů na jiné webové stránky, zpravidla řazené do tématických oblastí. Mezi nejznámější české katalogy patří Seznam, Centrum a Atlas, mezi zahraniční Yahoo! a ODP.

Záznam do katalogu je prováděn registrací do příslušné sekce (ta která se věnuje danému tématu) a po zkontrolování editorem příslušné sekce je záznam zařazen do databáze katalogu. V katalogu je vyhledávání prováděno buď uživatelovým procházením jednotlivých sekcí anebo pomocí jednoduchého dotazu do vyhledávacího políčka. Pokud uživatel vyhledává pomocí dotazu, katalog porovnává zadaný dotaz s URL, nadpisy a popisky stránek, které jsou v katalogu registrovány a také s kategorií do níž je odkaz zařazen. Pokud je stránka například v kategorii Inzertní servery je pravděpodobné, že ji katalog při vyhledávání slova inzerce přidá „plusové body“.

Možností jak ovlivnit pozici stránky při vyhledávání v katalogu je tedy hlavně klíčové slovo v nadpisu stránky (popisek má nižší i když nezanedbatelnou váhu). Problémem ve většině případů je, že autoři WWW prezentací dosti často ponechávají v nadpisu pouze název firmy nebo URL. Výhodu v katalogu tak v drtivé většině mají

[34] slovo či termín, které má z hlediska tématu stránky mimořádný význam

webové stránky, které mají obecný název domény (`knihkupectvi.cz`, `bazar.cz`, `pojistovna.cz`, apod.).

V popisku stránky je tedy vhodné zadávat co nejvíce klíčových slov, které mají s daným tématem něco společného. Je tedy nutné, aby se autoři webů vyhnuli slovům typu: internetové stránky, nejrychlejší, nejlepší apod. Editoři katalogů je z názvu a popisku většinou odstraňují, protože nepřináší žádnou užitečnou informaci. Klíčová slova a fráze v popisku se zvažují stejně důkladně jako v nadpisu. Katalogy v drtivé většině totiž neumí skloňovat ani hledat podobná slova. Co není v přesném tvaru jak uživatel vyhledává, katalog nenajde. Samozřejmě popisek musí být složen z vět a ne jen z prostého výčtu klíčových slov. Takovou stránku by editor sekce do katalogu zařadil.

V katalogu může mít web v jedné kategorii vždy pouze jeden zápis (nebo by alespoň měl), avšak registrování podstránek, které mají unikátní obsah odlišný od již zaregistrované webové stránky je v jiné sekci většinou povoleno. Co všechno může být v nadpisu, v popisku a do kolika sekcí je možné stránku zařadit mají katalogy napsané v obecných pravidlech registrace. Zlepšování pozice v katalozích je nazýváno optimalizace pro katalogy.

Oproti katalogům má každý fulltextový vyhledávač svého robota (software), který hledá a stahuje celé dokumenty (webové stránky, dokumenty textových editorů, PDF soubory, apod.) na Internetu a následně je indexuje (zpracovává a ukládá) do své databáze.. V této databázi pak umožňuje návštěvníkům svých webových stránek vyhledávat pomocí jednoduchých i složitých dotazů. Mezi nejznámější zahraniční vyhledávače patří Google a Yahoo!, mezi české pak Jyxo a Morfeo.

Každý fulltextový vyhledávač se tedy skládá ze dvou částí. První část je robot (také se nazývá crawler, spider, bot nebo pavouk) a druhá část je webové rozhraní. Robot má na starosti procházení webu, stahování souborů a jejich indexaci. Robota lze dále dělit ještě na getter, který stahuje soubory a na indexer, který má za úkol soubory zpracovávat a ukládat do databáze. Metody používané jednotlivými vyhledávacími nástroji se liší, ale indexující program v podstatě prohlédne různé součásti stránky, včetně jejího titulu, obsahu jejích meta značek, komentáře, názvů odkazů, textu v nadpisech a hlavní části textu [33]. Robot se po webu pohybuje úplně samostatně a nelze mu nařídit jak často má na stránku chodit (vyhledávači lze samozřejmě zaplatit

za pravidelnou indexaci), ale pouze to, které stránky nesmí indexovat. Zároveň si pamatuje kde byl a v jeho algoritmu je naprogramováno kdy se na danou stránku má opět vrátit.

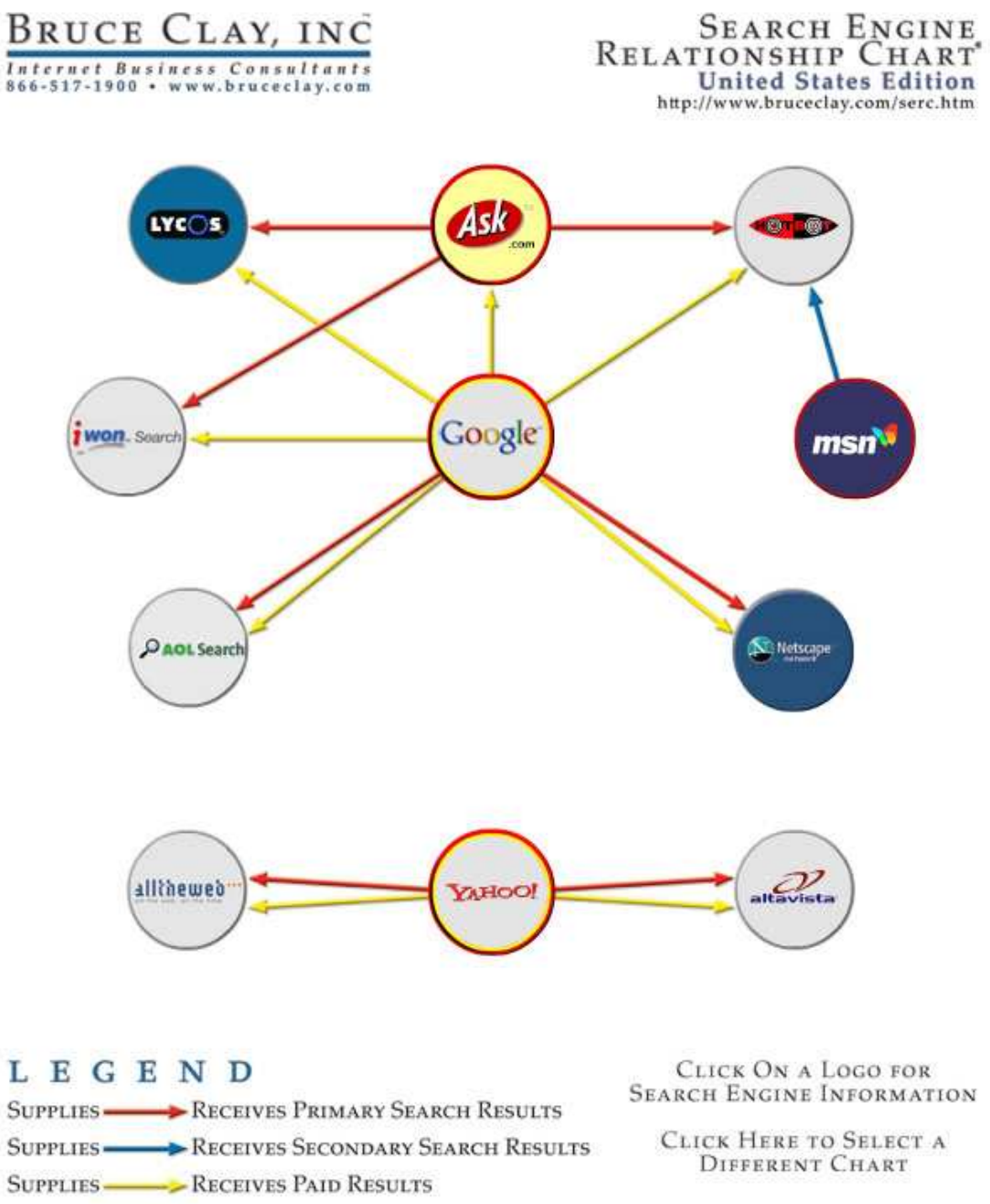
Do vyhledávačů je možné registrovat stránky, ale většinou takové úsilí postrádá smysl. Robot vyhledávače je schopen stránku najít sám a registrace bez toho aby na stránku vedl nějaký odkaz většinou nepomůže. Výjimku tvoří české stránky, které mají národní doménu jinou než `.cz` (například `.com` nebo `.net`). Takové stránky české vyhledávače (Jyxo[35], Morfeo) automaticky indexují na základě registrace do některého z katalogů (například pokud je `www.example.com` registrován v katalogu Atlas, je vyhledávačem Jyxo považován za českou stránku a indexován). Pokud stránka není registrovaná v katalozích je možnost ji zaregistrovat přímo do vyhledávače[36].

Při vyhledávání je na webové stránce vyhledávače pokládán dotaz, který vyhledávač porovnává se svou databází. Jelikož nikdo, kromě tvůrců příslušného vyhledávače, nezná přesný algoritmus podle něhož se řadí výsledky vyhledávání, lze se jen domnívat, že je to v současné době u moderních vyhledávačů kombinace toho kolikrát a v jaké html značce se hledaný výraz (klíčové slovo) na stránce nachází a popularity stránky. Popularita stránky je označován její ranking, Google používá PageRank, Seznam S-Rank.

Celá řada vyhledávačů ve skutečnosti nejsou vyhledávače, i když se tak na první pohled tváří. Pouze používají cizí technologii a přebírají výsledky, které pak kombinují s katalogem (nejčastěji s ODP1[37]). Současnou situaci na trhu nejlépe vykresluje obrázek č. 4 od Bruce Clay. Jedná se v postatě o outsourcing vyhledávacích a katalogových služeb [35].

[35] Jyxo hledá od května 2004 i české stránky, které nemají doménu .cz a nejsou v katalogu Atlas
[36] http://jyxo.cz/d/submit

[37] Open Directory Project

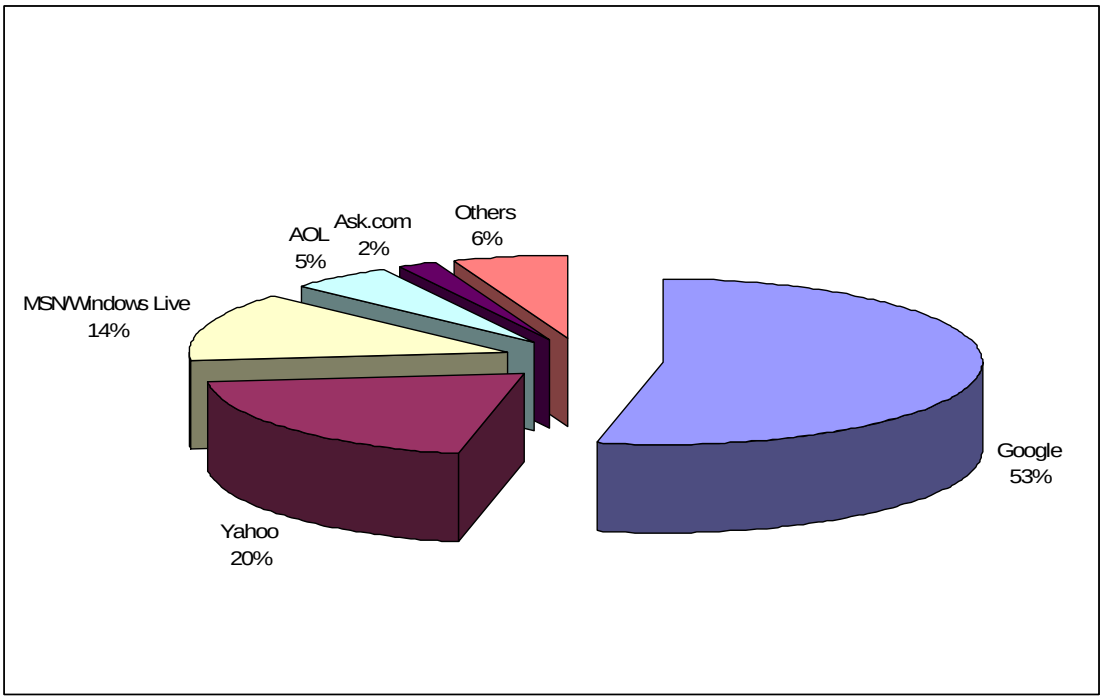


Obrázek 4: Vztahy mezi předními světovými vyhledávači [6].

Vzhledem k tomu, že návštěvníci vyhledávacích serverů používají v různých situacích oba výše popsané způsoby, integrovala postupem času většina vyhledávacích serverů oba mechanismy a nabízejí vyhledávání pomocí katalogu i fulltextu.

5.2.2 Nejpoužívanější vyhledávače a katalogy

Dle zveřejněné zprávy „Top 10 Search Providers, July 2007“ [7] je dle počtu vyhledávání anglicky mluvících uživatelů nejpoužívanějším vyhledávačem Google.



Graf 1: Nejpoužívanější světové vyhledávače (dle počtu vyhledávání - červenec 2007) [7].

Současní největší hráči na trhu tedy jsou: Google, Yahoo!, Microsoft a ASK. O AOL, které používá technologii Google, se spekuluje, že bude koupeno některým „hráčem z velké trojky“. Podíl vyhledávačů je samozřejmě obtížně měřitelný a existují různé přístupy. O tom, že ale globálně vede Google žádné spory nejsou. Od dubna 2007 do července 2007 sice Google necelá 2 procenta z celkového počtu vyhledávání ztratil, absolutní počet vyhledávání však stoupl o více než 370 mil [7] [8].

Poskytovatel	IV.07		VII.07		Změna v tis.	% změny
	Počet vyhledávání v tis.	Podíl z celkového vyhledávání (%)	Počet vyhledávání v tis.	Podíl z celkového vyhledávání (%)		
Google	3 773 032	55,18%	4 143 752	53,31%	370 720	9,83%
Yahoo	1 497 154	21,90%	1 559 745	20,07%	62 591	4,18%
MSN/Windows Live	612 526	8,96%	1 057 064	13,60%	444 538	72,57%
AOL	371 038	5,43%	407 988	5,25%	36 950	9,96%
Ask.com	126 286	1,85%	143 513	1,85%	17 227	13,64%
Ostatní	457 657	6,69%	461 112	5,93%	3 455	0,75%
Celkem	6 837 693	100,00%	7 773 174	100,00%	935 481	13,68%

Tabulka 1: Počty vyhledávání za duben a červenec 2007 [7] [8].

České vyhledávače se vyvíjely se zpožděním za těmi světovými. V roce 1997 naprogramoval Ivo Lukačovič a Štěpán Škrob fulltextový vyhledávač Kompas, který používal katalog Seznam jako vyhledávací technologii. V určité podobě fungoval Kompas až do roku 2001, kdy Seznam začal používat Google. Později

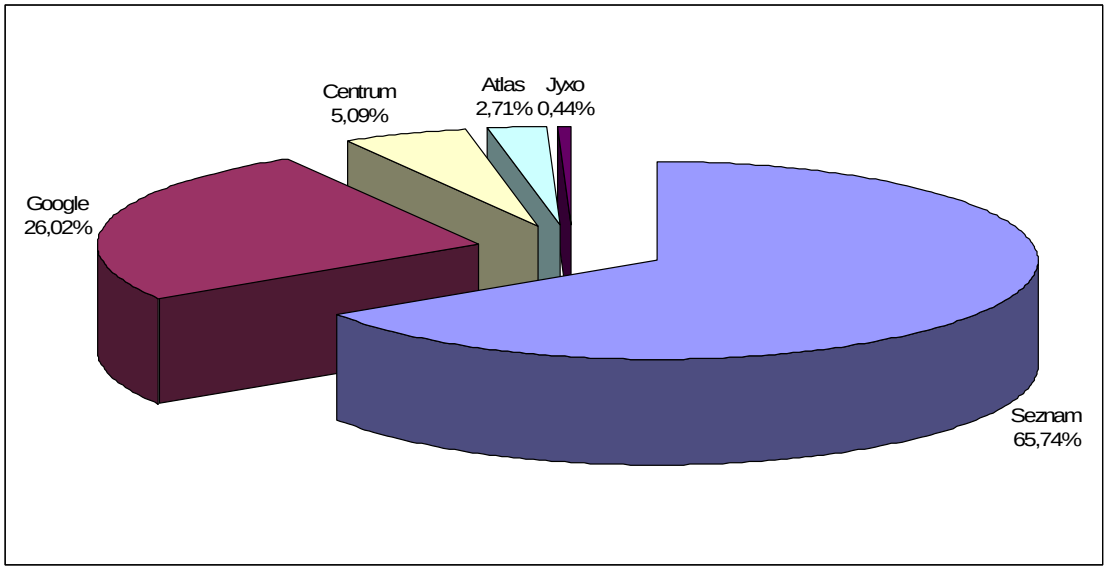
zakoupil technologii Jyxo, kterou na začátku roku 2005 nahradil svou vlastní – Seznam Fulltextem. Seznam, tak již primárně nevyhledává v katalogu, ale ve fulltextu.

Centrum začalo fulltext vyvíjet v roce 1999 společně s Martinem Marešem z MFF UK[38]. V té době se používal pouze na serveru (katalogu) Centrum.cz. V srpnu 2002 spustilo Centrum službu Webfast, který už měl samostatnou doménu. Začátkem roku 2003 Centrum provedlo ve fulltextu zásadní změny (specializace na češtinu – ohýbání slov, synonyma, překlepy), a proto byla změněna značka vyhledávače na Morfeo. V témže roce se z důvodu zkvalitnění výsledků fulltext propojil s katalogem Centra. Na podzim 2003 začalo Morfeo používat pro řazení výsledků také zpětné odkazy (obdoba Google PageRanku[39]).

Další vyhledávač na českém poli je Jyxo, které vzniklo na jaře 2002 (založení firmy). Z velkých katalogů ho od začátku roku 2003 používá Atlas a Quick a od února 2004 do března 2005 pak Seznam. Stejně jako Morfeo zvládá ohýbání slov, synonyma, překlepy a používá JyxoRank (obdoba Google PageRanku).

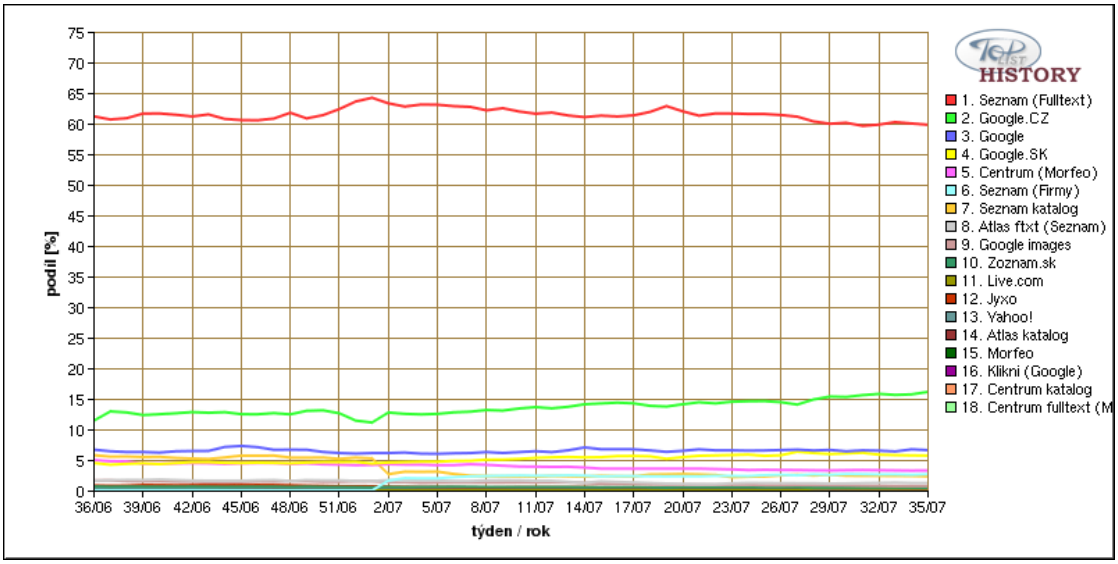
Dle ČSÚ[40] v roce 2006 využívalo služeb vyhledávačů 41,1% populace České republiky starší deseti let, přičemž nejvíce uživatelů (73,8%) připadá na věkové rozmezí 16 – 24 let [14]. Seznam drží téměř dvoutřetinový podíl návštěv webů z vyhledávačů v ČR měřených službou Navrcholu.cz[41]. Od konce srpna 2006, kdy začala fungovat národní verze Google na doméně Google.cz, se jeho celkové zastoupení mezi vyhledávači sice zvýšilo, přesto však Seznam udržuje výrazný náskok. Za těmito dvěma silnými hráči se řadí vyhledávače, jejichž podíl aktuálně nepřesahuje ani pětiprocentní hranici [19].

[38] Matematicko-fyzikální fakulta, Univerzita Karlova
[39] spočívá v tom, že se pro každou indexovanou stránku rekurzivním postupem vypočítá specifická hodnota, závislá na tom, kolik (a jak významných) stránek na tuto stránku odkazuje
[40] Český statistický úřad
[41] www.navrcholu.cz



Graf 2: Podíl jednotlivých vyhledávačů v ČR, leden 2007 [19].

Rovněž z údajů serveru Toplist vyplývá, že největší význam z hlediska odkazujících serverů má v ČR Seznam a Google. Seznam má podobnou pozici jako měl v anglicky mluvících zemích Yahoo!, ze kterého i ideově vychází. I když to není možná na první pohled patrné, jsou vyhledávače přímou konkurencí katalogům. Google poslední dobou stále ukrajuje větší část, a to jak Seznamu tak i Yahoo!.



Graf 3: Podíl vyhledávačů v refereru[42] [40].

V současné době naprostá většina uživatelů hledá požadované informace na Internetu pomocí vyhledávačů a jen malá část pak pomocí procházení katalogových sekcí. To ovšem neznamená, že není důležité odkazy na WWW stránky

[42] URL odkud přišlí návštěvníci WWW stránek

do katalogů registrovat. Stále platí, že základní registrace do katalogů jsou ve většině případů zdarma a registrací do správných sekcí katalogů jsou získávány relevantní zpětné odkazy.

5.2.3 Google

Google přišel v roce 1998 a autoři v něm recyklovali vyhledávací službu BackRub a název „Google“ zvolili jako hříčku vzniklou z anglického slova „googol“, tedy výrazu pro číslo tvořené jedničkou a sto nulami. Už od začátku se Google lišil od všech ostatních populárních vyhledávačů, mezi něž patřili například služby AltaVista, Excite, HotBot a další [10]. Základní myšlenkou vyhledávače BackRub z roku 1996 byla analýza zpětných odkazů, technologie předcházející současný PageRank. V roce 1998 Google odpovídal na 10 000 dotazů denně a o půl roku později již do bylo 500 000. Raketový vzestup a obsáhlejší popis dalších let společnosti Google Inc. je dostupný stránkách společnosti Google, v sekci „Google Milestones“[43]. Vzhledem ke špatnému ekonomickému prostředí (splasknutí bubliny „dotcom“) vstoupil Google s akciemi na burzu až v srpnu roku 2004.

Graf 4: Vývoj akcií společnosti Google (1. září 2004 – 31. srpen 2007) [24].

V současné době společnost Google neprovozuje pouze fulltextový vyhledávač, ale poskytuje celou řadu aplikací. Od dubna 2002 mohou zájemci využívat službu aplikačního rozhraní Google API, díky kterému mají tvůrci dalších webových aplikací možnost legitimně pracovat s vyhledávacím algoritmem Googlu a vše automatizovat podle svých potřeb[44].

[43] http://www.google.com/corporate/history.html
[44] automatické zadávání dotazů jakoukoli jinou cestou odporuje podmínkám, za nichž Google své služby poskytuje

5.2.3.1 Základy práce uživatele s Googlem

V případě, že uživatel napíše do dotazu vyhledávání více než jedno slovo, je vždy stanoven standardní postup zpracování dotazu. Každý vyhledávač je nastaven tak, aby mezi zadaná slova vložil buď logický operátor `AND` (vyhledávaný dokument musí obsahovat všechna zadaná slova) nebo `OR` (postačuje, aby dokument obsahoval jedno ze zadaných slov. Pokud vyhledávač používá standardně například operátor `AND`, existuje i možnost explicitně uvést operátor `OR` a naopak. Vždy však musí existovat implicitní nastavení pro případ, že uživatel žádný operátor neuvede.

Základní logické výrazy

Při zadání více slov používá Google standardně operátor `AND`. To znamená, že pokud uživatel napíše do dotazu několik slov bez operátoru, získá odkazy na dokumenty obsahující všechny zadané termíny. Dotaz může být ve tvaru:

`freon atmosféra „ozonová díra“`[45]

Google vyhledá stránky, na nichž se vyskytují všechna použitá slova. Pokud chce uživatel vyhledávači naznačit, že jsou pro něj přijatelné dokumenty obsahující libovolné z těchto slov, vloží mezi ně operátor `OR`:

`freon OR atmosféra OR „ozonová díra“`

Pokud by chtěl najít dokumenty obsahující jedno slovo a kromě něj ještě některé z dalších dvou slov, může část hledaného výrazu uzavřít do závorek:

`freon ( atmosféra OR „ozonová díra“ )`

Výsledkem takového dotazu budou stránky obsahující termíny *atmosféra* nebo *ozonová díra* a zároveň slovo *freon*. Operátor `OR` lze nahradit znakem `|` převzatý z programovacích jazyků:

`freon ( atmosféra | ozonová díra )`

Chce-li uživatel naopak vyhledávači Google sdělit, že určité slovo nebo fráze se ve výsledných dokumentech vyskytovat nesmí, použije znak `-`:

`freon atmosféra –„ozonová díra“`

[45] pokud uživatel uzavře několik slov do uvozovek, je vyhledávaná fráze, která přesně odpovídá výrazu v uvozovkách

Výsledkem pak budou stránky obsahující slovo *freon* a *atmosféra*, žádná z nich však nebude smět zároveň obsahovat výraz „*ozonová díra*“.

Jednoduché hledání a tlačítko Zkusím štěstí

Tlačítko s názvem Zkusím štěstí[46] nabízí uživatelům Googlu poměrně zajímavou funkci. Místo seznamu nalezených stránek vyhledávač přesměruje uživatele přímo na dokument, který považuje za nejvhodnější, tedy na ten, který by byl při běžném hledání uveden na prvním místě seznamu nalezených stránek (dokumentů).

Velká a malá písmena

Některé vyhledávače berou při zpracovávání dotazu v úvahu velikost písmen. Bude-li dotaz od uživatele ve tvaru „*JAN ŽIŽKA*“, nebudou se vůbec hledat dokumenty obsahující toto jméno v jiné kombinaci velkých a malých písmen, např. „*Jan Žižka*“ či „*jan žižka*“. Google naproti tomu ohled na velikost písmen nebere. Dotaz na slovo *PTÁK*, *pták*, *Pták* či *pTáK* povede vždy ke stejné odpovědi.

Speciální parametry

Mimo základních logických operátorů `AND` a `OR` a možnosti vyhledávat specifickou posloupnost slov uzavřením do uvozovek nabízí Google řadu speciálních parametrů, pomocí kterých lze dotaz upřesnit.

Google dokáže vyhledávat zadané výrazy v celém textu dokumentu. To také znamená, že musí celý dokument indexovat, pouze index nadpisu či anotace není postačující. Existují však zvláštní parametry, které lze použít jakou součást dotazu a omezit pomocí nich hledání jen na určité části webové stránky nebo na určitý typ informací. Zvláštní parametry jsou velmi využitelné ve chvíli, kdy běžný dotaz vrátí velké množství nalezených dokumentů a uživatel potřebuje nalézt jakékoli další pravidlo, kterým by počet nálezů omezil. Ideální je například určit slovo, které se musí vyskytovat v adrese (URL) dokumentu. Tím se podstatně zúží okruh prohledávaných dokumentů, aniž by uživatel musel upravovat samotný dotaz.

Parametr `intitle:`

Parametr `intitle:` zajistí, že bude vyhledáváno jen v titulku webové stránky. Existuje také varianta `allintitle:`, která vyžaduje, aby byla v titulku

[46] v anglické verzi je tlačítko pojmenováno „I’m feeling lucky“

uvedena všechna zapsaná slova. Parametr `allintitle:` by měl být používán jen ve velmi výjimečných případech, protože se poměrně špatně kombinuje s ostatními parametry.

```
intitle:“Jaroslav Vrchlický“
allintitle: „Daň z přidané hodnoty“ podnikání
```

Parametr `inurl:`

Pomocí parametru `inurl:` jsou vyhledávány zadané výrazy jen adrese (URL) dokumentu. Poměrně dobře může být využit při vyhledávání nápověd a stránek uživatelské podpory, protože ty jsou obvykle přehledné a jednotně koncipované. Varianta `allinurl:`, obdobně jako `allintitle:`, zajistí, že se v nalezených adresách budou vyskytovat všechna následující slova. Neměla by se ale kombinovat s dalšími parametry.

```
inurl:nápověda
allinurl:hledat nápověda
```

Parametr `intext:`

Použitím parametru `intext:` lze prohledávat jen samotný text dokumentu, tedy samotné tělo dokumentu. Při vyhledávání se tak ignoruje titulek, adresa a další specifické části stránky. Znovu je k dispozici varianta `allintext:`, ale i v tomto případě platí, že se velmi obtížně kombinuje s dalšími parametry. Parametr `intext:` sice nemá široké využití, ale pro určité typy dotazů se přesto hodí velmi dobře, například při vyhledávání výrazů, které se často vyskytují v adresách, ale málo v textu dokumentů.

```
intext:“email.cz“
intext:html
```

Parametr `inanchor:`

Parametr `inanchor:` lze použít tehdy, pokud chce uživatel vyhledávat v textech hypertextových odkazů. V případě, že v HTML stránce je zápis: `<a href=”http://fyzika.net”>Vstupte do světa fyziky</a>`, textem takového odkazu je „Vstupte do světa fyziky“. V něm by se také vyhledávala zadaná slova.

```
inanchor:“laserový paprsek“
```

Parametr `site:`

Parametr `site:` lze použít v případě, pokud chce uživatel omezit vyhledávání na jeden server nebo doménu. Vyhledávač AltaVista používá ke stejnému účelu dva parametry, `host:` a `domain:`, Google však jen jeden.

```
site:centrum.cz
site:www.pef.czu.cz
site:cz
```

Parametr `link:`

Parametr `link:` vrací seznam stránek, které odkazují na zadanou adresu. Pokud uživatel zadá dotaz například ve tvaru `link:www.google.com`, získá seznam stránek, které odkazují na vyhledávač Google. Definovat protokol `http://` se v tomto případě do dotazu nemusí a pokud je uživatelem uveden, Google jej ignoruje. Parametr `link:` funguje stejně dobře při hledání odkazů na dlouhé adresy (např. `http://www.mujweb.cz/www/xjmeno/stranka.html`) i na adresy obsahující pouze názvy domén (např. `mujweb.cz`).

Parametr `cache:`

Parametr `cache:` umožňuje uživateli najít kopii stránky, která se už nevyskytuje na původní adrese, nebo zcela změnila svůj obsah. Pokud Google vrátí na určitý dotaz adresy dokumentů, které očividně nesouvisí s právě zadaným dotazem, je vhodné vyhledat jejich starší verze v záznamech Googlu právě pomocí parametru `cache:`. Parametr `cache:` lze tedy využít pro vyhledávání stránek, které se velmi často mění.

```
cache:www.quick.cz
```

Parametr `daterange:`

Použitím parametru `daterange:` lze omezit hledání na stránky, které Google indexoval v zadaném časovém období, tedy kdy ji Google zařadil do svého indexu. Stránka vytvořená 5. května a indexovaná 15. května se objeví ve výsledcích dotazu jedině tehdy, bude-li uživatel parametrem `daterange:` vyhledávat stránky indexované 15. května. Google samozřejmě indexuje všechny stránky opakovaně, ale to, jestli se jeho záznam o poslední aktualizaci indexu změní, záleží na tom, jestli se změní obsah stránky samotné. Důležitým faktem také je,

že parametr `daterange:` akceptuje juliánské datum, které se od gregoriánského data liší. Na Internetu existuje aplikace, které zajistí převod data mezi oběma formáty. Pokud se uživatel nechce konverzí data zdržovat, může použít grafické rozhraní FaganFinger[47], v němž lze datum pro parametr `daterange:` nastavit pomocí roletového menu ve známější gregoriánské soustavě.

`„Jaromír Jágr“ daterange:2453121-2453135`
`otorinolaringologie daterange:2453137`

Parametr `filetype:`

Parametr `filetype:` bere v potaz při vyhledávání dokumentů přípony souborů. Různé přípony obvykle (nikoli však nutně) odpovídají různým typům souborů. Tento fakt je nutné si uvědomit, neboť dotaz s parametrem `filetype:htm` vrátí jiné výsledky než dotaz s parametrem `filetype:html`, i když jde v obou případech o dokumenty ve stejném formátu. Dotaz je možné omezit dokonce i na stránky využívající různé skriptovací jazyky (ASP, ASP.NET, PHP či CGI), ovšem pouze v případě, že je správce stránek nemaskuje přesměrováním apod. Google dokáže indexovat a tím pádem i prohledávat dokumenty v některých formátech definovaných firmou Microsoft. Mezi ně patří např. MS PowerPoint (`.ppt`, `.pps`), MS Excel (`.xls`) či MS Word (`.doc`).

`„požární ochrana“ filetype:doc`
`„ekonomický rozvoj“ filetype:ppt`

Parametr `related:`

Pomocí parametru `related:` lze vyhledávat stránky, které souvisejí se zadaným dokumentem. Tímto způsobem je možné nacházet stránky patřící do určité kategorie, například při hledání stránek souvisejících s Googlem (`related:google.com`) najde uživatel řadu dalších vyhledávačů, mezi nimi HotBot, Yahoo! či Northern Light.

`related:www.seznam.cz`
`related:www.idnes.cz`

[47] http://www.faganfinger.com/engines/google.shtml

Parametr `info:`

Dotaz s parametrem `info:` vrací seznam odkazů na dokumenty informující o zadané adrese (URL). Mimo jiné tu lze nalézt odkaz na obraz zadaného dokumentu v archivu Google, seznam stránek obsahujících odkazy na zadanou adresu, seznam souvisejících stránek a stránek, které uvedenou adresu obsahují ve svém textu. Spolehlivost uvedených informací samozřejmě závisí na tom, dokázal-li již Google zařadit uvedenou stránku do svého indexu. U nových a ještě neindexovaných stránek bude poskytnutých informací méně.

```
info:pocasi.centrum.cz
info:www.idnes.cz
```

Parametr `phonebook:`

Parametr `phonebook:` slouží k vyhledávání telefonních čísel, ale bohužel pouze v telefonních seznamech Spojených států.

```
phonebook:John Doe CA
phonebook:(510) 555-1212
```

Google průběžně přidává nové parametry a tím zpřístupňuje nové funkce. Uživatel může také využít formulář pro pokročilé vyhledávání[48], jímž nahradí jeden komplexní dotaz.

Filtrování

V anglické verzi uživatelova prostředí[49] nabízí Google filtrovací program SafeSearch, který se snaží odstranit ze seznamu nalezených dokumentů ty položky, které by mohly urazit citlivější uživatele. Implicitně je filtrování nastaveno na střední úroveň (moderate), které vylučuje dokumenty s nevhodnou grafikou. Plný filtr (strict) dokáže odstranit i stránky s nevhodným textem.

Volba slov – slang a terminologie

Řekne-li dospívající člověk o něčem, že je to „hustý“, používá slang, tedy zvláštní druh výrazů charakteristických pro určitou skupinu lidi nebo určitou kulturu. Napíše-li geolog do své práce výraz „stratotyp“, není to sice slang, ale přesto jde o zvláštní výraz, který používá jen určitá snadno definovatelná komunita.

[48] http://www.google.cz/advanced_search?hl=cs
[49] http://www.google.cz/preferences?hl=en

Při vyhledávání informací na Internetu mohou zvláštnosti uživatelova vyjadřování velmi dobře posloužit. Stačí, když přidá do dotazu méně obvyklý termín a může si být téměř jist, že nalezené dokumenty budou mnohem lépe odpovídat jeho představám.

Díky slangu může uživatel snadno upravit většinu dotazů tak, aby vracely především stránky z určité geografické oblasti. Hranice používání slangu sice nejsou úplně přesné, ale tento přístup funguje velmi dobře, především v angličtině, kterou se hovoří na velkém území. Hledá-li uživatel v anglicky psaných dokumentech, kde je občas třeba odlišit různý původ článků, může pro určení slangu využít určité volně dostupné internetové zdroje.

Probert Encyclopedia – Slang[50]

Stránky Probert Encyclopedia se dají procházet i prohledávat. Vyhledávač ovšem prochází celou encyklopedii a slangové výrazy se objevují až u konce. Sebraná slova pocházejí z celého světa a jsou doplněna křížovými odkazy (především u termínu s drogovou tematikou).

Dictionary of Slang[51]

Dictionary of Slang se zaměřuje především na slang používaný ve Velké Británii a lze jej volně procházet i prohledávat. Slova, která nepocházejí z Velké Británie, jsou doplněna místem původu uvedeným v hranatých závorkách. U slov je uvedeno i jejich vyznění (humorné, vulgární, hanlivé apod.)

Surfing for Slang[52]

Seznam Surfing for Slang se věnuje především dokumentům, které se zabývají slangem v angličtině a skandinávských jazycích.

Odborné výrazy se používají v různých oblastech lidské činnosti. V některých oborech je nalezení vhodného odborného výrazu poměrně jednoduché, jinde však dosti obtížné. Pokud se jedná o anglicko terminologii, může uživatel využít stránky Glossarist[53], které obsahují index zhruba 6000 různých slovníčků věnovaných desítkám různých témat.

[50] http://www.probertencyclopedia.com/slang.htm
[51] http.//www.peewish.co.uk/slang/
[52] http://www.linkopp.com/members/vlaiko/slanglinks.htm
[53] http://www.glossarist.com

On-Line Medical Dictionary[54]

Slovník On-Line Medical Dictionary obsahuje výrazy používané v biochemii, biologii rostlin, buněčné biologii, fyzice, chemii, lékařství, mechanice, molekulární biologii, radiobiologii a technologii. V současné době obsahuje více než 46000 záznamů. Slovník lze procházet dle abecedy nebo v něm vyhledávat. Uživatel také ve slovníku může vyhledat termín, který zná, a nalézt k němu ekvivalent používaný častěji v lékařství. Slovník vychází z britského prostředí, pokud je uživatel zvyklý na americkou angličtinu, musí mít na zřeteli určité rozdíly.

MedTerms.com[55]

Server MedTerms.com má sice ve slovníku méně slov (kolem 10000), ale zato disponuje rozsáhlým archivem článků z projektu MedicineNet. Nachází-li se uživatel na samém počátku hledání, může zde nalézt články, které ho seznámí se základní terminologií.

Law.com – Legal Dictionary[56]

Právní slovník Law.com může být pro uživatele velmi přínosný, protože v něm může vyhledávat nejen termíny, ale také jejich definice. Může tedy zadat jedno slovo a získat seznam všech definic, které dané slovo obsahují.

Omezení na 10 slov

Uživatelé, kteří potřebují pro svou práci používat složité vyhledávací dotazy, pravděpodobně narazili na vstupní omezení deseti slov. Google skutečně omezuje maximální délku dotazu na deset slov a počítá v to i všechny operátory a parametry. Zbytek dotazu jednoduše ignoruje Běžné uživatele takové omezení patrně nijak nelimituje, ale ti náročnější mohou mít problémy.

Pokud však uživatel vynechá v dotazu běžná a často používaná slova, ušetří dostatek prostoru pro svůj vyhledávaný řetězec, popř. může využít tzv. hvězdičkovou konvenci. Google totiž nepočítá hvězdičky jako slova, a tak je lze pomocí nich ušetřit místo pro vyhledávaný dotaz. Pokud má tedy vyhledávaný dotaz více než deset slov, je možné některé jeho méně důležité části nahradit hvězdičkami a ušetřený prostor

[54] http://cancerweb.ncl.ac.uk/omd/
[55] http://www.medterms.com
[56] http://dictionary.law.com/lookup2.asp

využít k zapsání upřesňujících výrazů. Hvězdičkou se však nedají nahradit části slov, ale jen slova celá[57].

Pořadí slov a opakování slov

Pořadí, ve kterém jsou zapsána slova do dotazu je stejně důležité, jako slova samotná. Změna pořadí slov může ovlivnit nejen počet výsledků, ale také pořadí jednotlivých odkazů v seznamu výsledků. U frází uzavřených v uvozovkách to je jednoznačné, ale pořadí slov má skutečně vliv i na postup hledání slov uvedených samostatně. Google na svých stránkách potvrzuje, že pořadí slov má svůj význam, ale nenabízí žádný návod, podle kterého by se uživatel mohl rozhodnout, jaké pořadí slov zvolit. Google má sklony přikládat pořadí slov určitou váhu a dokumenty se slovy zapsanými v uvedeném pořadí upřednostňovat. Mimo to přichází v úvahu i fakt, zda jsou požadovaná slova uvedena vedle sebe a nakonec i na význam jednotlivých slov jako takových.

Pokud uživatel použije jedno klíčové slovo několikrát. může tím rovněž do značné míry ovlivnit pořadí a počet odpovědí, které dostane. Google se o opakování vyhledávaných slov na svých stránkách nezmiňuje, nicméně pravděpodobně hledá co nejvíce výskytů uvedených slov nebo frázi. Skončí teprve ve chvíli, kdy už nemůže nalézt další. Ovšem, pokud Google nenajde hledané slovo v dokumentu tolikrát, kolikrát ho uživatel zadal do dotazu, nic se neděje a dokument je považován za vhodný. Uživatel tedy může opakování slov snadno určovat jejich váhu.

Z vyhledávacích experimentů vyplynulo:

- Při opakování podstatných jmen či názvů produktů se na přední místa dostávají především stránky prodejců. Tento efekt je ještě silnější tehdy, je-li uvedeno podstatné jméno v množném čísle.
- Pokud se výsledky vyhledávání neliší při dvojnásobném a trojnásobném opakování zadaného slova, neznamená to, že se výsledky nebudou lišit, bude-li vyhledávané slovo opakováno čtyřikrát či pětkrát.

[57] Běžně používaná slova jako „to“, „i“, „a“, „and“, „how“, „1“, a další nepřinášejí při hledání žádný užitek, neboť se jedná o tzv. „stop slova“ a Google je ve skutečnosti zcela ignoruje. Chce-li uživatel vyhledávací algoritmus přinutit, aby stop slovo při hledání použil, musí před něj přidat znak +.

Vlastní vyhledávací formuláře

Požaduje-li uživatel zadat jednoduchý vyhledávací dotaz, plně mu postačuje základní rozhraní Googlu, tedy domovská stránka. Ovšem, pokud ví, že on nebo jeho kolegové budou často klást podobné a relativně složité dotazy, může si připravit svůj osobní vyhledávací formulář.

```
<!-- Search Google -->
<form method="get" action="http://www.google.com/search">
<input type="text" name="q" size="31" maxlenght="255" value="" />
<input type="submit" name="sa" value="Vyhledat Googlem" />
</form>
<!-- Search Google -->
```

Příklad 2: Základní vyhledávací formulář Googlu.

Takový formulář je velmi jednoduchý a nedělá nic jiného, než že vezme zadaný dotaz a odešle jej na Google ke zpracování. Nic jej neupravuje. Ovšem není složité do formuláře doplnit další prvky (ať skryté či viditelné) a tak rozšířit nebo specializovat jeho funkci.

Pokud je nutné změnit některý z parametrů hledání, lze do vyhledávacího formuláře doplnit skrytou proměnou[58].

Typ souboru

Jednou z možností je filtrovat výsledky hledání a vybírat z nich pouze ty, které odpovídají zadanému typu souboru, například pouze soubory `.doc`, `.xls`, `.ppt`, `.pdf` nebo `.txt`. Pokud by tvůrce formuláře chtěl do dotazu doplnit parametr, který zajistí, že budou vyhledávány jen prezentace v MS PowerPoint, může zdrojový kód doplnit o skrytou proměnnou a názvem `as_filetype` a hodnotou `PPT`.

```
<input type="hidden" name="as_filetype" value="PPT" />
```

Vyhledávání v dané doméně

Vyhledávání v doméně omezuje hledání na konkrétní doménu nebo konkrétní server. K poměrně obecnému hledání je možné použít doménu nejvyšší úrovně, například `.cz`, ale prohledávaný rozsah se dá omezit i radikálněji. Vyhledávání

[58] Skrytá proměnná znamená, že uživatel ji nebude schopen měnit a bez nahlédnutí do zdrojového kódu HTML nebude znát ani její hodnotu.

v dané doméně může mít pro tvůrce WWW stránek význam, pokud chtějí vytvořit rychlé fulltextové hledání nad svým serverem.

```
<input type="hidden" name="as_sitesearch"
value="idnes.cz" />
```

Vyhledávání podle stáři

Snadno je možné vyhledávat stránky aktualizované jen v průběhu několika posledních měsíců, přičemž použitelné hodnoty jsou mezi 1 a 12.

```
<input type="hidden" name="as_qdr" value="m7" />
```

Počet výsledků

Ve formuláři lze také určit počet výsledků, které se zobrazí společně na jedné stránce. Google akceptuje hodnoty v rozsahu 1 až 100 [10].

```
<input type="hidden" name="num" value="50" />
```

5.2.3.2 PageRank a Hilltop

PageRank je algoritmus, který Google používá k hodnocení dokumentů. Algoritmus tak staví na předpokladu, který se běžně uznává v akademické sféře, tedy publikace je tím významnější, čím více dalších publikací na ni odkazuje. Ze statistiky publikací vzniká tzv. citační index, který Sergey Brin s Larrym Pagem[59] převedli do webové podoby. Význam dokumentu se podle nich určuje především s ohledem na počet odkazů, které vedou k hodnocení dokumentu z jiných stránek [10].

Dle Googlu samotného je PageRank objektivní měření důležitosti webových stránek, které je realizováno pomocí hlasů. PageRank interpretuje odkaz ze stránky A na stránku B jako hlas pro stránku B od stránky A. Algoritmus pak sečte počet hlasů a podle toho určí její důležitost na stupnici od 1 do 10. Čím vyšší hodnotu daná stránka má, tím je důležitější a tím více stránek na ní odkazuje. Mimoto, Google bere v úvahu důležitost hlasujících stránek, jelikož některé stránky mají větší hlas než jiné. Google používá „kolektivní inteligenci Webu“, tedy vypočítává na základě miliónu hodnot a miliard klíčových slov.

PageRank je veličina skalární, tedy pro každou stránku (URL, dokument) jde o jediné číslo. V rámci některých pojetí je PageRank označován jako vektor v n-rozměrném prostoru, kde n je počet dokumentů. Přesná metoda výpočtu PageRanku není známa – je obchodním tajemstvím společnosti Google. Je mnoho

[59] autoři původní myšlenky výpočtu PageRanku

názorů jak se PageRank počítá, sám Google tvrdí, že při výpočtu hlasu, který stránka předává pracuje i s analýzou HTML a CSS, například tučným odkazům dává vyšší váhu apod.

Pagerank v klasickém pojetí dosahuje hodnot od nuly do jedné. Vzorec výpočtu hodnoty PageRanku dle původní dokumentace [4] Sergeye Brina a Larryho Page je:

$$PR(A) = (1-d) + d * (PR(T_1)/C(T_1) + ... + PR(T_n)/C(T_n))$$

PR(A) – hodnocení (PageRank) stránky A

$PR(T_1)$ – hodnocení stránky T_1

C(T1) – počet odkazů vedoucích ze stránky T1

d – tlumící faktor, pro nějž platí $0 < d < 1$ a jeho hodnota bývá 0,85

Hodnocení stránky se tedy vypočítá jako součet hodnocení všech stránek, které na ní ukazují, vydělených celkovým počtem odkazů, které z těchto stránek vedou. Dle autora serveru „Jak psát web“ [20] je však výše uvedený vzorec výpočtu chybný a sám uvádí vzorec doplněný o proměnnou *m*:

$$PR(A) = (1-d)/m + d * (PR(T1)/C(T1) + ... + PR(Tn)/C(Tn))$$

m – celkový počet zaindexovaných stránek

Tím, že stránka obsahuje odkazy, o svůj PR nepřichází. Spíše než o „předávání“ by se dalo hovořit o „kopírování“ nebo "přeposílání" PageRanku. Čím méně má stránka odkazů, tím více Page Ranku se každým odkazem přeposílá. PageRank nijak nezávisí na hledaném slově (je to veličina skalární) a každá jednotlivá stránka (URL) má svůj vlastní Page Rank, nikoliv celý web dohromady (site, doména).

Čím má stránka vyšší PageRank, tím bude pravděpodobně výše ve výsledcích vyhledávání, ovšem Page Rank zdaleka není jediné kritérium pro nalezení stránky v Google. Page Rank je jistě velmi důležitý, ale jeho význam bývá v určitých případech přeceňován. Google však zřejmě používá PageRank zejména pro řízení robota (crawlera) a stránky s nízkým rankem navštěvuje zřídka. Některá nová URL ani nenavštíví, dokud jejich PR nedosáhne určité hodnoty (PR se dá počítat i pro ještě nenavštívené stránky).

Nejjednodušší metodou jak zjistit aktuální velikost Google PageRank pro danou stránku je použití jednoduchých rozšíření pro nejpoužívanější prohlížeče, tzv. toolbary[60]. Na nainstalovaném toolbaru se objevuje zelené měřítko Page Ranku od 0 do 10 a jehož hodnota je označována za Google Toolbar PageRank. PageRank zobrazený na Toolbaru není skutečný PageRank, ale jeho odvozenina (patrně logaritmická) [20].

Zatímco PageRank určuje důležitost webové stránky obecně, Hilltop (LocalScore) určuje důležitost webové stránky v relaci s dotazy na vyhledávače [17] a autory jsou Krishna Bharat a George Mihaila. Krishna Bharat stanovil metodu, při níž je pro výpočet PR nutné brát v potaz „směrodatnost“ dokumentu. Hodnota důležitosti webové stránky je tak určena s ohledem na dokumenty, které se týkají tématu, tzv. „expert documents”, tedy stránky, vytvořené za účelem navádění uživatelů ke zdrojům informací[61]. Odkazy z expertních dokumentů (webových stránek) pak určují „směrodatné skóre“[62] [2]. Jako příklad těchto expert dokumentů lze uvést zařazení v populárních katalozích – DMOZ, Yahoo, apod. Zjednodušený výpočet LocalScore dle algoritmu Hilltop je:

- Hilltop provede běžné vyhledávání aby vyhledal korpus63 expert dokumentů. Kvalifikační předpoklady pro tyto dokumenty jsou velmi přísné, tedy korpus je technicky zvládnutelné množství webových stránek
- Hilltop z výsledného korpusu odfiltruje „spřízněné website“64 a stránky s duplicitním obsahem. Veškeré „spřízněné website“ definuje Hilltop velmi exaktně. V zásadě se jedná o stránky ze stejné domény 2. úrovně, ze stejné C-třídy IP adres65 a spřízněné weby spřízněných webů (spřízněnost druhé úrovně).
- Stránka získá LocalScore což je hodnota „směrodatnosti“ založená na počtu a kvalitě hlasů, které získá od „expertních dokumentů“. Stránky jsou pak řazeny podle výsledků samotného LocalScore.

[60] http://toolbar.google.com
[61] ...pages that have been created with the specific purpose of directing people towards resources
[62] Authority Score, AS, LocalScore
[63] soubor, corpus
[64] affiliate sites
[65] Kolik bajtů z IP-adresy tvoří adresu sítě určují počáteční bity prvního bajtu IP-adresy. IP-adresy se dělí do pěti tříd přičemž u třídy C nevyšší tři bity prvního bajtu mají hodnotu 110_2.

Hilltop ovšem zakládá své výpočty pouze na „expert documents“. Jeho algoritmus potřebuje alespoň dva hlasy těchto dokumentů pro danou webovou stránku aby vrátil výsledky. V důsledku to znamená, že pokud dané dokumenty nenalezne vrací „nulový“ výsledek. Většinu SEO[66] komunity to vedlo k domněnce, že Google používá seznam klíčových slov pro který v SEO světě zevšedněl název „Money Words List“. Tento seznam rozděluje dotazy na ty, pro které bude použit algoritmus Hilltop a na ty pro které bude využit klasická metoda PageRank. Domněnka byla víceméně potvrzena skupinou „Google Hate Group“, která nalezla chybu v algoritmu Googlu. Chyba umožňovala díky přidání příkazu pro odstranění neexistujícího slova (např. „-dakjsjdksadk“) do dotazu porovnat výsledky podle starého a nového algoritmu. Google tedy nenašel relevantní expertní dokumenty a proto zobrazoval staré výsledky. GHG dokonce vytvořila webový nástroj, který celou věc automatizoval a to na adrese `www.scroogle.org`. Zavedením Hilltopu do metod výpočtů Google proběhlo 15. listopadu 2003.

„Money Word List“ ovšem podle posledních průzkumů není pevně stanoven. Je stanoven na počtu instancí daných klíčových slov ve vyhledávacích dotazech. Google má pravděpodobně nastavenou určitou hranici a pokud dané klíčové slovo dosáhne limitní frekvence, pak teprve přichází tento filtr do hry. Celý algoritmus pro výpočet GoogleRanku je následující:

Starý GoogleRank = [(1- d) + a(RS)] [(1- e) + b(PR* z)]*

Nový GoogleRank = [(1- d) + a(RS)] [(1- e) + b(PR* z)]* [(1- f) + c(LS)]*

RS – Relevance Score – Skóre relevantnosti odkazu, které je založené na klíčových slovech, které jsou součástí dokumentu. Různá váha je přiřazována klíčovým slovům uvedených v hlavičce, nadpisech, tučných výrazech a běžném textu WWW stránky.

PR – PageRank-

LS – LocalScore – Skóre vypočítané na základě expert dokumentů. Klíčové slovo získává různé hodnoty, pokud se zobrazuje v titulku, nadpisu, odkazu. Dále se bere v potaz hustota klíčových slov apod.

a, b, c – hodnoty pro ladění váhy jednotlivých částí výrazu

[66] Search Engine Optimization – optimalizace stránek pro vyhledávací roboty

d, e, f – hodnoty pro ladění útlumu jednotlivých částí výrazu. Předpokládá se, že v současnosti je f=0.

z – faktor základu – PR uvedený v rozsahu 1 až 10 nemá ve skutečnosti lineární průběh, ale exponenciální. Tato hodnota pouze převádí hodnotu PR v tomto formátu na skutečnou hodnotu PR.

Otázkou zůstává, jaké je rozdělení váhy mezi jednotlivé proměnné. Literatura většinou uvádí následující rozložení vah:

- RelevanceScore = 20%
- PageRank = 40%
- LocalScore = 40% [17]

6 Search Engine Optimization

Nekonvenčnímu marketingu, který má za cíl s minimem nákladů dosáhnout maximálních výnosů se někdy říká guerillový marketing (autorem tohoto pojmu je Jay Conrad Levinson). Zatímco tradiční způsob marketingu je především o penězích, guerillový je hlavně o důvtipu. I oblast SEO byla ve svém počátku guerillovým marketingem. Situace se však, jak už to v oblasti ICT bývá, kvapně mění a proto už dnes lze považovat metody SEO a SEM[67] za běžné marketingové metody vhodné pro nasazení v prostředí sítě Internet.

O tom co přesně se pojmy SEO a SEM rozumí se vedou v internetových diskuzích nekonečné spory. Někdo tvrdí, že SEM je součástí SEO někdo zase, že SEO je jen zlomkem SEM. V každém případě Search Engine Optimization je věda a umění, která má následující cíle:

- posunout webovou stránku na přední místa v internetových vyhledávačích na ta slova, přes která ji hledají potenciální zákazníci,
- přilákat tyto potenciální zákazníky na tyto stránky a motivovat je k nákupu.

V určitých případech se o SEO hovoří jako o organické optimalizaci a SEO samotné je hraniční (multioborová) disciplína, která pokrývá zhruba následující oblasti:

- webdesign
- kódování a programování
- marketing
- copywriting
- statistika a analytika

6.1 Marketing

Odkaz na stránku, který se objeví ve vyhledávači po zadání určitého vyhledávacího dotazu, je v podstatě reklama. Reklama, jejíž hlavní cíl je, řečeno slovy Sergio Zymana, *"prodávat mnohem častěji více zboží většímu množství zákazníků a utržit za to více peněz"*.

[67] Search Engine Marketing

Nejviditelnějším cílem SEO je „zviditelnit“ určitý website ve vyhledávačích, tedy pokud uživatel hledá ve vyhledávači zboží či službu, co prodávající „prodává“, aby ho také uviděl/našel. To však není hlavní marketingový cíl SEO. Pro onu většinu z uživatelů, kteří pracují v tržním hospodářství, je hlavní cíl SEO přilákat na klientovy stránky potenciální zákazníky a prodat jim produkty. Jediné faktické hodnocení úspěšnosti SEO je zvýšení (nebo alespoň udržení, pokud konkurence tvrdě útočí) prodeje.

To je tedy hlavní cíl SEO. Ne vždy je však dobře měřitelný. Pokud například je prodávána sportovní obuv, pak se může stát, že uživatel objeví značku přes vyhledávače, stránky jej přesvědčí, ale půjde si je koupit do kamenného obchodu. Pak zbývají jenom dvě věci:

- podrobně zkoumat návštěvnost,
- propojit stránky s offline prostředím (speciální slevové kupóny, zpětná registrace, zvláštní prodejní telefonní čísla apod.).

Může se však stát, že toto není hlavní cíl SEO. Rozhodně je však důležité se zamyslet před startem SEO projektu, co je pro provozovatele WWW projektu (investora) nejdůležitější a jak důležité jsou ostatní body. Je velmi vhodné, pokud lze zadané kvantifikovat.

Pomocí SEO lze tedy dosáhnout:

- zvýšení příjmů,
- zvýšení počtu návštěvníků stránek,
- zvýšení povědomosti o značce,
- kvalitní public relations.

Jestliže je jako hlavní cíl zvýšení příjmů, existují dvě možné strategie. V první z nich je brán jakýkoli příjem, je tedy cíleno široce. Pokud se jedná o prodej luxusní běžecké obuvi, optimalizuje se kromě na „luxusní běžecké boty“ i na „boty“, protože také přes toto klíčové slovo[68] mohou přijít zákazníci a pro jistotu i na „jogging“. Tímto způsobem mohou sice značně stoupnout příjmy, ovšem výdaje budou také značně vysoké.

[68] Key Words

Druhá strategie se zaměřuje výhradně na určitou cílovou skupinu. Website se optimalizuje pouze na „luxusní běžecké boty“. Tím je získán perfektní (internetový) marketing, kterým lze pro koupi obuvi přesvědčit takřka kohokoli, ale mnohem efektivnější je se zaměřit na cílovou skupinu, kterou není třeba moc přesvědčovat.

Pokud je žádoucí zvýšit návštěvnost stránek, opět je nutné se rozhodnout, zda jsou zajímavější jakýkoliv návštěvníci, nebo jenom určitá skupina. Pokud bude postačovat kdokoli, stačí skoupit nejbližší novinový stánek, udělat klastrovou analýzu a ihned je možné dojít k závěru, co lidi, uživatele, potenciální zákazníky, nejvíce zajímá.

Cílená návštěvnost je naproti tomu něco jiného. Většina produktů se přes Internet neprodává ovšem zákazníci tyto produkty na Internetu vyhledávají. Provozovatel website chce, aby ho tito zákazníci našli. Opět může zvolit, zda se bude zaměřovat na co nejširší cílenou návštěvnost, nebo na úzce specifikované zvyšování návštěvnosti.

Ač se to nezdá, SEO lze použít i pro budování značky [25]. A to hned dvojím způsobem. Jednak samotnou přítomností ve výsledcích vyhledávání. I když si na odkaz zákazník neklikne, přesto značku více či méně vnímá [11]. Účinnější způsob je, pokud provozovatel website dokáže přitáhnout návštěvníky přes vyhledávače na své stránky, kde na ně může působit mnohem silněji. Opět platí, že může cílit buďto široce, nebo cíleně.

Značku lze také posílit nebo poškodit tím, co je „kolem website“ napsáno ve výsledcích vyhledávání. A to jak v titulku, popisku a URL, tak mezi ostatními výsledky vyhledávání [25]. Například, jistý lokální poskytovatel telekomunikačních služeb se objevoval na předních místech ve vyhledávačích na celoamerické úrovni. Lidé si pak mysleli, že operuje po celém území USA [11].

Public relations přes vyhledávače souvisí s brandingem přes vyhledávače – jde tedy především o dlouhodobou záležitost. Provozovatel website by ale měl mít i připravené pozitivní a negativní plány. Až tedy budou lidé hledat něco pozitivního (obecně), aby ho nalezli v přítomnosti toho pozitivního. A naopak, v případě negativní situace, aby jeho problémy jen tak někdo nenašel – nebo aspoň byly něčím vykompenzovány.

Pokud provozovatel website chce mít úspěšnou SEO strategii, musí co nejlépe poznat své zákazníky. Každý typ zákazníků má jiné „spouštěče“, na které reaguje. Když zákazník zadá klíčové slovo, provozovatel website chce, aby kliknul právě na jeho odkaz – a pokud možno na žádný jiný. Zákazník má přitom například v českých výsledcích Googlu až 20 (deset přirozených a deset placených) výsledků, na které může kliknout.

Základní otázky, které je vhodné, aby provozovatel website probral, než stanoví svoji SEO strategii, jsou:

- Na jaký typ zákazníků se chce zaměřit?
- Na jaké země (ČR, USA, UK, Německo, ...)?
- V jakých jazycích budou stránky?
- Jací zákazníci jsou nejziskovější?
- Zaměřuje se na koncové nebo firemní zákazníky?
- Jaké jsou konkurenční výhody, proč by se měli zákazníci zakoupit zrovna konkrétní produkty?
- Co hledají zrovna provozovatelovi zákazníci? [25]

6.2 Analýza klíčových slov

Analýza klíčových slov je jednou z nejdůležitějších součástí Search Engine Optimization. Teoreticky lze optimalizovat website na jakékoli slovo. Ovšem po čase každý provozovatel website zjistí, že to stojí spoustu peněz a času, a tudíž se mu to nevyplatí. Proto je zde analýza klíčových slov, která má vypovědět, na co se vyplatí optimalizovat a co je nutné udělat.

Analýza klíčových slov:

- zjistí, jaká slova zákazníci přesně vyhledávají
- zjistí, která slova jsou nejvíce zisková, s největší návratností investic (ROI),
- zjistí veškeré náklady (čas, finance, energie, apod.).

Analýza klíčových slov má dvě fáze: v první fázi je nutné získat co nejvíce klíčových slov, na které lze website optimalizovat. Cílem je, aby nebylo opomenuto

žádné slovo, které by se mohlo stát velmi důležitým. Na hodně konkurenčních trzích ve vyhledávačích totiž dnes spíše provozovatelé website hledají tržní niku, na které by se mohli uchytit. Protože optimalizovat na „velké fráze“ typu „*New York hotels*“ či „used cars“ je velice těžké a pro nováčka na trhu by se to zřejmě ani nevyplatilo.

Analýza klíčových slov se skládá ze dvou částí: vertikální analýzy a laterální analýzy. Vertikální analýza dodává různé tvary daného slova (např. ze slova *spor*t udělá *sportovní*, „*sportovní oddíl*“, *sportu*, „*historie sportu*“). Laterální analýza naopak vyhledá taková slova, která vyhledávají lidé, kteří hledají něco souvisejícího se slovem *sport* (např. se objeví *tenis*, *golf*, ale i *plavky* nebo „*jak na gambit*“).

V druhé fázi naopak z těchto často stovek vybraných slovních spojení jsou hledána právě ta, na která se vyplatí optimalizovat. Jak již bylo napsáno výše, je třeba si nejdříve uvědomit, kdo jsou zákazníci. Pokud na tuto otázku je známa alespoň částečná odpověď a provozovatel website zná své produkty, je snadné odvodit, co zákazníci hledají. Tyto výrazy představují základ klíčových slov.

Zde ovšem existuje nebezpečí v podobě zvolení slov, která jsou používána v žargonu, a ne těch, která vyhledávají zákazníci. Provozovatel website se však může zeptat svých zákazníků, obchodních partnerů, jaká slova by při vyhledávání na Internetu použili. Obecně se vyplatí využít „třetí“ strany.

Zajímavé informace může přinést i analýza log souboru. Ovšem je nutné si uvědomit, že toto jsou slova, na která zákazníci již přicházejí. Tedy ne nutně ta nejvyhledávanější, nejziskovější, s nejlepší návratností investic. Na druhou stranu však provozovatelé website zjistí, co lidé na jejich stránkách již vyhledávají. Mnohem přínosnější je ale analýza pomocí příslušných databází. Databáze mají tři významné vlastnosti:

- pomohou rozšířit a upřesnit klíčová slova,
- sdělí, jak jsou která slova vyhledávána,
- najdou tržní niky.

Nejrozšířenější SEO databází je zřejmě Wordtracker[69] (jedná se o placenou službu). Umožňuje mnoho zajímavých služeb, např. vyhledat příbuzná slova (vertikální i laterální), počet vyhledávání v některých vyhledávačích (včetně Googlu

[69] http://our.affiliatetracking.net/wordtracker/a/11251

a Yahoo! v angličtině), odhadne počet navštívení stránek, pokud budou optimalizována na dané slovo, a spočítá i konkurenci. Právě kombinace posledních dvou faktorů je zajímavá. Vychází z ní totiž tzv. Keyword Efectiveness Index (KEI), který vymyslel Sumantra Roy. Je to podíl mezi počtem vyhledávání a počtem stránek, které obsahují toto slovní spojení v daném vyhledávači.

KEI umožňuje hledat tržní niky. Čím je vyšší (lidé dané slovo hodně vyhledávají a přitom se vyskytuje na málo stránkách), tím je výhodnější na toto slovo optimalizovat. Zkreslující je, že Wordtracker používá přesné výrazy, tedy dá výraz do uvozovek. Proto je potřeba ověřit i zadání bez uvozovek tak, jak to zadává většina lidí. Navíc počet stránek, na kterých se dané slovní spojení vyskytuje, nemusí přesně vypovídat o konkurenci.

Overture Search Term Suggestion Tool[70] se uplatní u vertikální analýzy a s počtem hledání v síti Overture. Pokud je zapotřebí rychlý údaj, pak je vhodné použít podíl sítě Overture na celosvětovém vyhledávání, kterým se získá odhad, kolikrát je dané slovo vyhledáváno celosvětově. Overture Search Term Suggestion Tool pro vyhledávání v němčině má i svoji vlastní stránku[71]. Overture Search Term Suggestion Tool má také drobnou nevýhodu, nerozlišuje totiž jednotné a množné číslo.

V případě Google AdWords Keyword Suggestions[72] mezi seřazenými slovy jak ve vertikální tak i v laterální analýze může existovat propastný rozdíl. Mnozí provozovatelé website pak tato slova vloží do reklamních sestav a domnívá se, že zobrazený počet interpretuje informace o počtu vyhledávání. To však je omyl. Toto číslo jenom představuje odhad kliknutí na reklamu v AdWords[73]. Přitom však platí dvě tendence. Pokud jsou přirozené výsledky více relevantní než placené, budou lidé klikat více na přirozené výsledky. Pokud budou přirozené výsledky nerelevantní a placené budou naopak odpovídat tomu, co uživatel hledá, budou tato čísla vysoká. Přesný poměr se nedá zjistit. Několikrát to bylo vyzkoušeno, ale protože výsledky přirozených i placených pozic se v čase mění, mění se i tento poměr.

[70] http://inventory.overture.com/d/searchinventory/suggestion/
[71] http://inventory.overture.com/d/searchinventory/suggestion/?mkt=de
[72] https://adwords.google.com/select/main?cmd=KeywordSandbox
[73] Google AdWords jsou cestou, jak nakoupit vysoce cílenou reklamu placenou za prokliky (CPC), bez ohledu na velikosti rozpočtu. Reklamy z AdWords se zobrazují podél výsledků hledání v Google a také na vyhledávacích a obsahových stránkách jeho partnerů.

V případě českých vyhledávačů je situace trochu specifická. Důvod je zřejmě takový, že proč by měly poskytovat něco, co má hodnotu, někomu, kdo jim za to nic nezaplatí – nebo bude dokonce přímo konkurovat v přirozených výsledcích? Na druhou stranu, pokud jim přináší nějaké peníze a má pro ně hodnotu, lze na základě dohody získat přístup k počtu vyhledávání na jednotlivých vyhledávačích. eTarget[74] ukazuje počet hledání denně na menších českých vyhledávačích.

Seznam, Centrum a Jyxo publikují (pravidelně nebo nepravidelně) seznamy nejvíce hledaných slov. Neobsahují sice většinou přesná čísla, takže mezi 10. a 11. místem může být rozdíl několika řádů. Nicméně slouží pro základní orientaci. Základem je ale získat čísla alespoň od jednoho reprezentativního zdroje. Zbytek pak lze dopočítat podle poměru jednotlivých vyhledávačů[75]. Ideální postup to samozřejmě není, ale určitou vypovídací schopnost to má.

Obecně mají ale databáze jednu velkou nevýhodu, ukazují aktuální údaje. Vzhledem k tomu, že mnoho prodejů je sezónních, je tedy k ničemu vědět v březnu, jak je vyhledáváno spojení např. „dárky k Vánocům“ nebo v listopadu „Chorvatsko last-minute“. To pak značně degraduje všechna udávaná čísla při dlouhodobém sezónním plánování [25].

6.2.1 Profesionální nástroje

Hitwise[76], především pak jejich služba Hitwise Search Terms, je špičkovým nástrojem téměř každého SEO konzultanta. Hitwise totiž monitoruje routery velkých ISP – a tudíž i chování provozovatele website. Z toho se pak dá vyzkoumat mnoho důležitých informací: jaká slova jsou vyhledávána, zda je klikáno na placené nebo neplacené výsledky, na jaké přesně, zda se provozovatel dostává ke konkurenci přes slovo „ABC“ nebo „BCA“, zda ke konkurenci chodí lidé přes bannery nebo přes vyhledávače – možností je mnoho.

Hitwise není čistě americká služba. Ve Spojených státech monitoruje internetové chování deseti milionů uživatelů a celosvětově 25 milionů. Bohužel, český trh samozřejmě nezahrnuje. Respektive: je možné si nechat udělat analýzu o českých websites, ale jenom z pohledu zahraničních návštěvníků.

[74] http://www.etarget.cz/
[75] http://www.toplist.cz/global.html
[76] http://www.hitwise.com/

qSearch[77] je služba firmy comScore, která má reprezentativní panel 1,5 milionů uživatelů, většinu samozřejmě opět v USA. Podobně jako Hitwise je qSearch silný ve výzkumu konkurence. Opět je možné se dozvědět, kolik lidí, přes co a jak navštěvuje konkurenci či jaký je konverzní poměr u konkurence. Konkurence například neví, že největší obrat jí dělají návštěvníci přicházející přes „lcd monitors“ přes Overture na MSN Search [25].

6.2.2 Váhy jednotlivých slov

Ne každé slovo je pro provozovatele website, respektive pro jeho zákazníky, stejně zajímavé. Např. může jít o serverhosting a v analýze vyšlo, že lidé, kteří hledají serverhosting, mohou zadat i slovo „webhosting“. Tito potenciální zákazníci nemusí vědět, že hledají serverhosting, myslí si, že webhosting je nadmnožina serverhostingu. Je tedy možné je z webhostingu přemluvit na serverhosting apod. Na druhou stranu však nemusí jít o nejlepší zacílení. Spousta lidí hledá opravdu webhosting, jiní nevědí, co serverhosting znamená, a zbytek nemá na serverhosting peníze.

Aby bylo možné určit, zda má význam optimalizovat WWW stránky na slovo webhosting, je nutné využít snížení váhy tohoto slova. Existují dva přístupy. V prvním je procentuálně snižována váha slova. Tedy např. pokud „serverhosting“ má hodnotu 100 procent, „webhostingu“ je přisouzeno 80 procent. V druhém je sledována cílová stránka/website takovéto optimalizace. Na základě váhy slova, cílové stránky/website, zkušeností a trhu je stanoveno, kolikátý každý návštěvník se stane zákazníkem (konverzní poměr).

Existuje ještě jedna váha, související s předešlou, na základě které se musí snížit váha jednotlivých slov: a to, jak je dané slovo vyhledáváno pro koupi dané komodity nebo alespoň podporu prodeje. Např. klíčové slovo „jet fighter“ bylo v síti Overture vyhledáváno v červnu 2005 32.432krát. S trochou nadsázky by bylo možné tvrdit, že se jedná o „zlatý důl“ pro výrobce stíhacích letounů.

Zde je proto namístě snížit váhu slova podle toho, jak se používají tato slova pro komerční účely. Váhu je třeba uplatnit i u běžnějších slov – obzvláště pak v podmínkách České republiky, kde jsou lidé hodně zvyklí zadávat jednoslovné výrazy. Např. pokud někdo hledá *Chorvatsko*, nemusí hledat nutně ubytování nebo dovolenou v Chorvatsku. Může hledat informace o Chorvatsku, kam pojede autem

[77] http://www.comscore.com/metrix/search.asp

a má již zaplacenou dovolenou, informace do seminární práce, cestopis Chorvatska nebo sezónní práce v Chorvatsku.

Nejpozději v tomto bodě je nutné nasadit filtr a výše uvedená slova profiltrovat. Pokud by byla všechna použita v dalším bodu, analýza by trvala příliš dlouho. Filtr je možné stanovit různým způsobem – na základě pravděpodobného počtu zákazníků, KEI (tu je ale vhodné upravit), dle počtu hledání sníženého o danou váhu nebo vše dohromady. Z důvodu rozsahu práce budou zmíněny jen ty nejdůležitější. Následně by mělo zůstat maximálně několik málo desítek klíčových slov, která budou dále analyzována (záleží na zadání).

Analýza konkurence

Aby bylo možné zjistit, jak náročné je optimalizovat WWW stránku na konkrétní slova, musí se provozovatel website zaměřit na konkurenci. Bez toho neodhadne:

- co všechno bude muset udělat pro úspěšnou optimalizaci WWW stránky,
- jak velká bude finanční náročnost úspěšné optimalizace WWW stránky.

Nyní jsou brána jednotlivá klíčová slova, jsou zadávána do vyhledávačů a výsledky jsou „proklikávány“. V tomto případě jsou výsledné stránky analyzovány z hlediska optimalizace (počet zpětných odkazů, Title, Description, Ranky, bohatost obsahu, použití nadpisů, CSS a dalších cca 90 faktorů[78]).

Marže u jednotlivých produktů

Provozovatel website u každého produktu vezme marži, kterou na něm má. Marži vynásobí odhadem počtu zákazníků u každého slova a porovná s cenou optimalizace. V tomto okamžiku by již měl mít jasno, do čeho se pustit a do čeho ne. Samozřejmě, může se pokusit o křížový prodej: tedy že bude optimalizovat na slovo A, ale následně pak souběžně nebo namísto toho prodá zákazníkovi produkt, na kterém vydělá více.

Doba trvání optimalizace

Většinou čím je slovo náročnější na optimalizaci, tím také déle bude trvat samotná optimalizace. V případě slova „Chorvatsko“ není vhodné optimalizovat

[78] http://www.google.com/intl/cs/webmasters/4.html

website na toto slovo v květnu, neboť dokončení by přišlo někdy po sezóně. Proto je výhodnější pustit se do „menších slov“, která třeba nejsou tak výhodná a nepřichází přes ně tolik zákazníků – ale lze oslovit trh, dokud mají potenciální zákazníci zájem.

S vylepšením psychodemografického sledování uživatelů Internetu[79] bude možné zadat, koho vlastně provozovatel website sleduje. Tím přesně zjistí, že např. zatímco majorita vyhledává „*chorvatsko dovolená*“, jeho cílová skupina „*chorvatsko cestovní kancelář*“ [25].

6.3 Úprava serveru

Internet pracuje v modelu klient/server. Při provádění určitého úkolu tak spolupracují dva počítače, které si vzájemně zasílají různé potřebné informace. Nejběžnějším scénářem takovéto dvoustranné spolupráce je komunikace mezi serverem, tedy počítačem, na němž jsou uloženy jisté informace, a klientem, čili počítačem, který tyto informace požaduje [27].

„Přívětivé“ URL je takové, které neobsahuje otazník, za kterým si script předává tzv. parametry mezi WWW stránkami. Příkladem „klasického“ URL je např. `http://www.example.com/clanek.php?clanek=1506`. Zde parametr identifikuje, že se má zobrazit článek číslo 1506. „Přívětivé“ URL potom může být například `http://www.example.com/nadpis-clanku.html`.

Důvody pro užívání „přívětivých“ URL jsou:

- většina vyhledávačů totiž vyhledává klíčová slova také v URL a přikládá jim významnou váhu,
- vyhledávače většinou neindexují stránky s mnoha parametry (například Google nevyhledává ve stránkách s více jak třemi parametry a do začátku loňského roku dokonce nevyhledával ve stránkách s více jak dvěmi parametry a ještě k tomu se ani jeden parametr nesměl nazývat id).

Další výhodou „přívětivých“ URL je také:

- menší riziko, že se při ručním přepisování někdo překlepne,
- z odkazů je ihned patrné, o čem stránka je.

Apache a modul Rewrite

[79] http://ihned.cz/2-12979110-000000_d-e5

Server Apache obsahuje neobyčejně stručný modul `mod_rewrite.c`, který sám o sobě tvoří takřka úplný samostatný softwarový produkt. Modul Rewrite uplatňuje na URL přepisovací masku, díky čemuž lze použít „přívětivá“ URL, aniž by bylo nutné provádět velké zásahy do již hotových website (měnit celý redakční systém apod.) [21]. Pokud například jsou nyní používány adresy jednotlivých článků ve tvaru `http://www.example.com/clanek.php?clanek=1208`, pomocí `Mod_Rewrite` ji lze změnit na `http://www.example.com/clanek/1208`.

Internet Information Server a modul ISAPI_rewrite

U Internet Information Serveru (IIS) je možné přepisovat URL pomocí `ISAPI_rewrite`[80], ovšem za předpokladu, že je tento modul do IIS doinstalován. Modul naprogramovala firma Helicon Tech a je ke stažení na stránkách modulu `ISAPI_rewrite` ve verzích full a lite. ISAPI_rewrite Full je zdarma k dispozici pro otestování na 30 dnů, poté se musí zaplatit. Verze lite je k dispozici zdarma a je identická s verzí Full s několika omezeními.

Přesun stránek na novou adresu a změna redakčního systému

S přesunem stránek či změnou redakčního systému je většinou mnoho komplikací. Z pohledu SEO je nejzávažnější ztráta zpětných odkazů, podle kterých většina vyhledávačů spoluhodnotí úroveň stránek. Pokud jsou tedy stránky přesouvány na novou doménu či subjekt přechází na jiný redakční systém, měla by všechna stará URL být přesměrována na nová.

V zásadě se naskytují dvě možnosti. V prvním případě je nutné zjistit, kde všude se vyskytují stará URL, a požádat o změnu na nová. V druhém případě se jedná o pokus o přesměrování odkazů vlastními silami provozovatele website, popř. SEO konzultanta. První volba je lepší, ale poněkud nepraktická, zdlouhavá a hlavně nespolehlivá, protože moc serverů nebude ochotných měnit odkazy. Nevýhoda druhé je ta, že je nutné uchovat poměrně dlouho stará URL, za kterými se bude vyskytovat přesměrování na nová, což je velký problém při přesunu stránek na novou adresu, přičemž starou je třeba co nejdříve opustit [25].

[80] http://www.isapirewrite.com/

6.4 Kód WWW stránky

Jak „vypadá“ WWW stránka včetně neviditelných atributů ALT rozhoduje z velké míry o tom, jaké budou výsledky ve vyhledávačích. Je to často důležitější než zpětné odkazy. Vždycky pomůže mít více smysluplného textu než toho nic neříkajícího – značek (tagů).

Nejdůležitější pravidlo SEO zní: „Tvůrce WWW stránek musí myslet jako vyhledávač“. Jak ale myslí vyhledávače? Odpověď je jednoduchá: myslí tak, jak myslí jeho tvůrci a provozovatelé. A jak myslí tvůrci a provozovatelé? Tvůrci a provozovatelé website myslí tak, aby poskytli uživatelům ty nejlepší výsledky při vyhledávání. Pokud neposkytnou nejlepší výsledky při vyhledávání, uživatelé navštíví jiné WWW stránky. Od návštěvnosti WWW stránek se mohou odvíjet i finance za reklamy umístěné právě na navštívených WWW stránkách.

Podobné myšlenky zřejmě vedly i tvůrce nového MSN Search. Chtěli, aby lidé dostávali především kvalitní přirozené výsledky. Proto přesunuli většinu reklam z horních pozic do pravého sloupce. I když to bude znamenat menší CTR[81] a tím zdánlivě méně peněz, zřejmě předpokládají, že díky zvýšenému počtu uživatelů jejich vyhledávače bude peněz ve skutečnosti více. Takto uvažoval i Yahoo!, když změnil své primární vyhledávání ze svého placeného katalogu na fulltext.

Vyhledávač nemá žádný zájem na tom, aby byla zrovna konkrétní stránka na vrcholu výsledků vyhledávání. Jeho zájmem je, aby poskytoval ty nejlepší výsledky vyhledávání pro své uživatele. Samozřejmě si každý myslí něco jiného o tom, co je nejlepší pro jeho uživatele a jak toho dosáhnout.

Zdrojová kód

Jak vypadá zdrojový kód je důležité. Obecně platí, že čím více obsahuje zdrojový kód skutečného textu, tím lépe. Obecně se vyplatí používat kaskádové styly (CSS) pro formátování WWW stránky. Ty navíc umožní „vytáhnout“ nejdůležitější text co nejvýše, neboť co je výše, má větší váhu, než to, co je níže.

S tímto tématem souvisí i používání externích souborů: vše, co je možno a co přímo nesouvisí s obsahem stránky, je převedeno do externích souborů. To platí samozřejmě primárně pro CSS a dále pro JavaScript. Kód stránky se opět „zahustí“. Samozřejmě se vyplatí, když je kód validní a splňuje kritéria přístupnosti.

[81] Click Through Rate

Prvky kódu

Vyhledávače sledují téměř vše. Samozřejmě, jsou věci důležitější a méně důležité. Ale když už se provozovatel website pouští do optimalizace, je vhodné optimalizovat vše.

- `title`,
- `description`,
- `keywords`,
- ostatní meta značky,
- nadpisy (`h1`, `h2`...),
- začátek stránky,
- tvar URL dané stránky,
- tvary URL na stránce (např. u obrázků),
- alternativní popisky u obrázků,
- `title` u odkazů,
- texty ve FLASH animacích,
- odkazy ze stránky (jejich podoba),
- text samotné stránky,
- používání tagů `<strong> </strong>` a `<em> </em>`.

Váhy u jednotlivých vyhledávačů se liší (mezi sebou i v čase).

Náhodně generovaný obsah

Náhodně generovaný obsah může být velmi dobrý „SEO trik“. Pokud bude na stránkách stále něco nového,

- roboti vyhledávačů budou na stránky chodit častěji, protože se na nich „stále něco děje“ a oni chtějí mít přehled o aktuálním dění,
- posunou stránky výše ve výsledcích vyhledávání, protože lidi více zajímají nové, aktuální a aktualizované informace než něco starého.

Trik tedy spočívá v tom, aby na stránkách byl určitý generátor náhodného obsahu. Vyhledávací roboti si při každé návštěvě budou myslet, že se jedná o nový obsah.

Hustota klíčového slova

Velké polemiky se vedou kolem hustoty klíčového slova. Ta se samozřejmě mění čas od času a vyhledávače od vyhledávače. Vhodné je použít výsledky vyhledávání na dané slovo a odstranit ostatní faktory. Na hustotu klíčového slova existují speciální nástroje. Ostatně, sledování konkurence je velmi účinná SEO technika [25].

6.5 SEO copywriting

Copywriting, tedy vlastní text a jeho podoba, rozhoduje z velké části o skutečném úspěchu celá optimalizace pro vyhledávače. Samotná pozice totiž nezaručí, že na stránku někdo přijde. A i když už na stránku někdo přijde, že si něco koupí.

Jak již bylo popsáno výše, je nutné nejprve vědět, koho stránky budou oslovovat. Každý člověk má totiž jiné „spouštěče“ a reaguje jinak. Pokud je např. prodáván prostředek na potenci, bude jiný text oslovovat lékaře, jiný důchodce ze střední třídy a jiný pro přepracované manažery. V případě, že se to vyplatí, může existovat pro každou cílovou skupinu jiný web. Na všech se bude prodávat to samé, ale pokaždé se bude klást důraz na něco jiného.

Je nutné používat výrazy a symboly skupiny, která je oslovována. Ideální je nechat text přečíst někomu z dané skupiny a sledovat jeho reakce.

Title

Title, neboli, co uživatelé uvidí jako nadpis ve vyhledávačích, rozhoduje podle různých studií z 80 procent o tom, zda na WWW stránky uživatel klikne, nebo ne. Uživatel většinou zadává do vyhledávače podstatná jména (případně přídavná jména) – a podstatná jména potom také vyhledává, když „skenuje“ stránku s výsledky vyhledávání. Více nutící k akci jsou ale slovesa. Například Google opravdu zvýhodňuje titulky stránky, které jsou větami nebo se aspoň větám podobají (to platí alespoň pro angličtinu). Proto je vhodné psát titulky jako větu, kde bude jak podstatné jméno (klíčové slovo), tak k akci pobízející sloveso. Velkou výhodu zde má

angličtina, která zabere zhruba o jednu třetinu méně místa se stejným významem než čeština.

Jiný „trik“ je umístit do titulku užitečnou informaci. Pokud je to možné, vhodné je do titulku umístit i název firmy nebo výrobku. Pokud tak provozovatel website neučiní, 80 procent lidí se nikdy nedozví, co to vlastně propaguje.

Description

Description je další text, který uživatel uvidí ihned ve vyhledávačích. Je vhodné v něm rozvinout myšlenku z titulku včetně klíčových slov.

Dlouhý či krátký text

O tom, zda by text měl být krátký nebo dlouhý, se vedou velké spory. Přitom autoři zapomínají, že je důležité, v jaké fázi se potenciální zákazník nachází. Existují totiž tři fáze nákupního procesu:

- prvotní seznámení,
- zjišťování podrobností,
- nákupní rozhodnutí.

V první fázi je nutné zákazníkovi nabídnout krátký text, aby jej zaujal. Poprvé text jen prolétne. Podívá se i na konkurenci. Ale pokud krátký text nabídl to, co on potřebuje, vrátí se (nebo bude přímo pokračovat) k dlouhému textu. Nyní chce totiž zákazník vědět všechno o tom, co bude vlastně kupovat. A protože si zboží nemůže „osahat“, potřebuje maximum informací.

V poslední fázi je třeba zákazníkovi nabídnout zase krátký text. Nyní se již jen potřebuje utvrdit ve svém rozhodnutí nakupit a přesvědčit sám sebe, že dělá dobře. Nebo se rozhoduje mezi produktem provozovatele website a konkurencí.

Cena

Slovo *cena* (v našich podmínkách i *ceník*) je na Internetu velice vyhledávané. Proto je vhodné na toto slovo cely website optimalizovat a uvádět cenu i v textu. Pokud zákazník neuvidí cenu příslušného produktu, pravděpodobně půjde jinam. Ostatně konkurence je jenom o jeden klik vedle. Pokud provozovatel website operuje na americkém trhu, vyplatí se používat i slovo „discount“.

Grafická úprava

I na grafické úpravě záleží. Vůbec neuškodí, pokud stránky budou správně přístupné. Navíc ale je vhodné:

- používat krátké věty a krátké odstavce. Někdy stačí i dvě věty na odstavec,
- na webu používat bezpatkové písmo (u tisku je tomu přesně naopak),
- nepoužívat pro hlavní text tmavé pozadí a světlé písmo
- nepoužívat v nadpisech kapitálky (velká písmena),
- v navigaci používat text, nikoli primárně grafické symboly.
- konec nadpisu použít bez tečky. Tečka značí konec. Nadpis je ale teprve začátek,
- hlavní text by neměl být ani moc široký, ani příliš úzký. Úzký je ale obvykle lepší. Pokud bude text příliš široký, budou muset čtenáři kroutit hlavou. U užšího budou muset aspoň „kroutit“ očima.
- mezititulky použít zhruba každých pět až sedm centimetrů textu. Pomůže to i SEO jako takovému,
- nejdůležitější informace vložit mezi `<strong> </strong>` a `<em> </em>` nebo do tabulky,
- odsadit první řádky nového odstavce od kraje. Ideální by bylo, kdyby první písmeno každé stránky bylo zformátováno zvláštním písmem,
- když je nutné pospojovat řadu vzájemně nesouvisejících informací, je dobré použít odrážek nebo číslování

Hustota, vzdálenost, frekvence a tvary slov

Hledání správné hustoty bylo vysvětleno v předcházející části. Problém však může nastat, když je text optimalizován např. na klíčové slovo „počítače cena“. Věty, kde by vedle sebe byla slova „počítače“ a „cena“, se píší dosti obtížně. Proto by tato dvě slova měla figurovat co nejblíže vedle sebe. Klíčové slovo, na které je website optimalizován, by také mělo mít v textu nejvyšší frekvenci.

Google a další vyhledávače bohužel neumějí skloňovat ani časovat. A i když vyhledávač umí ohýbat slova, stejně dává přednost tvarům, které uživatel přímo vyhledává. Proto je nutné zjistit, jak přesně zadává uživatel slovo

do vyhledávače. Na ten samý tvar se pak optimalizuje. Na druhou stranu neuškodí, pokud je v textu občas použit jiný než základní tvar. Tím se získají ti, kteří zadají do vyhledávače jiný tvar [25].

6.6 Zpětné odkazy

Zpětné odkazy jsou odkazy, které odkazují na konkrétní WWW stránku provozovatele. V SEO se získávají většinou výměnou. V případě, že na stránce je vysoce kvalitní obsah, budou lidé na tento web odkazovat sami, ovšem je nutné, aby provozovatel WWW stránky poskytoval velmi kvalitní obsah. Obzvláště to platí pro firemní stránky. Ze začátku tak musí s velkou pravděpodobností získat nějaké odkazy výměnou, tedy provozovatel website odkážete na stránku, ze které by rád získal zpětný odkaz. Získá tím již výše zmíněný rank, hodnocení, které používá většina vyhledávačů (Google PageRank, Yahoo WebRank[82], Jyxo JyxoRank[83] apod.).

Google však nerozlišuje (i když se již objevují opačné názory) mezi interními odkazy (v rámci jednoho webu) a externími odkazy (z jiných domén). Z toho vyplývá, že velké weby mají výhodu, a pokud mají vhodně vyřešenou strukturu stránek, budou mít vyšší PageRank než menší websites.

Vysoký PageRank má do určité míry vysoký význam, avšak nikdy nebyl nejdůležitější při řazení výsledků, ale přeci podstatnou váhu. Při vyhledání libovolného slova lze zjistit, že např. stránka na čtvrtém místě má PageRank 5/10 a na sedmém 3/10.

Google Toolbar PageRank je velice nepřesný (pokud vůbec udává správné hodnoty) a široký. Mezi GTPR 4/10 a 5/10 může být velice nepatrný (pokud mají stránky např. hodnotu 4948 a 5033), ale také propastný rozdíl (při hodnotách 915 a 27.213) hodnot skutečného PageRanku, který Google používá při svých výpočtech.

V každém případě platí, že čím méně je odkazů na stránce, tím vyšší hodnota se bude předávat v rámci jednoho odkazu. Jediný odkaz ze stránky 4/10 tak může mít vyšší hodnotu než odkaz ze stránky 5/10, kde je odkazů sto.

Při vytváření zpětných odkazů na WWW stránku provozovatele je nejdříve vhodné zvolit správný titulek s popiskem WWW stránky a pokračovat registrací do většiny katalogů, neboť jsou stále uživatelé, kteří je používají. Navíc jsou získány

[82] http://help.yahoo.com/help/us/companion/webrank
[83] http://jyxo.cz/d/faqw

zpětné odkazy obvykle zdarma. Převážná většina katalogů sice nemá extrémní hodnotu PR (na stránce se vyskytuje velké množství odkazů, mezi které se PR dělí), nicméně v žádném případě se taková registrace neprojevuje na website provozovatele negativně. Při hledání zpětných odkazů Google zobrazuje jenom zpětné odkazy ze stránek s PR 4/10 a vyšším. Pro zkoumání všech zpětných odkazů je tedy lepší využít Jyxo nebo Yahoo!.

Velkou váhu má text, který na příslušný website odkazuje. Pokud je použit text ve tvaru „Přijďte k nám nakoupit“, dostane stránka sice zpětný odkaz, ale zároveň se přiřadí jistá váha při vyhledávání na dotaz „Přijďte k nám nakoupit“, což je z hlediska optimalizace pro vyhledávače nedostatečné, neboť výraz „Přijďte k nám nakoupit“ nikdo nehledá. Proto je vhodnější použít titulek stránky, který obsahuje klíčová slova.

Dalšími kroky mohou být:

- nákup specializovaného softwaru (Arelis[84], Zeus[85] apod.),
- pracné ruční hledání,
- přidání odkazu do „zaručených seznamů“ ochotných vyměnit odkazy.

První a třetí možnost nemusí být nejvhodnější. Specializovaný software se zaobírá kvantitou a ne kvalitou odkazů. Bere stránky na předních pozicích, ale nevšímá si ochoty vyměňovat odkazy, poslední aktualizace, případně PR stránky, na které bude uveřejněn odkaz na website. „Zaručené odkazy“ jsou zase často linkové farmy, které jsou Googlem penalizovány. Ruční hledání je sice nejpracnější, ale také nejvhodnější. Pokud provozovatel website zvažuje, že by si s nějakým website vyměnil odkazy, pak by si na stránce, kde se objeví jeho odkaz, měl zkontrolovat:

- zda jsou stránky vhodně řešené (např. nemají parametry, které většina vyhledávačů nepodporuje),
- meta tagy,
- soubor `robots.txt` na daném serveru (např. `www.example.com/robots.txt`).

Zřídka se stává, že mají stránky s odkazy zakázané sledovat odkazy vyhledávači (`<meta name="robots" content="nofollow" />`), případně

[84] http://www.ibusinesspromoter.com/linkbuilding/link-builder.htm
[85] http://www.cyber-robotics.com/zeusindex.htm

indexaci celé stránky (`<meta name="robots" content="noindex" />` v meta tagu a/nebo `Dissallow: meno-stranky-s-odkazy.htm` v `robots.txt`). Odkaz na takové WWW stránce nemá prakticky žádnou cenu, i když ne všechny vyhledávače dodržují standardy (ty největší však ano).

Provozovatel website by si tedy měl vést seznam, s jakými servery má vyměněny odkazy. V drtivé většině se nevyplatí vyměňovat odkazy se servery, které mají PR dané stránky 4/10. V krajním případě může stačit 3/10, ale méně rozhodně ne. Proto platí: pokud nemá stránka (ta s odkazy, ne hlavní, 4/10), pravděpodobně si nikdo odkazy nevymění.

Rozhodně by provozovatel website neměl vyměňovat odkazy s „adult“ obsahem. Pokud je na pochybách, raději by měl výměnu odmítnout. Google totiž stránky s porno a sexuální stránky tvrdě penalizuje. I když může být stránka o turistice, odkazem na porno stránku, za který může stoprocentně provozovatel (i diskuse lze monitorovat a čistit), se dostává do „špatného sousedství“. Tato snaha postihovat porno obsah jde tak daleko, že vyhledávače penalizují celé IP adresy. Pokud sídlí na stejné adrese jako porno server, můžou se vyskytnout problémy.

Kolem výměny zpětných odkazů existuje mnoho mýtů:

- Penalizace za příliš mnoho odkazů na stránce. Na druhou stranu, s odkazy to není třeba přehánět.
- Penalizace za Links ve jméně souboru, odkud vedou odkazy.
- Penalizace za cross-linking – stránky s podobným obsahem, které vymění rychle velké množství odkazů s jinými stránkami.
- Sandbox effect – podle této teorie mají stránky po spuštění prudký vzestup ve vyhledávání. To je pravda. Dále však tato teorie uvádí, že po prudkém vzestupu ve vyhledávání stránka na dva až čtyři měsíců zmizí z výsledků vyhledávání, bez ohledu na cokoli jiného [25].

6.7 Kontinuální SEO

Občas je SEO přirovnáváno k běhu na dlouhou trať. Nevýhodou však je, že samotná optimalizace pro vyhledávací roboty nikdy neskončí. Pokud se však určitý website bude držet na předních místech ve výsledcích vyhledávaní, bude zabírán „televizními kamerami“ a sponzoři budou platit.

Log soubor je hlavní materiál, se kterým tvůrci a provozovatelé website pracují. Jak často je nutné s ním pracovat závisí na povaze projektu. U některých projektů stačí jednou za měsíc, u některých například každý den. Sleduje se hlavně:

- odkud lidé přicházejí,
- na jaká slova přicházejí,
- odkud přicházejí lidé přes jaká klíčová slova,
- z jakých zemí návštěvníci pocházejí (lze lokalizovat i jednotlivá města).

Při použití cookies[86] je dále možné zjistit:

- kteří zákazníci jsou nejziskovější, což je nejdůležitější údaj. Přes slovo A mohou na website sice proudit davy, ale produkt či službu si objedná málokdo. Na slovo B přijde jenom několik lidí měsíčně, ovšem velké procento z nich nakoupí produkty za hodně peněz a ještě s velkou marží,
- jak se návštěvníci chovají. Je možné např. zjistit, že lidé, kteří přicházejí přes slovo X, se zastavují před objednávkou a třikrát se vrací. Kdežto lidé přicházející přes slovo Y navštíví web dvakrát a udělají nakonec ve 27 procentech objednávku za 353 dolarů. U zbývajících 63 procent je nutné provést další analýzu, co není v pořádku.

Podle výše uvedených analýz je možno upravovat celou optimalizaci. K analýzám se používá speciální software. Pro skutečně hloubkové a detailní analýzy to mohou být například ClickTraks[87].

Občas ovšem všechny analýzy selžou a velmi lukrativní klient může přijít přes slovo, popř. slovní spojení, které by provozovatele website nikdy nenapadlo. Může optimalizovat své WWW stránky na různá „typická“ slova, ale tohoto klienta by nikdy nezískal. Nebo potenciální klienti mohou používat výhradně jeden konkrétní vyhledávač. Pokud odkaz na provozovatelovi WWW stránky při vyhledávání na určité slovo nebude figurovat na předních místech, nikdy na něj „nekliknou“.

[86] Jako cookie se v protokolu HTTP označuje malé množství dat, která WWW server pošle prohlížeči a ten je uloží na počítač uživatele.
[87] http://www.clicktracks.com/entrypoint.php?a=22457

Dalším důležitým činitelem jsou zpětné odkazy. O zpětné odkazy se musí provozovatel website starat. A to aktivně! Neustále je buduje a sledujte, jak je na tom konkurence se zpětnými odkazy. Podle toho ví, jak intenzivně musí ještě zapracovat.

Čas potřebný na optimalizaci je velmi variabilní. Záleží na trhu. Na snadná slova jsou zapotřebí týdny (do ukázání výsledků ve vyhledávačích). Na těžká měsíce. Na velmi těžká třeba rok i déle. Ostatně, Google sám uvádí, že jenom změna URL mu zabere šest až osm týdnů. SEO není žádná kouzelná hůlka, kterou SEO konzultant mávne a web se objeví na prvním místě ve vyhledávačích. Je to velice náročná a dlouhodobá práce [25].

7 Search Engine Marketing

Pojem Search Engine Marketing (SEM) je často překládán jako marketing ve vyhledávačích a je jedním z možných způsobů propagace provozovatelova website. SEM je možné představit zejména jako výhodný nákup placených odkazů ve vyhledávačích. Obrázkové nebo animované reklamy (netextové) jsou zatíženy tzv. reklamní slepotou, což znamená, že se je uživatelé naučili přehlížet. Textovým odkazům uživatelé stále ještě důvěřují a klikají na ně, proto má nákup textových odkazů smysl a může se vyplácet [20].

Zatímco tradiční SEO je orientováno spíše na dílčí úpravy webových stránek, které mají za cíl co nejlepší umístění odkazů ve fulltextových vyhledavačích, SEM na základě pečlivé analýzy nejprve formuluje účinnou strategii a tu pak aplikuje nejen v oblasti typických fulltextových vyhledavačů, ale i na katalogy stránek a vyhledavače typu pay-per-click.

Význam i obliba SEM neustále stoupá a zejména v anglicky mluvících zemích dnes zaujímá v oblasti internetového marketingu bezkonkurenčně přední místo. V USA veliké firmy utrácejí za průběžnou aplikaci SEM až stovky tisíc dollarů ročně, ale i ty nejmenší projekty obvykle se SEM počítají již při svém zrodu.

I v České republice se zájem o SEM pomalu ale jistě zvyšuje, i když jeho důsledné a komplexní využívání je spíše výjimkou. Z části je to zřejmě dáno mnohem menším počtem stránek v českém jazyce. České katalogy nejsou zdaleka tak přeplněny jako ty mezinárodní, anglicky psané. Provozovatelům website proto přinášejí určitou návštěvnost i bez pečlivější optimalizace.

Komplexní SEM se orientuje do tří hlavních oblastí, resp. využívá 3 typů internetových služeb:

- katalogy stránek
- fulltextové vyhledavače
- pay-per-click vyhledavače

Především pro anglicky psané stránky jsou pay-per-click (PPC) vyhledavače velmi efektivním nástrojem zvyšování návštěvnosti. Typickým zástupcem

a průkopníkem PPC modelu je vyhledavač Overture[88] (dříve GoTo.com) a poměrně rychle přibývají další.

V případě PPC vyhledavačů na optimalizaci stránek pro vyhledavače nezáleží. Namísto toho provozovatelé website za dobré umístění platí. Princip fungování je velmi jednoduchý: je-li provozovatel website ochoten zaplatit 10 centů za každého návštěvníka, který se z výsledků vyhledavače proklikne na jeho stránky a zároveň jeho konkurent je ochoten zaplatit jen 9 centů, budu ve výsledcích na prvním místě a konkurent bude až za ním. Ovšem jen do okamžiku, než se rozhodne platit za proklik 11 centů.

PPC vyhledavače tedy uplatňují 2 základní principy:

- platí se jen za proklik
- pozice pro určité klíčové slovo či frázi se získává v dražbě

Významnou výhodou PPC vyhledavačů je jejich operativnost. Velmi snadno lze s jejich pomocí organizovat krátkodobé kampaně a výsledky se projeví velmi rychle. Výhodou je i transparentní řízení nákladů na návštěvníka. Vykazují-li provozovatelovi WWW stránky zároveň vyrovnaný konverzní poměr a tržby na návštěvníka, má ziskovost svého webu plně pod kontrolou.

Na druhou stranu je třeba vzít v úvahu, že správné použití PPC vyhledavače není tak triviální, jak by se na první pohled mohlo zdát. Snad ještě více než u tradiční SEO zde hraje roli perfektní analýza cílového segmentu klíčových slov. Jsou slova, která získají 20 návštěvníků denně, ale stojí dollar. Pečlivým výběrem ovšem lze nalézt jiná slova, která sice přivedou pouze 5 návštěvníků, ale cena je jen 10 centů za jednoho [32].

7.1 Matice FCB a faktory ovlivňující SEM

Provozovatel website se může rozhodnout pro Search Engine Marketing dle dvou základních modelů. První model je spíše kvalitativní a vyplatí se pro rychlé rozhodování. Druhý je na první pohled také kvalitativní, ale dá se velice dobře převést do čísel.

[88] http://www.overture.com/

Aby bylo možno zjistili, jaké produkty se vyplatí optimalizovat, lze nejprve použít matici, kterou v roce 1980 vymyslela agentura FCB[89], a která interpretuje, jaké faktory jsou ovlivňují koupi produktu.

Matice FCB		
	Rozhodnutí (motivy) **racionální**	Rozhodnutí (motivy) **emocionální**
Velká míra zaangažování kupujícího na rozhodnutí o nákupu	**Reklama informační** (např. životní pojištění, poloprofesionální digitální fotoaparát, motorový olej)	**Reklama afektivní** (např. automobily, značková kosmetika, značkové oděvy)
Malá míra zaangažování kupujícího na rozhodnutí o nákupu	**Reklama výrobků kupovaných automaticky** (např. prostředek na mytí nádobí, prášek na praní)	**Reklama, která představuje okamžitou satisfakci** (např. zmrzlina, sladkosti)

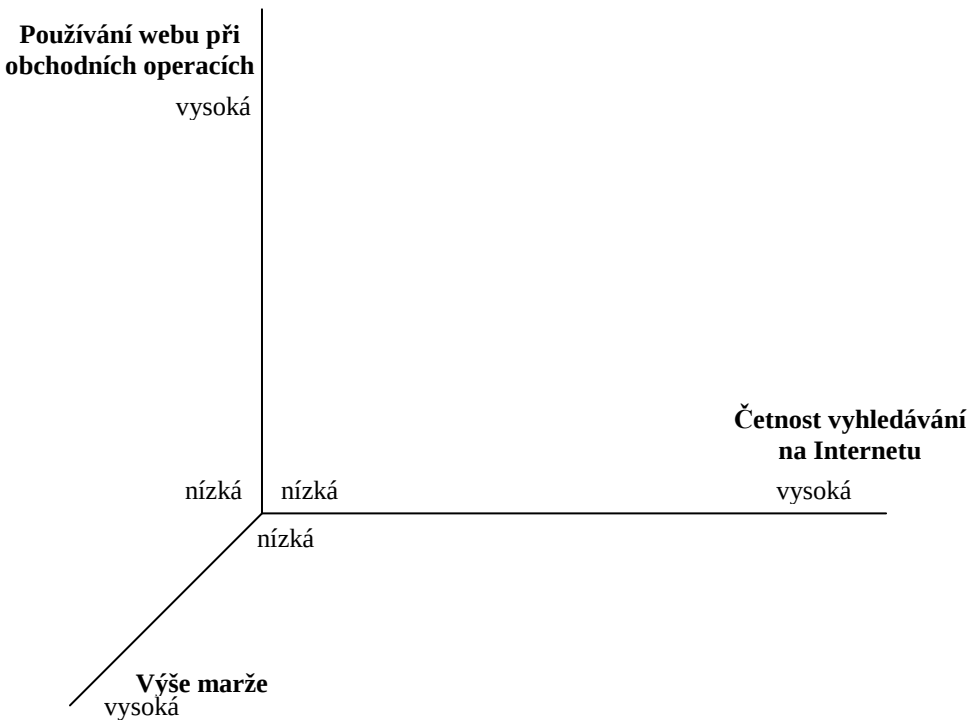
Obrázek 5: Matice FCB [25].

Nejvíce se pro Search Engine Marketing hodí produkty, kde:

- je velká míra zaangažovanosti na rozhodnutí (produkt je poměrně drahý, pokud člověk dobře vybere, hodně získá, jestliže vybere špatně, dost ztratí; tudíž má tendenci porovnávat a konkurenci vybírá na Internetu) a zároveň,
- se rozhoduje relativně racionálně – a informace opět hledá na Internetu.

Pole A1 se tedy vyplatí pro Search Engine Marketing nejvíce. Naopak pole B2 se nevyplatí pro search engine marketing vůbec. Firma Robert Němec znovu zanalyzovala prostředí Search Engine Marketingu a nakonec jim vyšly tři faktory, ze kterých sestavili trojrozměrný model. Podle něho se dá určit, zda se Search Engine Marketing vyplatí.

[89] http://www.fcb.com/



Obrázek 6: Tři faktory ovlivňující Search Engine Marketing [25].

Model se skládá ze tří os. První osa určuje, jak je produkt na Internetu vyhledáván. Čím více se lidé po něm ptají, tím více má Search Engine Marketing význam. Druhá osa ukazuje velikost marže na jednotlivém produktu (v nominální výši) a třetí osa potom prezentuje použití webu pro obchodování s produktem (myšlena celosvětová pavučina webových stránek).

Smysl dává teprve spojení těchto tří os. Mohlo by se zdát, že pokud produkt nepatří k nejvyhledávanějším, nemá pro něj Search Engine Marketing význam. V analýze může například vyjít, že přes Search Engine Marketing firma získá za rok jenom tři zákazníky. Pokud každý zákazník přinese vysoké tržby, pak se Search Engine Marketing zřejmě vyplatí. To ale předpokládá nutnost použít poslední osu – nakolik je pro uzavření obchodu používán web. Web není představován jen aplikací elektronického obchodu, ale nakolik se při různých fází uzavírání obchodu používá website, především pak ve spojení s vyhledáváním. Nemusí jít přímo o vyhledávání kontaktu. Potenciálnímu klientovi se může např. líbit konkrétní nabídka, ale bude ji chtít porovnat s konkurencí, kterou bude vyhledávat na Internetu. Jestliže však na dotaz „konkrétní produkt“ budou ve výsledcích vyhledávání určité domény (provozovatele website), pak mnoho konkurence neobjeví.

Slova jako Yahoo!, CNN, Google nebo Seznam patří k vysoce vyhledávaným. Bylo by však samozřejmě iluzorní předpokládat, že pokud budou chtít majitelé dotyčných podniků svůj podnik prodat, vyplatí se jim pro prodej podniku provádět Search Engine Marketing.

Čím blíže je produkt na modelu firmy Robert Němec pozici [0;0;0] neboli [nízká;nízká;nízká], tím méně se vyplatí Search Engine Marketing a naopak, čím více se přibližuje bodu [1;1;1] neboli [vysoká;vysoká;vysoká], tím více se Search Engine Marketing vyplatí. Při analýze situace pomocí výše uvedeného modelu je dobré si uvědomit, že se jedná o prostor. Body [0;0;0] a [1;1;1] jsou ideální typy, ke kterým se lze blížit, nikdy jich však nelze dosáhnout. Pozice produktu bude vždy nějaký bod v prostoru a podle jeho postavení se prodejce může rozhodovat, zdali je pro něj Search Engine Marketing vhodný, nebo ne. Bude-li v levé přední dolní polovině prostoru, pak ne, bude-li v pravé zadní horní polovině prostoru, pak ano.

Přesto se pro někoho výše uvedený model nemusí zdát dost dobrý. Byly identifikovány totiž ještě další dvě proměnné, které rozhodují o výhodnosti Search Engine Marketingu.

Čtvrtá proměnná určuje, jaké jsou náklady na vyřízení zakázky a je-li prodávající schopen zpracovávat velké množství zakázek. Obecně sice platí, že čím větší má marži, tím více si může dovolit investovat na získání jednoho zákazníka – a tedy se mu i více vyplatí investovat do Search Engine Marketingu. Na druhou stranu, dokáže-li stlačit své variabilní náklady a zvládnout – aby vykryl fixní náklady – velké množství zakázek, vyplatí se mu Search Engine Marketing více, než pokud to nedokáže.

Ale ani čtvrtá proměnná nemusí postačovat. I když prodávající má například velmi malou marži a vysoké náklady na vyřízení jedné zakázky a není schopen obsloužit velké množství zákazníků, přesto se pro něj může Search Engine Marketing hodit – pokud totiž dokáže prodat zákazníkovi něco dalšího s větší provizí [25].

7.2 SEM na Google AdWords

Pravý sloupek ve výsledcích vyhledávání na Google je naplněn reklamami z reklamního systému AdWords[90], přes který se očekává, že několik málo let půjde zhruba polovina obratu internetové reklamy. Pokud uživatel hledá na Google v češtině

[90] http://adwords.google.com/select/Login

během dopoledne po deváté hodině, dostává podstatně více AdWords inzerátů, než večer. S největší pravděpodobností příčinou je, že většina inzerentů přes den vyčerpá denní rozpočet a v noci už se inzeráty nezobrazují. Jakmile nastane v Kalifornii půlnoc, tak se rozpočty doplní a inzeráty se začnou zobrazovat znovu. Je na tom vidět důležitý aspekt inzerce v AdWords – možnost kontrolovat náklady. Kromě denního rozpočtu existují i měsíční a plus rozpočty na jednotlivé kampaně.

U AdWords zákazník platí za prokliknutí reklamy, inzerent má tedy větší jistotu, že mu jeho reklama něco přináší. Platba za prokliknutí reklamy je označována termínem Pay Per Click (PPC) a velkou zajímavostí AdWords je, že cenu za určité klíčové slovo si stanovuje sám zákazník. Za konkurenční slova je ale nutné nabízet více, aby Google reklamy zobrazoval.

Nakonec se Google při vypisování inzerátů ale nerozhoduje pouze podle ceny, kterou je inzerent ochoten zaplatit. Na stránce se inzeráty řadí také podle míry prokliku. Nabídnutá cena do toho vstupuje jen jako jeden faktor. Tím se AdWords liší od ostatních PPC systémů (např. eTarget[91] či Overture[92]). Google tím zřejmě sleduje dvě věci:

- Naučí hledající, že v pravém sloupku bývají zajímavé věci.
- Stimuluje inzerenty, aby co nejpřesněji cílili svoje reklamy na konkrétní klíčová slova, jejich kombinace a shody.

Když je určité slovo hodně konkurenční, Google nemá v pravém sloupku místo na všechny objednané reklamy a musí nějaké vyřadit. Kdyby vyřazoval podle ceny lacinější reklamy, nemusí v pravém sloupci zůstat žádná relevantní reklama. Lepší je zobrazovat výše a častěji ty reklamy, na které se více kliká. Úspěšnost klikání je označováno zkratkou CTR (Click Through Rate[93]).

Zájmem inzerenta je tedy mít reklamy s největší mírou prokliku. V České republice je míra prokliku zatím méně důležitá, protože konkurenčních slov je velmi málo a boj zatím není tak ostrý. Ale když někdo připraví reklamní text, který není proklikáván skoro vůbec, Google za nějaký čas přestává být ochoten jej zobrazovat, „naředí“ jeho zobrazování a případně ho pak úplně vyřadí [20].

[91] http://www.etarget.cz/
[92] http://www.overture.com/
[93] míra prokliku – poměr počtu prokliknutých inzerátů vůči počtu zobrazených inzerátů – vyjadřováno v procentech

Na některé dotazy nevrátí Google žádné reklamy a na jiné dotazy zobrazí inzerátů plno. Některá slova jsou tedy hodně konkurenční, čili drahá. Ta opravdu konkurenční slova zaplní celou pravou oblast stránky. Na českém hledání je zatím takových slov relativně velmi málo (například Webhosting nebo Chorvatsko), ale v budoucnu určitě přibudou. Google na nejvyšších místech zobrazuje inzeráty, které mají nejvyšší součin míry prokliku a ceny, kterou je inzerent ochoten zaplatit.

Na první pohled se zvyšování míry prokliku může jevit jednoduché, postačuje aby inzerent napsal slibné inzeráty, na které si každý uživatel klikne. Jenomže, za každé kliknutí se platí a pokud si návštěvník provozovatelova webu nic nekoupí, výborná míra prokliku nic neznamená. Je tedy nutné do inzerátů psát pravdu, aby přišli ti praví zákazníci.

Výhodnějším řešením než zvyšovat počet kliknutí je redukovat počet zobrazení. Pokud například uživatel zadá do dotazu vyhledávání klíčová slova „olympus opravna“ a v pravém sloupci se zobrazí inzeráty na klíčové slovo „olympus“, mohou být výsledky vyhledávání irelevantní, neboť inzeráty opravy neposkytují, chtějí pouze prodat fotoaparát. Majitelé inzerátů utrpěli zobrazení, za které sice nic neplatí, ale snižuje se jim míra prokliku.

Řešením může být, když inzerent přidá v účtu AdWords nechtěná slova mezi vylučující slova (negative keywords). Vyloučené slovo v dotazu potom znemožní zobrazení inzerátu. Nejen, že se může zvýšit míra prokliku, ale mohou se tak i uspořit peníze.

Další způsob, jak zvýšit míru prokliku, je používat co nejvíce specializovaná klíčová slova. Dokonce taková, o kterých neexistuje mnoho WWW stránek. Obecná slova, jako například *chorvatsko*, *webhosting* či *PDA*, jsou sice hodně vyhledávána, ale může být lepší jít směrem k většímu počtu méně hledaných slov. V češtině na ně není konkurence, a tak jsou extrémně laciná [20].

Proměna návštěvníka v zákazníka je označováno „konverze“ a „konverzní poměr“ vyjadřuje, kolik procent návštěvníků se stává zákazníkem. AdWords v sobě přímo obsahují nástroj na měření konverzí. Provozovatel website si může vygenerovat kousek kódu, který vloží do stránky potvrzení objednávky. Jakmile uživatel klikne na inzerát v AdWords, dostane cookie, a v případném uskutečnění prodeje si javascriptový kód cookie přečte a poznamená si úspěšnou konverzi [20].

Jak Google podle klíčových slov cílí reklamy do svého hledání, tak je schopen dodat relevantní textovou reklamu také na libovolnou stránku webu (pozná, na jaké téma je web zaměřen). Když si čtenář na reklamu klikne, platí inzerent stejný proklik jako z Google a majitel WWW stránky dostává provizi [20].

7.3 Agentura DoubleClick a Performics 50

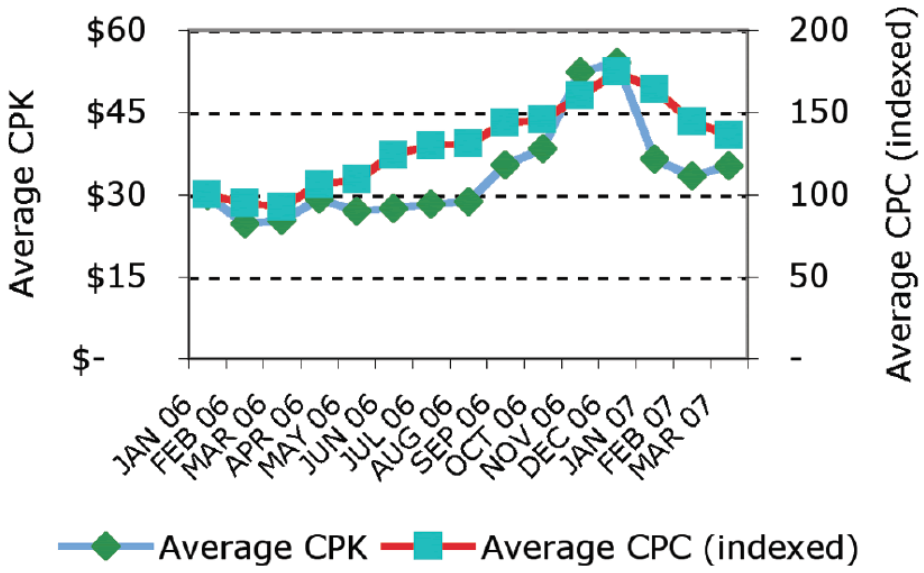
V polovině roku 2005 představila agentura DoubleClick Performics 50, index, který byl sestaven k zachycení vývoje vyhledávacích kampaní v dynamickém obchodním prostředí a k poskytnutí stálého základu srovnávacích analýz a hodnocení. Klíčové indikátory průměrné ceny za klíčové slovo (CPK[94]), průměrné ceny za klik (CPC[95]) a návratnosti investice (ROI) jsou poskytovány každé čtvrtletí s dodatečnými přehledy. Performics 50 prezentuje 50 aktuálních placených vyhledávacích kampaní řízených experty DoubleClick Performics SME používajících službu DART. Index byl stanoven v srpnu 2005, znamená míru marketingové kampaně na vyhledávačích a používá zpětně data z roku 2004.

Počáteční členství v indexu bylo určeno na základě velikosti inzertních kampaních v prvních třech měsících roku 2004 a posunuto na duben téhož roku. Nové kampaně byly přidány pouze, pokud se stávající kampaně v indexu významně odchýlily a byly opět založeny na velikosti pro aktuální měsíc a předchozí 3 měsíce. DoubleClick Performics řídí přibližně 280 aktivních kampaní napříč širokým rozsahem odvětví.

Pracovníci v DoubleClick Performics stále přemýšlí o vyhledávání a přidruženém marketingu. Nevytváří pouze budoucí předpovědi výkonu elektronického marketingu, ale jsou rovněž odhodláni k inovacím. Jejich ucelená kolekce poskytovaných služeb inzerentům (provozovatelům website) se světovou kvalitou řešení kombinuje technologie, inteligenci a zkušenosti. Zaměření na přístupy, placené vyhledávaní a marketing založený na vyhledávání bylo směrodatné k porozumění impulsů a akcí online spotřebitelů od roku 1998. DoubleClick Performics tak zabezpečuje, že marketingové investice jejich klientů jsou účinnější a efektivnější.

[94] Cost Per Keyword
[95] Cost Per Click

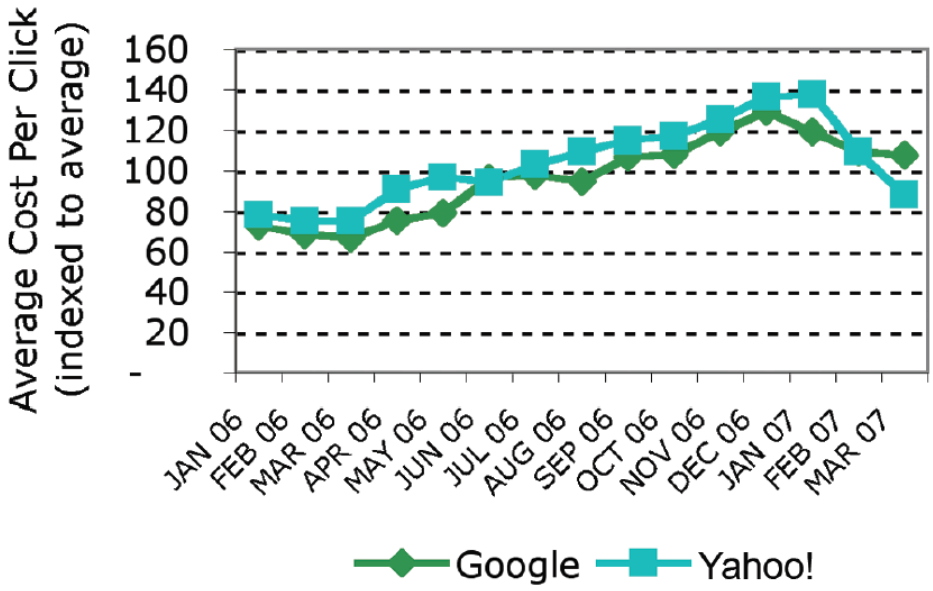
Ze zatím poslední zveřejněné zprávy DoubleClick Performics[96] vyplývá, že během prvního čtvrtletí roku 2006 rostl CPK v průměru o 28 % měsíčně, zatímco CPC o 55 % měsíčně. CPC klesl od prosince 2006 do ledna 2007 jen o 6 %. Ve srovnání s lednem 2006, kdy pokles byl 19 % z prémiových svátečních cen, se jedná o poměrně velký rozdíl.



Graf 5: Vývoj ceny za klik (CPC) a ceny za klíčové slovo (CPK) [15].

Pouze šestiprocentní pokles cen u CPC může být vysvětlen z důvodu migrace vyhledávače Yahoo! na novou platformu Panama, kde inzerenti můžou mít svá klíčová slova ve vyšších cenových hladinách než by byli ochotní platit na systému předchozím. Z grafu č. 6 vyplývá, že CPC na Yahoo! se skutečně v lednu zvýšil. Ačkoli má Panama určitou část na mírnějším trendu meziročního poklesu, existují zde určitě ještě další ovlivňující faktory. Důkazem je pokles CPC na Google v lednu 2007 oproti prosinci 2006 pouze o 8 %, ve srovnání s 18 % za stejné období v roce předešlém.

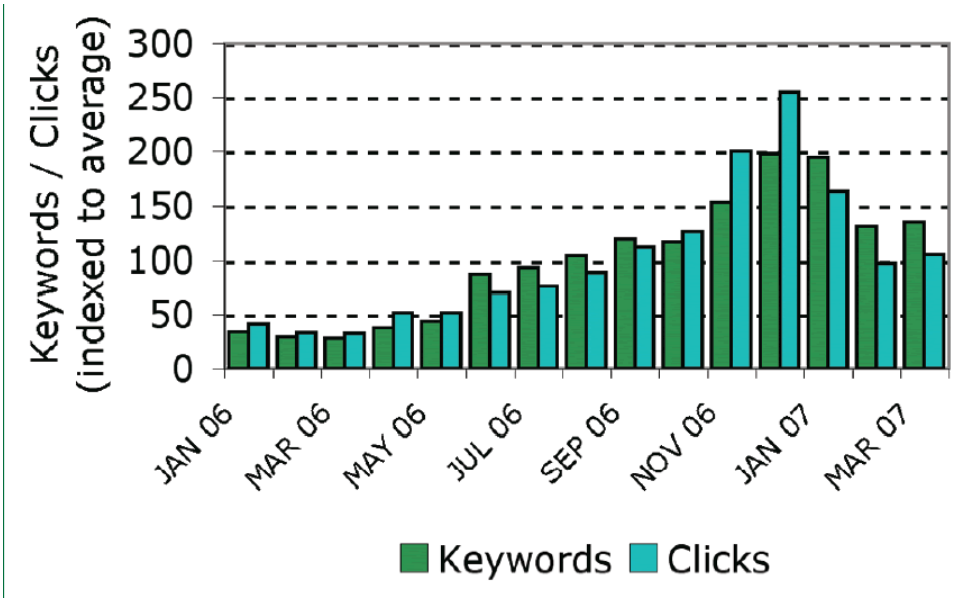
[96] DoubleClick Performics Q1 2007 Search Trend Report



Graf 6: Vývoj ceny za klik (CPC) za vyhledávače Google a Yahoo! [15].

Očividně tedy na diagramu 2 je příkrý pokles CPC od ledna do března pro klíčová slova na Yahoo!, která odpovídají zároveň nově představenému algoritmu výpočtu ranku z února 2007.

Největší složkou v meziročním nárůstu CPC a CPK je rapidní zvýšení drahých klíčových slov. První čtvrtletí 2007 ukazuje v průměru za měsíc šestinásobný počet klíčových slov dražších jak 1 USD v porovnání se stejným obdobím roku předcházejícího. Zatímco počet klíčových slov s cenou 0,20 USD a nižší v prvním čtvrtletí roku 2006 nerostl.

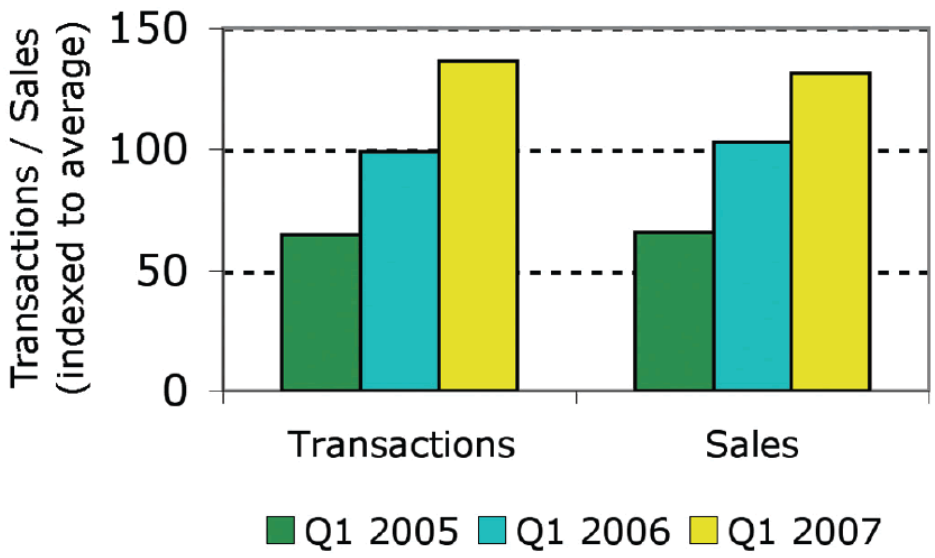


Graf 7: Nárůst drahých klíčových slov [15].

Podíl drahých klíčových slov na vyhledávacích kampaních nebyl o vánočních svátcích převládající a tedy tato klíčová slova nemohou po Novém roce kolísat s takovou mírou jako levná klíčová slova. Drahá klíčová slova tedy přispívají k vyšší základní čáře CPC, vycházející z vánočních svátků a snížení ve skutečných USD je reprezentováno malou procentní změnou.

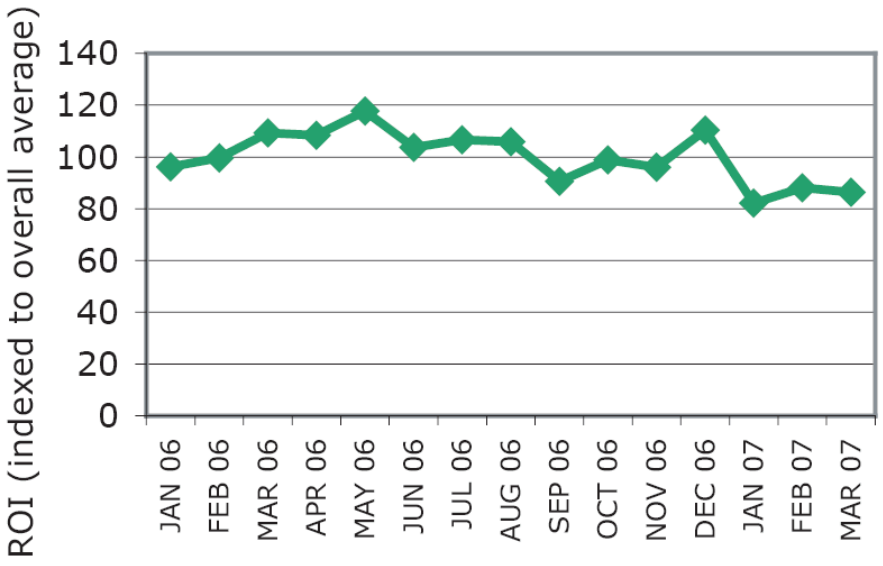
Zatímco drahá klíčová slova ovlivňují vysoké číslo zobrazení, obecně nejsou tak hodně „klikatelná“ jako „levnější“ slova. To vysvětluje, proč nárůst CPK není tak rychlý jako CPC a proč byl mezi prosincem a lednem vyšší pokles než u CPC.

První tři měsíce roku 2007 vynesly více dolarových prodejů než jakékoli první čtvrtletí v historii indexu Performics 50. Ve srovnání s posledním rokem, transakce narostly o 38 %, zatímco prodeje narostly o 28 %.



Graf 8: Meziroční změna transakcí a prodejů (pouze kampaně s přímými příjmy) [15].

S meziročním nárůstem ceny za klíčové slovo v průměru o 33 % měsíčně a nárůstem prodejů o 28 procent se očekává, že ROI bude nepatrně menší, možná i stagnovat. Podíl nákladů roste rychle a tím odvozený ROI je klesající. Ve srovnání s prvním čtvrtletím roku 2006, ROI poklesl o 16 %.

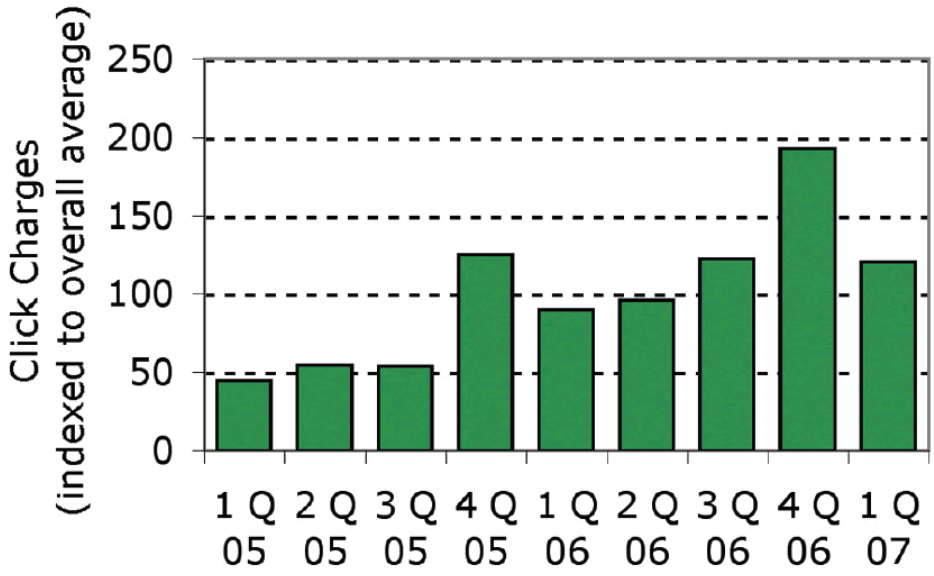


Graf 9: Návratnost investice (pouze kampaně s přímými příjmy) [15].

CPK a celkové prodeje v porovnání s cenami sice stoupají, počet klíčových slov však stoupá mnohem rychleji. Průměrný počet aktivních klíčových slov za měsíc pro první čtvrtletí 2007 bylo o 54 % vyšší než v roce 2006. Tato rapidní expanse klíčových slov, je zaměřena na klíčová slova, u kterých je cena za klik 1 USD a vyšší. Zatímco nárůst dražších klíčových slov je klíčovým faktorem v poklesu odvozeného ROI, hlubší analýza ukazuje, že tato klíčová slova mohou skutečně vylepšit výkonnost ostatních klíčových slov v kampani.

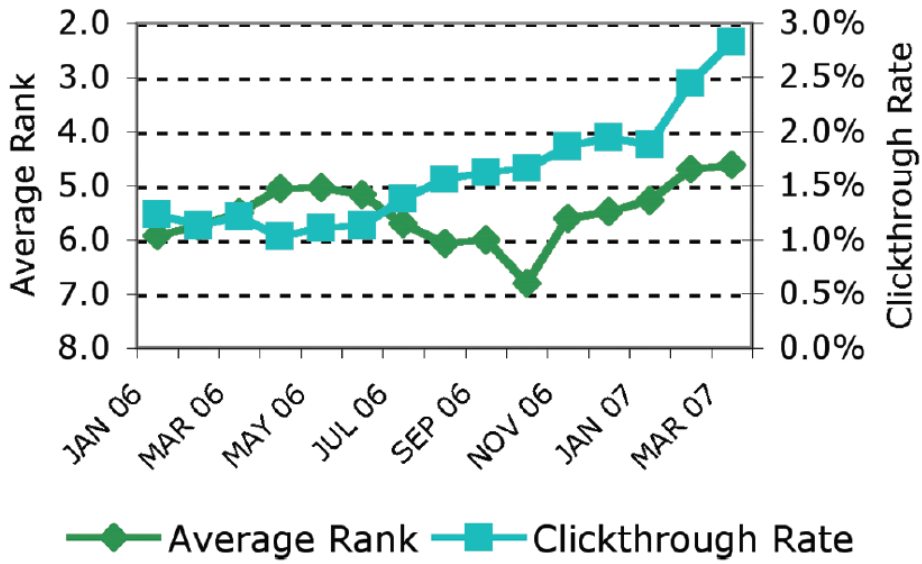
Ne všechny z trendů prvního čtvrtletí jsou body zvyšující ceny na klíčová slova. Nový algoritmus výpočtu ranku (pozice) na Yahoo! platformou Panama je navržen pro určité cenové protekce relevantním klíčovým slovům. Snížení CPC na Yahoo! bylo tedy očekáváno.

Celkově na vyhledávacích kampaních Yahoo! bylo utraceno pouze o 33 % více během prvního čtvrtletí než za stejné období roku 2006 (v porovnání s meziročním růstem o 124 % na Google). V současné době inzerenti na Yahoo! více spoří, s největší pravděpodobností v důsledku růstu cen na Google a MSN. Jak si inzerenti zvykají odstínům nové platformy, pravděpodobně začínají reinvestovat úspory zpět do svých programů Yahoo!.



Graf 10: Útraty inzerentů na Yahoo! [15].

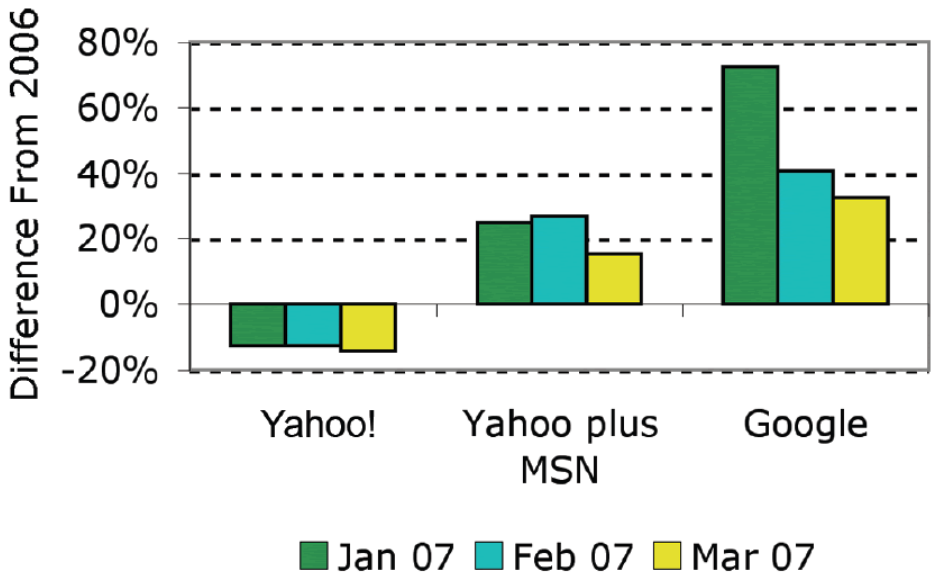
Podle Yahoo! by nový algoritmus pro výpočet pozice měl zlepšit výkon vyhledávače na programech placeného vyhledávání. Průměrný rank (pozice) se od změny algoritmu na Yahoo! zlepšil a výsledný přínos je v podílu kliků na zobrazené inzeráty. Nová platforma tedy prokazuje lepší přístup v zobrazování více relevantních výsledků vyhledávání.



Graf 11: Průměrný rank (pozice) a podíl kliků [15].

Meziroční nárůst v transakcích a prodejů byl většinou pomalý pro inzerenty v Performics 50 na Yahoo! Jedním pravděpodobným faktorem tohoto trendu je, že v srpnu 2006 začal MSN plně poskytovat své vlastní adCenter výpis a ne dále

používat placené vyhledávání na Yahoo! Tím pádem byl snížen počet vyhledávání na Yahoo! Nicméně, celkový nárůst na MSN, Yahoo! a Yahoo! plus je pomalejší než na Google [15].



Graf 12: Meziroční vývoj transakcí (pouze kampaně s přímými příjmy) [15].

8 Faktory ovlivňující algoritmus Google

Při hledání klíčového slova jako výsledek vyhledávání vrátí vyhledávač první adresy WWW stánek, na kterých má hledané slovo velkou váhu. Různé vyhledávače v počítání relevance klíčových velmi liší. Například Google zcela ignoruje klíčová slova ve značce `<meta name=keywords content="slovo" />`. Nebo existují specializovaní roboti, kteří hledají například pouze obrázky nebo počet odkazů. Stále je předmětem mnoha spekulací, jak je váha slov na WWW stránkách počítána. V tabulce č. 2 jsou uvedena nejčastější kritéria pro hodnoceni WWW stránky.

Co vyhledávače sledují	**XHTML značka**	**Důležitost (váha)**
titulek	`<title>slovo</title>`	obrovská
klíčová slova	`<meta name="keywords" content="slovo" />`	značná, v určitých případech však žádná (Google)
popis (description)	`<meta name="description" content="slovo" />`	různá
nadpis 1. úrovně	`<h1>slovo</h1>`	značná

ostatní nadpisy	`<h2>slovo</h2>`
	`<h3>slovo</h3>`
	`<h4>slovo</h4>`
	`<h5>slovo</h5>`
	`<h6>slovo</h6>`
		sporná

začátek stránky	`<body>slovo ..... </body>`	větší než malá
adresa URL	`jméno souboru včetně cesty`	různá
text odkazů s jiných serverů	`<a href="adresa">slovo</a>`	u některých vyhledávačů obrovská (Google, Jyxo)
alty obrázků	`<img src="soubor" alt="slovo" />`	malá
text stránky	`prostý text`	malá
katalogový popisek	`text v katalogu`	různá

Tabulka 2: Nejčastější kritéria pro výpočet váh klíčových slov na WWW stránce [20].

Nedostatečné a subjektivní kvalitativní ohodnocení faktorů ovlivňujících pořadí WWW stránky ve vyhledávači bylo na základě zkušeností 37 předních odporníků na problematiku SEO kvantifikováno. Jednotlivé faktory byly shrnuty do významově blízkých skupin a ohodnoceny dvěma čísly. První číslo, „Důležitost“, vyjadřuje závažnost faktoru (míru jeho významnosti) a je spočítáno jako průměr hodnocení daného faktoru všemi 37 účastníky průzkumu. Druhé číslo, „Neshoda“, vyjadřuje, nakolik se ve svém hodnocení tohoto faktoru jednotliví účastníci průzkumu shodli, a je spočítáno jako velikost standardní odchylky v hodnocení daného faktoru.

Stupnice hodnocení každého faktoru byla stanovena na:

1. bez vlivu
2. minimální vliv
3. částečný vliv
4. silný vliv
5. zásadní vliv

Negativní faktory byly ohodnoceny:

1. bez vlivu na crawling[97] a ranking[98]
2. minimální vliv na crawling a ranking
3. částečný vliv na crawling a ranking
4. silný vliv na crawling a ranking
5. zásadní vliv na crawling a ranking [35]

8.1 Faktory spojené s používáním klíčových slov

Faktory spojené s používáním klíčových slov jsou vztahovány k vlivu přítomnosti vyhledávaného výrazu ve WWW stránce na celkovém hodnocení její relevance vyhledávacím strojem.

Klíčové slovo použité v elementu `title`

Umístění klíčového slova či fráze v elementu title hlavičky stránky má zásadní vliv – 4,9 – na celkovém hodnocení relevance WWW stránky vyhledávacím strojem.

[97] Procházení a stahování stránek robotem. Stránky, které robot (crawler) neprojde a neuloží k pozdějšímu vyhodnocení, nemohou být následně zobrazovány ve výsledcích vyhledávání.
[98] Hodnocení stránek z hlediska relevance vůči vyhledávanému slovu či frázi, míra významnosti webu.

Míra neshody respondentů je 0,3, čili respondenti se výrazně shodují. Element title tak zůstává nejvýznamnějším on-page faktorem pro samotnou optimalizaci po několik uplynulých let a je velmi důležitý pro CRT[99] v SERP[100]. Mimořádný vliv samotného elementu při jeho vhodném použití samotné WWW stránce velmi pomáhá a naopak, nevhodné použití může WWW stránku přímo zničit. Je to tedy nejlepší příležitost, jak proměnit čtenáře SERP v návštěvníka webu provozovatele. Téměř v každém případě bylo zaregistrováno výrazné a rychlé zlepšení pořadí WWW stránky ve vyhledávači právě po umístění klíčových slov do titulku stránky, zejména u rozsáhlých firemních websites s tisíci IBL[101] vedoucími na jednotlivé stránky. Na druhou stranu, jako u všeho, i tento efekt je závislý na použití klíčových slov v textu stránek a v textu odkazů na ně. S největší pravděpodobností také záleží na umístění klíčového slova v titulku, neboť prominentnější (bližší počátku) umístění mělo výraznější efekt, zvláště u webů, které souhlasily s umístěním své značky až za klíčová slova.

Klíčové slovo použité v textu stránky

Umístění klíčového slova v rámci elementu `body`, tedy ve vlastním obsahu stránky má na hodnocení WWW stránky silný vliv – 3,7 a respondenti se shodli průměrně (1,0). Je však důležité používat klíčové fráze v textu stránky tam, kde dávají smysl. Jak roste sofistikovanost vyhledávačů, stává se pro určení významu stránky stále důležitějším nejen umístění hledané fráze na stránce, ale také celkový mix všech použitých slov a frází. Pokud bude na WWW stránce klíčové slovo či fráze použita mnohokrát, může takový přístup výrazně snížit ranking. V zásadě je možné získat vysoký ranking i pro stránky, které neobsahují klíčová slova v textu, ovšem při mnohem vyšším úsilí správce webu a množství odkazů s klíčovým slovem zevnitř i zevně webu.

Vztah obsahu stránky a klíčových slov (tématická analýza)

Tématická analýza (topic analysis) umožňuje posoudit míru vztahu mezi obsahem stránky a hledanými klíčovými slovy. Samotný vztah byl označen za faktor se silným vlivem – 3,4 a shoda respondentů byla průměrná (1,0). Jeho význam je

[99] Click Through Rate – Míra prokliku. Číslo udávající poměr mezi počtem zobrazení odkazu a počtem kliknutí na něj.
[100] Search Engine Results Page – Stránka zobrazující výsledky vyhledávání konkrétního termínu v podobě řady odkazů na různé WWW stránky.
[101] InBound Link – Příchozí odkaz. Odkaz vedoucí z cizího website na website provozovatele. V češtině bývá tento termín překládán jako zpětný.

umocňován ve chvíli, kdy je celý website považován za relevantní. Použití sémanticky souvisejících termínů usnadňuje propojení cílové stránky s dalšími tématicky příbuznými stránkami a pomáhá tak WWW stránkám ke kvalitnímu umístění při vyhledávání málo frekventovaných klíčových slov (long tail keywords). Očekává se, že vztah obsahu stránky a klíčových slov bude v budoucnu ještě důležitejší, podpořen tématy odkazů směřujících na cílovou WWW stránku.

Klíčové slovo použité v elementu `h1`

Vytvoření elementu `h1` (nadpis hlavní úrovně) s klíčovým termínem nebo frází je považován za faktor se silným vlivem (3,4) při průměrně shodě respondentů (1,0). V zásadě záleží na tématu stránky a slovu v elementu. Pokud na WWW stránce o průmyslových pračkách je vložen elementu `<h1>Britney Spears<h1>` ničemu to nepomůže. Naopak, pokud se nadpis první úrovně příliš neshoduje s titulkem stránky a s textem odkazů, může být vážně poškozen ranking WWW stránky. Pokud problémy se šablonami způsobí příliš mnoho duplicit ve velké částí website, může to vést také k omezení crawlingu. Z jiných poznatků vyplývá, že se v současné době již nejedná o tak důležitý faktor, jako v minulosti, nicméně stále zůstává nezbytným pro dobře optimalizované WWW stránky a velmi užitečným při organizaci obsahu stránky.

Klíčové slovo použité v názvu domény

Zahrnutí klíčového slova či fráze do registrovaného doménového jména, například `slovo.cz` či `keyword.com`, je považováno za faktor s částečným vlivem (3,0) s průměrnou shodou (1,0) respondentů. Použití klíčových slov v doménovém jméně je však prospěšné pouze tehdy, pokud je jsou oddělena pomlčkami. Všeobecně se však soudí, že použití příliš velkého počtu pomlček působí problémy s důvěryhodností domény, proto více než jednu či dvě pomlčky nelze doporučit. Dobrá značka je vždy lepší než doménové jméno plné klíčových slov. Pokud však doménové jméno může elegantně vystihovat hledaný výraz, lze mu přikládat mimořádný význam, protože se může jednat o navigační dotaz (navigation query). Navíc přesná shoda doménového jména a hledaného výrazu znamená, že stránky o tomto tématu byly založeny mezi prvními (jsou tedy pro toto téma nejdůležitější), nebo také, že vlastník zaplatil za doménu skutečně vysokou cenu.

Klíčové slovo použité v URL stránky

Zahrnutí klíčového slova či fráze do URL WWW stránky, například `http://www.domena.cz/klicova-fraze/` bývá označováno za faktor s částečným vlivem (2,8) na celkovém hodnocení relevance WWW stránky vyhledávacím strojem. Respondenti se na částečném vlivu shodují průměrně (1,0). Sice klíčové slovo v URL stránky není považováno za tak důležitý faktor jako klíčové slovo v doménovém jménu, pomáhá však zvýšit CTR (a tedy relevanci, pokud se CTR do relevance započítává). V některých případech je také používáno URL jako text odkazu na WWW stránku. Pro samotného uživatele mají klíčové fráze v URL (slova oddělena pomlčkami) rovněž velký význam, neboť jim usnadňuje navigaci.

Klíčové slovo umístěné v elementu `h2`, `h3`, .. `h6`

Umístění cílových termínů v kódu stránky dovnitř elementů nadpisů `h2`, `h3` a dalších mívá částečný vliv (2,8) na celkové hodnocení relevance WWW stránky vyhledávacím strojem s průměrnou shodou (0,9) respondentů. Pokud se text v nadpisech příliš shoduje s titulkem stránky a s textem odkazů, může být poškozen celý rank WWW stranky. Pokud problémy se šablonami způsobí příliš mnoho duplicit ve velké částí website, může to vést také k omezení crawlingu, podobně jako u klíčových slov v nadpise první úrovně. V určitých případech jsou elementy `h2` a další považovány za druhý nejdůležitější on-page[102] faktor. Významnost se z času na čas mění a může se stát, že elementy `h2` a `h3` budou mít stejnou či vyšší důležitost než `h1`. Styly písma a zdůraznění by teoreticky měly být objektivním měřítkem důležitosti slov či frází. Vyhledávací stroje však nejsou „hloupé“, znají CSS a zneužívání těchto elementů může vést k určitým protiopatřením.

Klíčové slovo použité v atributech `alt` a `title` obrázků

Použití cílových klíčových slov v atributech `alt` kódu stránky a v atributech `title` vložených obrázků má na vyhledávací stroje částečný vliv (2,5). Respondenti se na částečném vlivu shodli výrazně (0,8). Pokud je obrázek použit v odkazu místo textu, stává se „textem“ odkazu obsah atributu `alt`. Atributy alt jsou u běžných obrázků velmi důležité pro použitelnost WWW stránky (například zobrazení na mobilních zařízeních, čtečkách obrazovek apod.). Pro SEO je jeho význam

[102] Faktor ovlivňující ranking stránky, který je fyzickou součástí stránky samé. Souvisí s obsahem stránky. Typickým faktorem tohoto druhu je nadpis stránky.

poněkud menší a bývá důležitý pro vyhledávání obrázků[103] nikoli WWW stránek samotných.

Klíčové slovo použité v elementech `b` a `strong`

Umístění klíčového slova či klíčové fráze v textu stránky do elementů pro zvýšení důrazu má jako faktor částečný vliv (2,4) na výsledek vyhledávacích dotazů při průměrné shodě (1,0) respondentů. Vyčlenění slov mimo ostatní text stránky naznačuje, že jsou důležitá. Dává tedy smysl, že tato slova získají o trochu více pozornosti při analýze stránky vyhledávačem. Využitelnost tohoto faktoru může spočívat v použití jednotlivých elementů. Element `b` je pro vyznačování v kódu, element `strong` pro označení důrazu. Obvykle je používán element `strong` se zdůrazněním tří až pěti klíčových frází na každé stránce, v mírně odlišných variantách.

Klíčové slovo použité v elementu `meta` typu `description`

Použití klíčových slov v popisku stránky umístěném v elementu `meta` typu `description` (popisek) v hlavičce stránky je označován za faktor s částečným vlivem (2,1) na činnost vyhledávacích strojů s průměrnou shodou (1,1) respondentů. Vhodný popisek může pomoci ovlivnit uživatele, aby kliknul na odkaz v SERP. Obsah elementu `meta` stále bývá používán jako popisek webu na mnoha menších rozcestnících a vyhledávačích. Je však nutné, aby byl jedinečný pro každou WWW stránku v rámci celého website (podobně jako element `title`) a aby obsahoval cílová klíčová slova. Vyhledávače tak z popisku vstřícně zobrazí výňatek (`snippet`).

Klíčové slovo použité v elementu `meta` typu `keywords`

Použití klíčových slov ve výpisu „klíčových slov“ stránky umístěném v elementu `meta` typu `keywords` (*klíčová slova*) v hlavičce stránky ovlivňuje celkové hodnocení relevance WWW stránky vyhledávacím strojem minimálně (1,2). Shoda respondentů na tomto tvrzení je výrazná (0,5). Klíčová slova v elementu `meta` typu `keywords` se stávají významnějšími pro vyhledávač Yahoo! [35].

8.2 Faktory spojené se stránkou

Faktory spojené se stránkou se vztahují k interpretaci specifických dat spojených s WWW stránkou Googlem a která nejsou závislá na klíčových slovech.

[103] http://images.google.com/

Link popularity v rámci vnitřní struktury website

Link popularity se vztahuje k počtu a důležitosti odkazů, které vedou na cílovou WWW stránku zevnitř website. Na vyhledávací roboty má zásadní vliv (4,1) za výrazné shody respondentů (0,9). Bez odpovídající struktury odkazů však nezískají některé stránky patřičnou váhu. V oblastech extrémního „dlouhého ocasu“ (long tail) a nízké konkurenčnosti v SERP, vnitřní link popularity se stává významným faktorem při hodnocení stránek zanořených hluboko ve struktuře webu (deep content pages). Hodnotu navigačních odkazů určitě zvýší klíčová slova, popř. klíčová fráze, v odkazech v obsahu WWW stránky.

Kvalita a relevance odkazů na externí WWW stránky či websites

Kvalitní a relevantní odkazy na externí stránky nebo celé websites jsou odkazy ve stránce právě na kvalitní, tématicky související WWW stránky a respondenti se na silném vlivu (3,5) faktoru shodli průměrně (1,0). Faktor souvisí s celým konceptem autorit a uzlových webů. V zásadě, co je dobré pro uživatele, by se mělo promítnout do vah pro vyhledávače. Výhodné je také tzv. „tunelování textu odkazu“, tedy pokud WWW stránka A odkazuje na WWW stránku B klíčovým slovem a zároveň tím samým klíčovým slovem odkazuje WWW stránka B na WWW stránku C.

Stáří dokumentu

Stáří WWW stránky má silný vliv (3,4) na celkovém hodnocení její relevance vyhledávacím strojem, neboť starší dokumenty mohou být vnímány jako autoritativnější, zatímco novější mohou být vůči tématu dočasně relevantnější. Respondenti se na tomto poznatku shodli průměrně (1,2). Starší dokumenty tak mohou tedy působit důvěryhodněji, zvláště pokud jsou často citovány a neobsahují příliš mnoho nefunkčních odkazů. V případě blogů a zpravodajských serverů mívají novější stránky obvykle vyšší PageRank, vlivem interní struktury webu. Nové dokumenty mohou být také podpořeny svou aktuálností a aktuální dokumenty na důvěryhodné doméně jsou roboty ze začátku častěji procházeny. Stáří doménového jména je však důležitější než stáří dokumentu. Tudíž stejný dokument na starším důvěryhodném doménovém jméně bude hodnocen lépe než tentýž dokument na novém doménovém jméně. Sice je pravděpodobné, že Google od poslední aktualizace snížil váhu tohoto faktoru, aby omezil parazitní hosting a piggybacking,

ale stáří dokumentu, stáří odkazů na něj vedoucích a stáří domény je stále velmi důležitým faktorem.

Množství indexovatelného textového obsahu

Množství indexovatelného textového obsahu se přímo vztahuje k množství viditelného textu na stránce a je označováno jako faktor se silným vlivem (3,2) na pozici stránky ve vyhledávači s průměrnou shodou (1,1) respondentů. Nad určitou prahovou hodnotou ztrácí tento faktor na významu, ale stránky obsahující výhradně Flash nebo obrázky se málokdy ve vyhledávačích výrazně umístí. Pokud se velikost příliš mnoha stránek blíží nebo rovná velikosti mnoha dokumentů považovaných za spam, může být konkrétní dokument s vysokou pravděpodobností také považován za spam. Google také neustále zpřísňuje svá kritéria pro stránky v primárním indexu požadavkem na určitou minimální hodnotu jejich PageRanku. Delší a autoritativnější dokumenty jsou lepší než články rozsekané na mnoho kratších stránek.

Kvalita obsahu dokumentu (z pohledu algoritmu)

Tento faktor s částečným vlivem (3,0) na pozici stránky ve vyhledávači by mohl mít jistý význam, vzhledem k předpokladu, že vyhledávače mohou používat textové, vizuální a další metody analýzy k určení platnosti (validity) a hodnoty obsahu. Respondenti však v případě faktoru kvality obsahu dokumentu našli jen nízkou shodu (1,3). Obsah dokumentu je však významným faktorem při získávání odkazů, ale je těžké jej změřit a zhodnotit algoritmem. Pro měření kvality obsahu by měly být používány externí faktory.

Organizace toku dokumentu

Uspořádání toku textu v dokumentu, například v publicistických článcích (v USA) se vyžaduje úvod zaměřený na nějaký konkrétní detail, následovaný co nejširším přehledem informací o problematice s postupným zvyšováním detailnosti textu ke konci dokumentu, je považováno za faktor s částečným vlivem (2,8) na celkovém hodnocení relevance vyhledávacím strojem. Respondenti se na částečném vlivu shodli průměrně (1,2). Transparentní organizace nejen napomáhá vyhledávači pochopit, jak dokumenty souvisí navzájem, ale také pomáhá efektivně přenášet PageRank mezi jednotlivými WWW stránkami website (link authority). Struktura website je tedy důležitá, nicméně, každá WWW stránka by měla být schopna obstát před vyhledávacími roboty sama o sobě.

Frekvence aktualizace stránky

Množství změn učiněných v dokumentu a časový interval, v němž k těmto změnám dochází má částečný vliv (2,4) na relevantnost dokumentu u vyhledávacích strojů a respondenti se na tomto částečném vlivu shodli průměrně (1,1). Na Internetu je mnoho starých stránek, které nebyly nikdy aktualizovány a které se dobře umisťují na konkurenční termíny. Z toho vyplývá, že mnohem důležitější je frekvence aktualizace website jako celku, tedy faktor by měl být posuzován ve vztahu ke vzorci obvyklému pro skupinu webů, do níž dokument patří, a k tomu, co četnost aktualizací (či jejich absence) pro tuto skupinu znamená.

Počet lomítek „/“ v URL

Čím vyšší je počet lomítek, tím hlubší je zanoření (nebo též vyšší úroveň zanoření) WWW stránky ve struktuře website. Faktor má při průměrné shodě respondentů (1,0) na relevantnost výsledků indexování vyhledávači minimální vliv (1,9), nicméně, doporučením pro tvůrce WWW stránek a správce serverů je používat maximálně čtyř lomítek.

Gramatická správnost

Gramatická správnost dokumentu je přímo vázána na jazyk dokumentu (angličtina, čeština a další) a při průměrné shodě respondentů (1,0) má minimální vliv (1,8) na umístění WWW stránky ve vyhledávači. Vysvětlením minimálního vlivu faktoru je, že Google začal k nejčastěji chybně zapisovaným výrazům nabízet odkaz na alternativní verze. Tím ztratil tento faktor na významu. Nicméně, zatímco špatná gramatika nemá vliv na ranking stránek, může ovlivnit jejich návštěvníky – negativně! Chybně napsané výrazy (překlepy, špatné hláskování) někdy bývají ve stránkách schválně kvůli uživatelům, kteří chybně zapsaná slova zadávají do vyhledávače – v tom případě ovšem nejde o gramatické chyby, ale o cílové klíčové fráze!

Validita dokumenty (dle standardu W3C)

Validitou dokumentu se rozumí správnost kódu dokumentu či stránky vzhledem k doporučením standardizační organizace W3C, která zastává roli oborové autority v této oblasti. Faktor má minimální vliv (1,4) na umístění dokumentu ve vyhledávači při výrazné shodě respondentů (0,6). Faktor by mohl být důležitější v případě, pokud provozovatel website (tvůrce) dokáže přesvědčit ostatní webdesignéry, že jeho obsah website je lepší, protože jsou WWW stránky validní dle

určité normy. Tím mohou WWW stránkám věnovat pozornost, sledovat jejich informace a odkazovat na ně [35].

8.3 Faktory spojené s website (doménou)

Google rovněž hodnotí WWW stránky vzhledem k jejich umístění v rámci konkrétního website či domény.

Globální link popularity cílového website

Globální link popularity cílového website se vztahuje k počtu a důležitosti odkazů, které vedou na cílový web ze všech ostatních webů na Internetu. Faktor má zásadní vliv (4,5) na umístění dokumentu ve vyhledávači při výrazné shodě (0,9) respondentů. Nejdůležitější je kvalita cílové WWW stránky. Mechanismus výpočtu PageRank sám zajistí přirozené rozložení globální link popularity mezi jednotlivé stránky website.

Stáří webu

Při stáři webu se nejedná ani tak o datum registrace domény, jako spíše o zpřístupnění indexovatelného obsahu čitelného pro vyhledávač (může se měnit se změnami vlastníka domény). Faktor má při výrazné shodě respondentů (1,1) na relevantnost výsledků indexování vyhledávači zásadní vliv (4,0). Faktor měl vždy určitý význam, ale během posledních dvou let se jeho význam ještě zvýšil, přičemž velmi nové weby mohou pomaleji získávat pozici ve vyhledávači v hodně konkurenčních oblastech. Novým webům může pomoci výrazné napojení na „rodiče“ v podobě důvěryhodného webu.

Tématická relevance odkazů směřujících na web

Tématickou relevancí odkazů směřujících na web je vyjádřena tématická příbuznost mezi odkazující stránkou či webem, cílovým webem a klíčovým slovem. Na významném vlivu (3,9) faktoru na pozici ve vyhledávači se respondenti shodli výrazně (1,1). Tématická relevance odkazů je s největší pravděpodobností konečným cílem vyhledávače Google. Ať už relevantní odkazy přinášejí či nepřinášejí zvýšení významu stránky, jsou dobrým obchodem a pomáhají získat vyšší pozice v SERP.

Link popularity webu v rámci skupiny tématicky souvisejících webů

Faktor se zásadním vlivem (3,9) na relevanci celkového hodnocení WWW stránky vyhledávacím strojem se vztahuje k počtu a důležitosti odkazů, které vedou

na cílový web ze všech tématicky příbuzných webů na Internetu. Na zásadním vlivu faktoru se respondenti výrazně shodli (1,1). Vysoce specializovaný website nemusí mít velké množství odkazů, ale několik odkazů od „autorit“ ze sousedství obvykle dostačuje k tomu, aby získal vysokou pozici na klíčová slova ze svého specializovaného oboru.

Četnost přibývání nových odkazů na web

Rychlost a množství, v jakém přibývají nové odkazy na cílový web je považováno za faktor se silným vlivem (3,5) na relevanci celkového hodnocení WWW stránky vyhledávacím strojem. Respondenti se na tomto silném vlivu výrazně shodují (0,9). K velkému vzestupu v pozicích WWW stránek na vyhledávačích u webů s množstvím nových odkazů může dojít, pokud zároveň dojde i k zvětšení objemu obsahu a dalších faktorů. Tento vzestup může být zpomalen, pokud většina odkazů přichází z bezvýznamných webů.

Relevance hlavního tématu webu a hledaného slova či fráze

Míra vztahu mezi obsahem webu a dotazem, který uživatel zadává do vyhledávače je faktorem se silným vlivem (3,2) na pozici WWW stránky ve vyhledávači. Ovšem, s nízkou shodou (1,4) respondentů. Význam faktor se projeví, pokud existují odkazy z cizích webů, které používají text shodný s dotazem zadaným uživatelem do vyhledávače, popř. jeho význam v současné době může být ještě vyšší,neboť se Google snaží zabránit různým google bombám apod.

Výkonnost webu z dlouhodobého hlediska, měřená časem stráveným na stránce, CTR ze SERP, přímými návštěvami, odkazy apod.

Statistiky CTR, času stráveného návštěvníky na stránce či webu, přímých přístupů návštěvníků skrze záložky (*bookmarks*) a další, které může Google měřit prostřednictvím svého toolbaru, WiFi sítí s volným vstupem, služby Google Analytics[104] a pod. Faktory mají částečný vliv (2,7) na relevanci celkového hodnocení WWW stránky vyhledávacím strojem při průměrné shodě (1,2) respondentů a jejich význam stále roste. Jisté obavy jsou poněkud na místě, neboť vyhledávače „vidí“, co uživatel dělá – zvláště v případě uživatelů, kteří používají některé specializované služby vyhledávačů.

[104] http://www.google.com/analytics/

Přímé nastavení autority webu zaměstnanci Google

Google bývá příležitostně podezírán či obviňován, že manipuluje s hodnotami některých webů či domén. Tento faktor se zvláště vztahuje na pozitivní vliv na hodnocení stránek a má na pozici WWW stránky ve výsledcích vyhledávaní částečný vliv (2,6) s nízkou shodou (1,7) respondentů. Obecně se předpokládá, že k umělému navyšování PageRanku WWW stránky dochází velmi ojediněle. Pokud k tomuto zásahu dojde, faktor se stává velmi významným pro ovlivnění pozice WWW stránky ve výsledcích vyhledávání.

TLD domény (`.edu`, `.gov`, `.us`, `.ca`, `.com` a další)

Top Level Domény mají na relevanci celkového hodnocení WWW stránky vyhledávacím strojem částečný vliv (2,5) při průměrné shodě (1,2) respondentů. Předpokládá se, že novější TLD mohou mít vliv pouze na nové weby, tedy znamenají pro ně jistou nevýhodu v podobě stáří TLD, nikoli v názvu TLD. Ovšem weby na TLD „`.gov`“ a „`.edu`“ jsou obvykle velmi významné, vzhledem k množství IBL, které na ně směřuje.

Četnost přidávání nových WWW stránek na website

Četnost výskytu nových WWW stránek na website je označováno za faktor s částečným vlivem (2,5) na relevanci celkového hodnocení website vyhledávacím strojem při průměrné shodě (1,0) respondentů. Velké, zřetelně automatizované přidávání velkého počtu nových stránek může spustit určitou reakci (pozitivní i negativní). Záleží hlavně na kontextu – zpravodajský server může snadno vyprodukovat 1 000 nových stránek denně, zatímco web o domácím mazlíčkovi může mít jednu týdně, ale je to normální pro typ webu. Právě odchylka od normálu zde může být tím podstatným.

Četnost hledání určitého webu v daném časovém intervalu

Četnost hledání doménového jména nebo značky společnosti či organizace, jak je možno zjistit z logů vyhledávacích serverů, bývá považováno za faktor s částečným vlivem (2,0) na pozici WWW stránky ve vyhledávači. Respondenti se na tomto tvrzení shodují průměrně (1,0). Faktor tedy může potenciálně ovlivnit ranking a časté hledání stejného webu může mít časem vliv na SERP.

Verifikace webu prostřednictvím Google Webmaster Central

Služby Google Webmaster Central[105] lze využívat, pokud provozovatel website se zaregistruje v síti Google a prokáže, že mu cílový web skutečně patří. Faktor má však na relevanci celkového hodnocení website vyhledávacím strojem minimální vliv (1,4) při výrazné shodě respondentů (0,8). Faktor tedy příliš nepomůže to ranku stránky, ale může zvýšit rychlost indexace. Požadavkům na znovuzahrnutí do indexu, které přicházejí prostřednictvím Google Webmaster Central, je věnována větší pozornost [35].

8.4 Faktory spojené s vlastnostmi zpětných odkazů

Google přikládá určitou váhu odkazům z cizích webů, které směřují na cílovou WWW stránku.

Text, jímž je tvořen zpětný odkaz

Text, který je uzavřen mezi značky a je považována za faktor se zásadním vlivem (4,4) na výslednou pozici WWW stránky ve výsledcích vyhledávání při výrazné shodě (0,8) respondentů. Text tvořící odkaz je tak tím nejjednodušším ukazatelem toho, o čem si odkazující myslí, že je obsahem WWW stránky či webu, avšak během posledních jednoho či dvou let Google přesunul důraz od textů tvořících odkazy k autoritě domény a počtu citací.

Globální link popularity odkazujícího webu

Globální link popularity odkazujícího webu se vztahuje k počtu a důležitosti odkazů, které vedou na odkazující web ze všech ostatních webů na Internetu. Jedná se o faktor se silným vlivem (3,6) na relevanci celkového hodnocení website vyhledávacím strojem, avšak při nízké shodě (1,3) respondentů. Nicméně platí, že čím lepší je odkazující web, tím lepší je odkaz.

Tématická příbuznost odkazujícího webu

Faktor se silným vlivem (3,5) na výsledné umístění WWW stránky ve výsledcích vyhledávání ukazuje, nakolik WWW stránka, ze kterého přichází odkaz, souvisí s tématem cíle odkazu. Respondenti se na silném vlivu shodují průměrně (0,9). Avšak, tématická příbuznost nemusí nutně udávat autoritativnost.

[105] http://www.google.com/webmasters/

Link popularity cílového webu v rámci skupiny tématicky souvisejících webů

Faktor se vztahuje k počtu a důležitosti odkazů, které vedou na cílový web ze všech tématicky příbuzných webů na Internetu a je považován za faktor se silným vlivem (3,5) na relevanci celkového hodnocení website vyhledávacím strojem při průměrné shodě (1,2) respondentů. Faktor však může být velmi významný pro weby působící jako komunitní centra, neboť pokud website nemá žádné odkazy ze své komunity, je jeho dobré umístění ve vyhledávání mnohem obtížnější.

Stáří odkazu

Doba, po kterou odkaz na cílový web existuje je faktor se silným vlivem (3,1) na pozici WWW stránky ve vyhledávači při průměrné shodě respondentů (1,0). V určitých případech může být obtížné odlišit vliv stáři odkazu od vlivu stáři webu. V budoucnu se však stáří odkazu může stát základem vyhledávacích algoritmů.

Text obklopující odkaz

Test stránky před a za odkazem má silný vliv (3,1) na relevanci celkového hodnocení website vyhledávacím strojem při výrazné shodě (0,8) respondentů. Podle nejnovějších studií se tento faktor stává stále více a více důležitým. Důkazem jsou výsledky vyhledávání Google Images, testy ukazují, že významnost obrázků je hodnocena stejně tak podle okolního textu, jako podle jakékoli vlastnosti obrázku samotného.

Interní link popularity odkazující stránky v rámci webu či domény

Interní link popularity odkazující stránky v rámci webu či domény se vztahuje k počtu a důležitosti odkazů, které vedou na stránku, z níž přichází odkaz, zevnitř webu. Faktor má umístění WWW stránky ve vyhledávači částečný vliv (2,9) při průměrné shodě (1,2) respondentů. Správně zvolený text vnitřního odkazu tak může být velmi užitečný, jakmile začne být web považován za důvěryhodný. Pokud se však provozovatel website nebo správce serveru rozhodně interní linky popularity odkazujících WWW stránek v rámci webu či domény zneužívat a website bude mít příliš málo IBL z jiných domén, může být website vyhledávači persekuován.

TLD odkazující domény (`.edu`, `.gov`, `.com`, `.ca`, `.co.uk` a další)

V tomto případě má Top Level Domain na relevanci celkového hodnocení website vyhledávacím strojem částečný vliv (2,5) za průměrné shody (1,2) respondentů. Všeobecně je známo, že odkazy z domén `.edu` a `.gov` jsou cenné, protože obsahují hodně kvalitních odkazů z významných webů.

PageRank (podle Google Toolbaru) odkazující stránky

Vliv hodnoty PageRanku je považováno za faktor s částečným vlivem (2,5) na pozici WWW stránky ve vyhledávači Google. Respondenti se na tomto tvrzení shodnou průměrně (0,9). Google Toolbar se chronicky zpožďuje, ale Google používá hodnot PageRanku pro nastavení priorit procházení stránek (crawling) a při rozhodování, zda dokument zařadit do primárního nebo doplňkového indexu (supplemental index). Ukazatel PageRanku podle Google Toolbaru není ve své podstatě špatný, nicméně provozovatelé website a správci serverů by se tímto ukazatelem neměli nechat příliš ovlivnit.

Vlastnosti odkazu spojené s časem (kdy byl odkaz vytvořen či změněn)

Datum vzniku odkazu, jeho změny apod. má částečný vliv (2,5) na relevanci celkového hodnocení website vyhledávacím strojem s průměrnou shodou (1,0) respondentů. Tento faktor má tedy svůj vliv v závislosti na tom, kdy byl odkaz vytvořen a jak dlouho existuje [35].

8.5 Faktory s negativním vlivem na crawling nebo ranking

Faktory, které mohou mít negativní dopad na schopnost robota vyhledávače procházet stránky nebo na ranking stránek v Google jsou čím dál tím významnější.

Server je často robotům nepřístupný

Dostupnost serveru, na kterém jsou stránky umístěny, je považováno za faktor se silným vlivem (3,8) na umístění WWW stránky ve vyhledávači, ovšem za nízké shody (1,3) respondentů. Pokud však obsah website nemůže být indexován, získávají ostatní (konkurenční) weby výhodu, protože jsou indexovány dříve. A navíc, pokud je web často mimo provoz nebo výpadek je delší 48 hodin, velice rychle zmizí z indexu vyhledávače.

Obsah velmi podobný nebo duplicitní k již zaindexovanému obsahu

Faktor se vztahuje k samotnému chování vyhledávače a jeho vliv na umístění WWW stránky ve vyhledávači při průměrné shodě (1,2) respondentů je silný (3,6). Jelikož vyhledávačům stačí jedna verze konkrétního obsahu, je velmi pravděpodobné, že duplicitní obsah zahazují. V dokonalém světě by index obsahoval mnohem méně duplicit a zřejmě to je cíl, ke kterému Google směřuje. Duplicitní obsah v zásadě nemusí být nějak penalizován, je ale zbytečný. Duplicitní obsah však může snížit viditelnost nebo nepřímo poškodit ranking dělením zpětných odkazů, v závislosti na povaze oné „duplicity“.

Odkazování na nekvalitní a spamovací weby

Vliv odkazů, které vedou z provozovatelova webu na jiný, méně kvalitní web, má na hodnocení provozovatelova webu silný vliv (3,5) při průměrné shodě (1,2) respondentů. Odkazováním na méně kvalitní weby v sousedství provozovatel website sám sebe označuje za součást tohoto sousedství, nicméně, podstatný je samotný poměr kvalitních a nekvalitních odkazů.

Použití duplicitních elementů `title` a `meta` na velkém množství stránek

Faktor se vztahuje ke zdrojovému kódu stránek u webů, které na větším množství svých stránek používají stejný text jako obsah elementů `title` a `meta` a je považován za faktor se silným vlivem (3,3) na relevanci celkového hodnocení website vyhledávacím strojem. Silný vliv tohoto faktoru však vyplynul z nízké shody respondentů (1,3). Filtry duplicitního obsahu jsou na vyhledávačích stále. Pokud website neposkytuje mnoho obsahu a navíc ještě obsahuje řadu duplicit, nejen že to snižuje ranking, ale také to může způsobit vypadnutí mnoha WWW stránek do doplňkového indexu (supplemental index). Duplicitní titulky mnoha WWW stránek tak vážně snižují schopnost WWW stránek získat dobrou pozici, zvláště u WWW stránek z oblasti dlouhého ocasu (long tail), které nemají mnoho zpětných odkazů. Tyto stránky vzájemně soupeří v oblasti vnitřní odkazovanosti a relevance elementu `title`, takže je důležité mít unikátní titulek každé stránky.

Účast v odkazovacích službách nebo prodej odkazů

Vliv odkazových bloků v patičkách stránek apod. silně ovlivňuje (3,3) pozici WWW stránky ve výsledcích vyhledávání. Nízká shoda (1,4) respondentů však

vypovídá o rozdílných názorech SEO expertů. Google tvrdí, že sleduje placené odkazy prostřednictvím pravděpodobnostních modelů a sofistikovaných analyzačních algoritmů. Některé weby jsou snadno identifikovatelné, takže jim lze zabránit v předávání PageRanku. Nicméně stále zbývá řada webů, na nichž se může vyplatit přítomnost placeného odkazu.

Nadužívání cílových klíčových slov (tapetování či spamování)

Četnost výskytu často vyhledávaných klíčových slov v textu nebo kódu stránky je považováno za faktor se silným vlivem (3,3) na relevanci celkového hodnocení website vyhledávacím strojem při průměrné shodě (1,0) respondentů. Při překročení určité míry četnosti klíčových slov na WWW stránce lze očekávat různé negativní vlivy na hodnocení samotné WWW stránky vyhledávačem.

Velmi pomalá reakce serveru

Jedná se o čas, který uplyne mezi odesláním požadavku na server a přijetím odpovědi vyhledávacím robotem. Faktor má částečná vliv (2,8) na pozici WWW stránky ve výsledcích vyhledávání s průměrnou shodou (1,0) respondentů. Faktor tedy zpomaluje crawling, avšak neovlivňuje ranking.

Zpětné odkazy ze spamovacích webů

Odkazy, které vedou na website z webů, které vyhledávače považují za nekvalitní mají částečný vliv (2,2) na relevanci celkového hodnocení website vyhledávacím strojem při průměrné shodě (1,0) respondentů. Zpětné odkazy ze spamovacích webů mohou ohrozit nově vzniklé weby. Starší weby s mnoha kvalitními zpětnými odkazy vydrží bez úhony mnohem více spam odkazů. Záleží spíše na poměru mezi kvalitními a nekvalitními odkazy, než na přesném počtu nekvalitních odkazů.

Malé počty návštěvníků webu (měřeno Toolbarem, počtem kliků v SERP a podobně)

Návštěvnost website má na relevanci celkového hodnocení website vyhledávacím strojem částečný vliv (2,1) a respondenti se na tomto částečném vlivu shodují průměrně (1,0). Frekvence používání dat je však jednou ze známek kvality. Pokud je nízká, musí ji website dohnat něčím jiným [35].

9 Metodický postup optimalizace dokumentu pro WWW vyhledávače

Navrhovaný metodický postup úspěšné optimalizace dokumentu (MPOD) částečně vychází z již publikovaných kroků a postupů SEO činností a částečně významnosti jednotlivých faktorů ovlivňujících pozici dokumentu ve výsledcích vyhledávání. Po několikaletých praktických experimentech a prostudování dostupných zdrojů na uvedené téma lze metodický postup optimalizace dokumentu shrnout do následujících činností:

1. Analýza klíčových slov
2. Analýza konkurence
3. Analýza stávajícího stavu (pokud existuje)
4. MPOD copywriting
5. Optimalizace dokumentu a serveru
6. Registrace do katalogů a tvorba zpětných odkazů
7. Analýza efektivity

Každá z těchto činností má své postupy a metody, které jsou popsány v dalších částech textu. Nejdříve je však nutné definovat primární klíčová slova z příslušného oboru a regionu, tzv. „startovací můstky“ celého procesu optimalizace dokumentu. Provozovatel website by si měl nejdříve pečlivě zvážit cílovou skupinu uživatelů, tedy potenciálních zákazníků a od této skupiny zvolit 1 – 10 primárních klíčových slov.

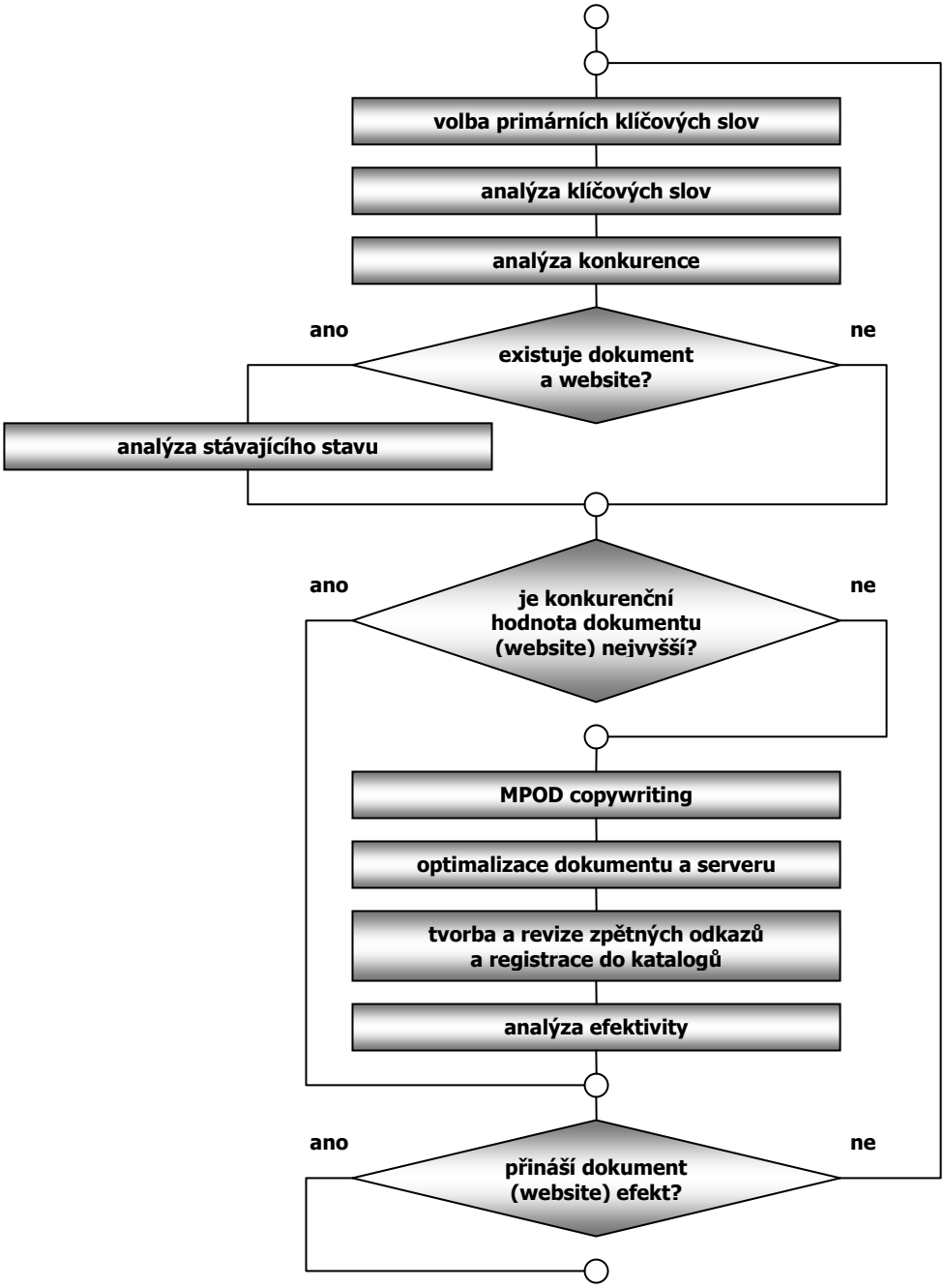


Schéma 1: Metodický postup optimalizace dokumentu pro WWW vyhledávače.

9.1 Analýza klíčových slov

Analýza klíčových slov je pro metodický postup optimalizace dokumentu zásadní, neboť správně vybraná klíčová slova významně ovlivňují úspěch všech kroků MPOD. Nejdříve je nutné získat co nejvíce klíčových slov, na které lze dokument, popř. celý website optimalizovat. Pro získání šířky klíčových slov lze použít celou řadu SW nástrojů, například Google AdWords, eTarget, Wordtracker či Overture

Search Term Suggestion Tool. Výběr příslušného SW nástroje není důležitý, důležitá je samotná práce se SW, rozdělení klíčových slov a výběr klíčových slov.

V každém SW nástroji pro získání synonym klíčových slov je nutné postupně zadat primární klíčová slova, která byla pečlivě stanovena samotným uživatelem nebo MPOD expertem. SW nástroj Google AdWords po zadaní klíčového slova nebo fráze a výběru jazyku a země či teritoria, vrátí celou řadu synonym s cennými daty o objemu vyhledávání, konkurenci inzerentů, průměrném objemu vyhledávání, trendech vyhledávání popř. nejvyšším objemu vyhledávání. Veškerá nabídnutá synonyma příslušného klíčového slova jsou řazena dle relevance a v některých případech jsou nabídnuta i tzv. další klíčová slova.

Samotný výběr šíře klíčových slov je závislý na provozovateli website nebo serveru či na MPOD expertovi. Konkrétní postupy výběru všech vhodných klíčových slov mohou být různé a většinou a většinou jsou tajeny pracovníky zabývajícími se v jakékoli podobě optimalizací WWW stránek, tudíž nelze objektivně zhodnotit, zda jsou vhodné či nevhodné, objektivní či neobjektivní apod.

Prvním krokem analýzy klíčových slov v MPOD je prostý a rychlý výpočet relevance jednotlivých nabídnutých klíčových slov, která je vypočítána z celkového počtu nabídnutých klíčových slov a frází a samotného pořadí konkrétního klíčového slova a fráze. Pokud je například celkový počet nabídnutých klíčových slov 150, první, nejrelevantnější klíčové slovo či fráze, má hodnotu 1 (150/150 – v naprosté většině případů se jedná právě o jedno z deseti primárních klíčových slov) a poslední, nejméně relevantní klíčové slovo či fráze, má hodnotu 1/150.

Druhým krokem je nutné rozdělit klíčová slova určitou bodovací metodou na „konkrétnější klíčová slova“ a „další klíčová slova“ (pokud byla aplikací Google AdWords nabídnuta), například „konkrétnější klíčová slova“ jsou ohodnoceny dvěma body a „další klíčová slova“ pouze bodem jedním.

Po stanovení relevančnosti jednotlivých klíčových slov a frází a jejich rozdělení dle konkréčnosti následuje trvale vyřazení naprosto nevhodných a nesmyslných klíčových slov a frází samotnou redukcí počtu provozovatelem website ve spolupráci se MPOD expertem. Odstraněná klíčová slova nebo fráze se totiž absolutně odchylují od původního zadaného klíčového slova (např. fráze sice obsahuje primární klíčové slovo, ovšem sama o sobě se vztahuje k jinému tématu)

nebo obsahují názvy domén prvního řádu, tzv. Top Level Domain či slova, která jsou fulltextovými vyhledávači ignorována, například spojky, číslovky, předložky apod.

Očištěná, relevantní, klíčová slova je nezbytné posoudit a přiradit jim prioritu, tzn. provozovatel website nebo MPOD expert z veškerých klíčových slov a frází sestaví pořadí, ze kterého budou jednotlivá klíčová slova a fráze použita při samotném procesu metodického postupu optimalizace dokumentu. Primárním úkolem je shrnout klíčová slova a fráze do tabulky a sestavit z nich matici pro aplikování metod vícekriteriální analýzy variant. V prvním sloupci tabulky jsou tedy jednotlivá klíčová slova a v prvním řádku jednotlivá kritéria. Z každého aplikačního SW pro analýzu klíčových slov vzejdou poněkud jiná klíčová slova (synonyma) a jiná kritéria. Postup při analýze a výběru klíčových slov a frází (stanovení jejich pořadí pro MPOD) je však stejný nebo velmi podobný. Z aplikace Google AdWords jsou v poslední době pro analýzu klíčových slov použitelná kritéria v podobě konkurence inzerentů, průměrného objemu vyhledávání, relevance a konkrétních slov.

Klíčové slovo	**Konkurence inzerentů k1**	**Průměrný objem vyhledávání k2**	**Relevance k3**	**Konkrétní slova k4**
slovo 1	hodnota 1,1	hodnota 1,2	hodnota 1,3	hodnota 1,4
slovo 2	hodnota 2,1	hodnota 2,2	hodnota 2,3	hodnota 2,4
slovo 3	hodnota 3,1	hodnota 3,2	hodnota 3,3	hodnota 3,4
slovo 4	hodnota 4,1	hodnota 4,2	hodnota 4,3	hodnota 4,4
slovo 5	hodnota 5,1	hodnota 5,2	hodnota 5,3	hodnota 5,4

Tabulka 3: Ukázková tabulka pro výpočet pořadí klíčových slov a frází.

Pro stanovení vah z kardinální informace o preferencích jednotlivých kritérií lze využít několika metod, přičemž se předpokládá, že provozovatel website nebo MPOD expert je schopen a ochoten určit nejen pořadí důležitosti kritérií, ale také poměr důležitosti mezi dvojicemi kritérií. Nejpoužívanějšími metodami této oblasti jsou metoda bodovací, která transformuje bodové hodnocení důležitosti kritérií do podoby váhového vektoru, a Saatyho metoda kvantitativního párového porovnávání, která odvozuje váhový vektor z informace o odhadu poměru vah.

Důležitost každé z variant dle bodovací metody je vyjádřena určitým počtem bodů v rámci určené bodové stupnice. Bodovací metoda je vhodná v případě, že kritéria hodnotí více expertů, ne pouze jeden člověk. Každý člověk ohodnotí kritérium určitým počtem bodů, čím je kritérium důležitější, tím více bodů dostane. Stupnice pro bodování může být vyjádřena i graficky pomocí úsečky. Na ní jsou pak zakresleny pozice jednotlivých kritérií vzhledem ke koncům úsečky, které vyjadřují

nejnižší a nejvyšší preferenci. Hodnoty váhového vektoru se pak normalizují podle vztahu:

$$v_j = \frac{b_j}{\sum_{j=1}^{n} b_j}, \text{ j = 1, 2, ..., n,}$$

kde b_j je součet všech bodů od jednotlivých MPOD expertů, které *j-tému* kritériu tito MPOD experti přidělili. Je ovšem otázkou, zda je vždy vhodné pevně stanovit rozsah stupnice již na začátku hodnocení. Tento postup je vhodný, pokud existuje hned na začátku poměrně jasná představa o tom, jak jsou ta která kritéria pro hodnocení variant důležitá.

Saatyho metoda slouží k určení vah kritérií, hodnotí-li je pouze jeden MPOD expert. Jde o metodu kvantitativního párového porovnávání kritérií. Pro hodnocení párových porovnání kritérií je používána 9-ti bodová stupnice, přičemž je možné využívat i mezistupně (hodnoty 2, 4, 6, 8):

1 – rovnocenná kritéria i a j

2 – slabě preferované kritérium i před j

5 – silně preferované kritérium i před j

7 – velmi silně preferované kritérium i před j

9 – absolutně preferované kritérium i před j

MPOD expert tedy porovná každou dvojici kritérií a velikosti preferencí *i-tého* kritéria vzhledem k *j-tému* kritériu zapíše do Saatyho matice $S = (s_{ij})$:

$$S = \begin{matrix} 1 & s_{12} & .. & s_{1n} \\ 1/s_{12} & 1 & ... & s_{2n} \\ ... & ... & ... & ... \\ 1/s_{1k} & 1/s_{12} & & 1 \end{matrix}$$

Matice je čtvercová řádu nxn, reciproční, tj. platí, že $s_{ij}=1/S_{ji}$, a vyjadřuje vlastně odhad podílů vah *i-tého* kritéria a *j-tého* kritéria. Dle Saatyho existuje několik jednoduchých způsobů, pomocí kterých lze odhadnout váhy v_j. Nejčastěji se používá postup výpočtu vah jako normalizovaného geometrického průměru řádků Saatyho matice (metoda logaritmických nejmenších čtverců).

$$b_i = \sqrt[n]{\prod_{j=1}^{n} s_{ij}}$$

Váhy jsou pak vypočítány normalizací hodnot b_i

$$v_i = \frac{b_i}{\sum_{i=1}^{n} b_i}$$

Případy, kdy je Saatyho matice nekonzistentní, jsou velmi časté zvláště u rozsáhlejších úloh. Nekonzistence může být způsobena chybou při zadávání odhadů poměrů vah, když MPOD expert neprováděl žádnou kontrolu svých odhadů. V tomto případě je nutné na základě odhadu vah překvantifikovat Saatyho matici tak, aby splňovala požadavek konzistence, a poté provést nový odhad vah. Tímto interaktivním způsobem lze dospět k velmi solidním výsledkům [4].

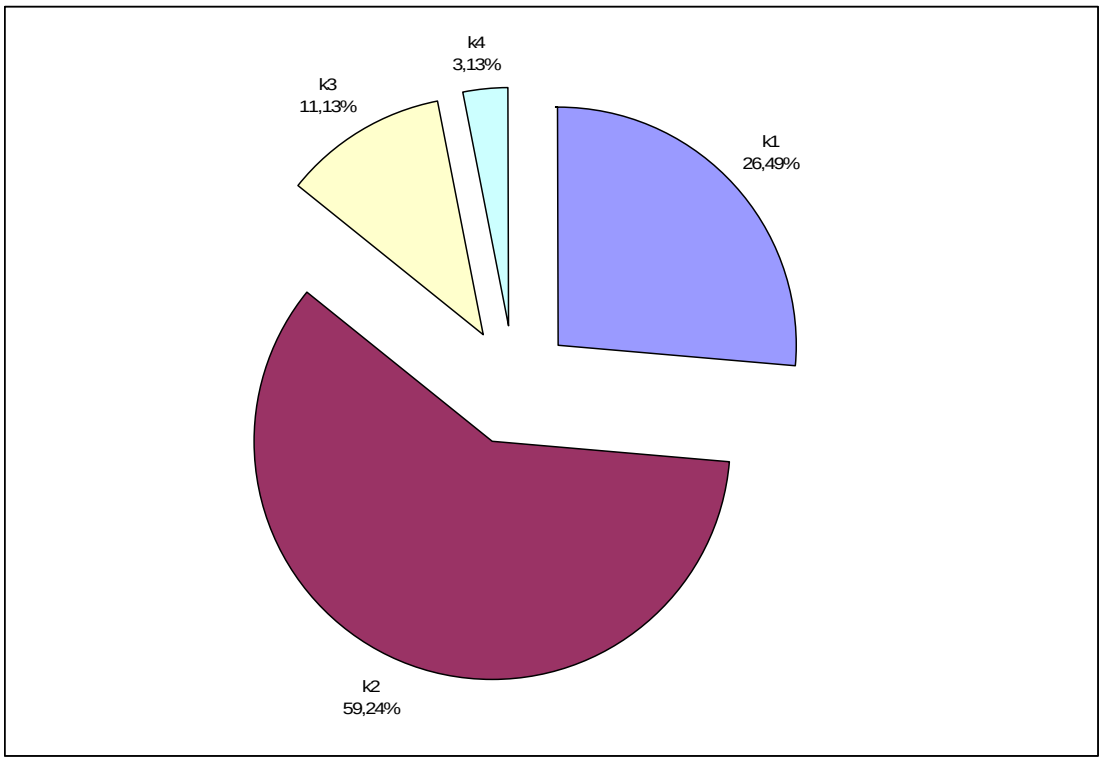
Saatyho metodou kvantitativního párového porovnání lze velmi snadno a rychle stanovit jednotlivé váhy kritérií pro sestavenou tabulku klíčových slov následovně:

k1 – konkurence inzerentů : 0,2650

k2 – průměrný objem vyhledávání : 0,5924

k3 – Relevance : 0,1113

k4 – Konkrétní slova : 0,0313



Graf 13: Grafické znázornění vah kritérií dle Saatyho metody.

Po stanovení vah kritérií a provedení kontroly, zda jsou všechna kritéria maximalizační, je vhodné, když MPOD expert využije pro hodnocení jednotlivých klíčových slov některé metody vyžadující kardinální informaci. Metod, které vyžadují zadání kardinální informace o kritériích v podobě vah a o variantách v podobě kriteriální matice s kardinálními hodnotami, je celá řada, např. metoda váženého součtu nebo metoda TOPSIS.

Metoda váženého součtu je speciálním případem metody funkce užitku a vychází z principu maximalizace užitku. Dosáhne-li varianta a_i podle kritéria j určité hodnoty y_{ij}, přináší tak užitek, který lze vyjádřit pomocí lineární funkce užitku. V prvním kroku výpočtu je určena ideální varianta H s ohodnocením $(h_1, ..., h_n)$ a bazální varianta D s ohodnocením $(d_1, ..., d_n)$. Následně je sestrojena standardizovaná kriteriální matice R, jejíž prvky jsou získány pomocí vzorce:

$$r_{ij} = \frac{y_{ij} - d_j}{h_j - d_j}$$

Matice R již představuje matici hodnot funkce užitku z *i-té* varianty podle *j-tého* kritéria, protože prvky této matice jsou lineárně transformovanými kriteriálními hodnotami tak, že $r_{ij} = \in \langle 0;1 \rangle$. Potom bazální variantě odpovídá hodnota nula

a ideální variantě hodnota jedna. Pro jednotlivé varianty pak lze vypočítat agregovanou funkci užitku:

$$u(a_i) = \sum_{j=1}^{n} v_j r_{ij}$$

Na základě výpočtu agregované funkce užitku pro jednotlivá klíčová slova lze sestavit pořadí podle hodnot $u(a_i)$.

Metoda TOPSIS posuzuje varianty (klíčová slova) z hlediska jejich vzdálenosti od ideální a bazální varianty. Primárně musí MPOD expert zkonstruovat normalizovanou kriteriální matici $R = (r_{ij})$ podle vzorce:

$$r_{ij} = \frac{y_{ij}}{\sqrt{\sum_{i=1}^{p} y_{ij}^{\ 2}}}$$

Sloupce matice R jsou vektory jednotkové délky. Následně dle vztahu $w_{ij} = v_j r_{ij}$ je vypočítána normalizovaná vážená kriteriální matice $W = (w_{ij})$ a vzhledem k hodnotám matice W je určena ideální varianta h s hodnotami $(h_1, ..., h_n)$ a bazální varianta d s ohodnocením $(d_1, ..., d_n)$. Dalšími nezbytnými kroky jsou výpočty vzdálenosti jednotlivých variant od ideální varianty

$$d_i^{\ +} = \sqrt{\sum_{j=1}^{k} (w_{ij} - h_j)^2}$$

a od bazální varianty

$$d_i^{\ -} = \sqrt{\sum_{j=1}^{k} (w_{ij} - d_j)^2} \ .$$

Předposledním krokem, před samotným sestupným seřazením variant (klíčových slov a frází), je výpočet relativních ukazatelů vzdáleností jednotlivých variant od bazální varianty podle vzorce

$$c_i = \frac{d_i^{\ -}}{d_i^{\ +} + d_i^{\ -}} \ .$$

Hodnoty těchto ukazatelů se pohybují mezi 0 a 1, přičemž hodnotu 0 nabývá bazální a hodnotu 1 ideální varianta [4]. Výsledkem metody TOPSIS a metody váženého součtu je tedy pořadí jednotlivých klíčových slov a frází, které je základem

pro pokračování metodického postupu optimalizace dokumentu. Pořadí z obou metod jsou velmi podobná a na nejlepších několika místech a na nejhorších několika místech ve většině případů naprosto shodná. Shodné výsledky mohou sloužit k ověření správnosti jednotlivých metod. S klíčovými slovy, jejiž hodnota ukazatele je pod 0,5 není přínosné pracovat v následujících krocích.

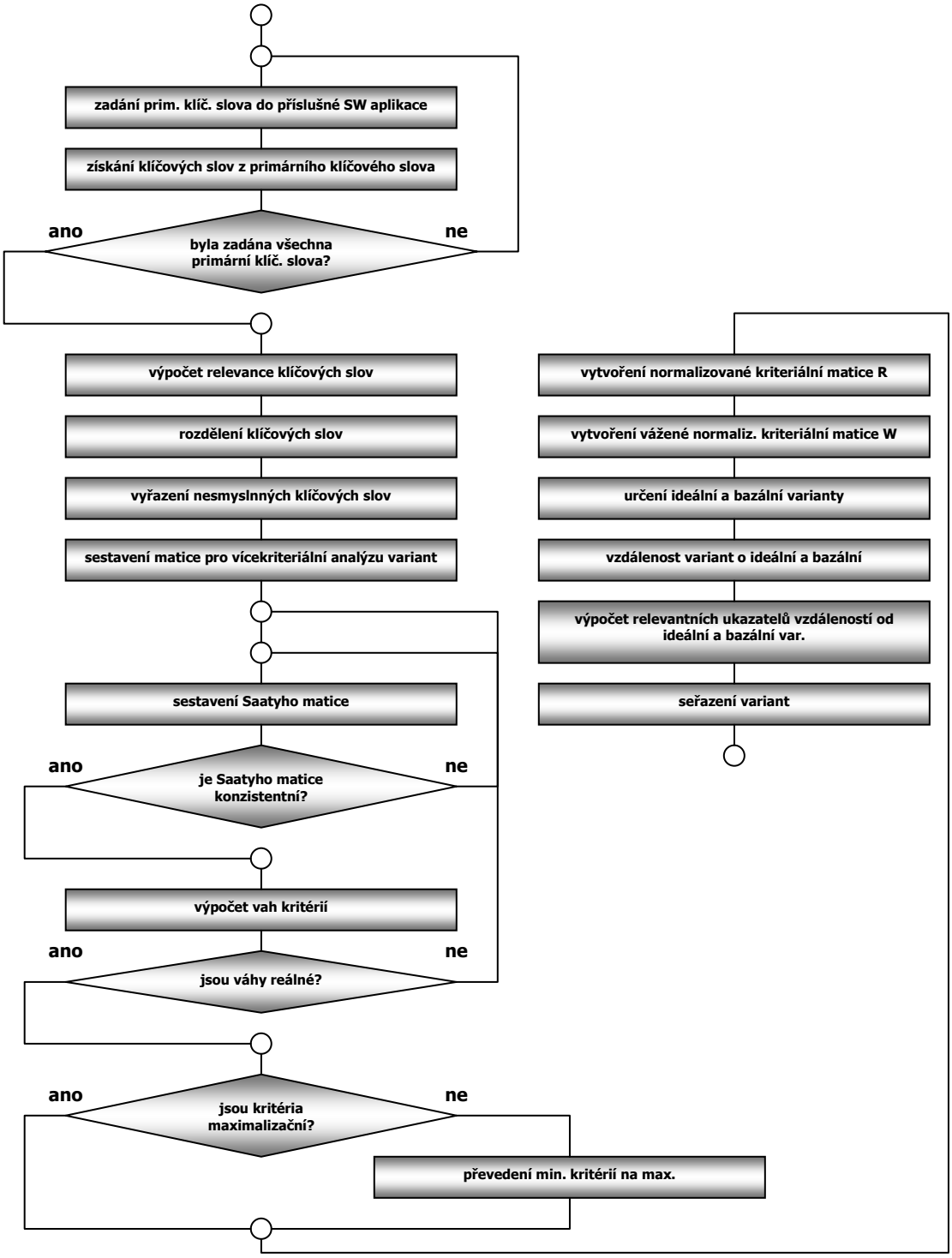


Schéma 2: Metodický postup pro analýzu klíčových slov.

9.2 Analýza konkurence

Druhým krokem metodického postupu optimalizace dokumentu po analýze klíčových slov je podrobná analýza konkurenčních dokumentů (websites) popř. nekomerčních dokumentů (websites) zaměřených na podobné téma. Základem pro analýzu konkurence jsou výsledná klíčová slova a fráze z analýzy klíčových slov. V zásadě je každé klíčové slovo či fráze vložena do vyhledávačů, např. Google a analyzovány jsou následně samotné výsledky vyhledávání. Výsledkem analýzy konkurence je soubor tabulek (za každé klíčové slovo či frázi je několik tabulek), které mají důležitou vypovídací hodnotu.

Provozovatel website nebo MPOD expert prozkoumá každý odkaz z minimálně prvních 3 stránek (cca 30 odkazů) z výsledku vyhledávání klíčového slova nebo fráze. Z důvodu objektivity vyhledávání je nutné, aby byly před zadáním klíčových slov ve vyhledávači vymazány dočasné soubory cookies, které se vztahují ke konkrétnímu WWW prohlížeči a používanému vyhledávači. Na konkurenčních dokumentech nabídnutými vyhledávačem jsou zkoumány a zapisovány do tabulky následující informace:

- Pozice dokumentu (o) – nabývá hodnoty 1 – 30.
- Google Toolbar PageRank stránky (GTPR) – nabývá hodnoty 0 – 10.
- `Title` (title) – nabývá hodnoty 0 nebo 1 a informuje o zastoupení klíčového slova v titulku dokumentu.
- `h1` (h1) – nabývá hodnoty 0 nebo 1 a informuje o zastoupení klíčového slova v nadpisu 1. úrovně v dokumentu. V ojedinělých případech může tento ukazatel nabývat i vyšších hodnot než 1. V takovém případě informuje o nekorektním používání vyššího počtu nadpisů nejvyšší úrovně na jednom dokumentu.
- `h2` (h2) – nabývá hodnot celých kladných čísel a informuje o počtu zastoupení klíčového slova v nadpisech 2. úrovně v dokumentu.
- `h3` (h3) – nabývá hodnot celých kladných čísel a informuje o počtu zastoupení klíčového slova v nadpisech 3. úrovně v dokumentu.
- `h4` (h4) – nabývá hodnot celých kladných čísel a informuje o počtu zastoupení klíčového slova v nadpisech 4. úrovně v dokumentu.

- `h5` (h5) – nabývá hodnot celých kladných čísel a informuje o počtu zastoupení klíčového slova v nadpisech 5. úrovně v dokumentu.
- `h6` (h6) – nabývá hodnot celých kladných čísel a informuje o počtu zastoupení klíčového slova v nadpisech 6. úrovně v dokumentu.
- `Description` (des) – nabývá hodnot 0 nebo 1 a informuje o zastoupení klíčového slova v popisku dokumentu.
- `KeyWords` (kw) – nabývá hodnot 0 nebo 1 a informuje o zastoupení klíčového slova v seznamu klíčových slov dokumentu.
- Stáří indexu (i.a.) – nabývá hodnot přirozených čísel a informuje o stáří indexace dokumentu v hodinách.
- Velikost souboru (B) – nabývá hodnot přirozených čísel a informuje o velikosti souboru v B[106].
- Množství slov (w) – nabývá hodnot přirozených čísel a informuje o celkovém počtu slov v těle dokumentu (pokud se jedná ho XHTML dokument, tak v těle dokumentu).
- Hustota klíčového slova (d.kw.) – nabývá hodnot reálných čísel a informuje o podílu zastoupení klíčového slova či fráze na celkovém počtu slov v dokumentu (počet klíčových slov/celkový počet slov).
- Doména (dom) – nabývá hodnot 0 nebo 1 a informuje o zastoupení klíčového slova v názvu doménového jména.
- Stáří domény (d.a.) – nabývá hodnot přirozených čísel a informuje o stáří doménového jména v měsících (v některých případech může být ukazatel zkreslující, zejména, pokud na doménovém jménu byl provozován obsah s několikaměsíčním, popř. několikaletým, zpožděním).
- URL (URL) – nabývá hodnot 0 nebo 1 a informuje o zastoupení klíčového slova v URL dokumentu (nikoli v doménovém jménu).
- Počet lomítek v URL (/) – nabývá hodnot přirozeného čísla a informuje o počtu lomítek v URL dokumentu. Ukazatel ztrácí důležitost v případě, že ukazatel URL nabývá nulové hodnoty.

[106] Byte

- `alt` a `title` (a.t.) – nabývá hodnot přirozeného čísla a informuje o počtu zastoupení klíčového slova v alternativních popiscích a titulcích obrázků umístěných v dokumentu (XHTML dokument).
- `b` a `strong` (b.s.) – nabývá hodnot přirozeného čísla a informuje o počtu zastoupení klíčového slova ve zvýrazněných částech textu dokumentu (XHTML dokument).
- validita (val) – nabývá hodnot přirozeného čísla a informuje o dodržení příslušné normy XHTML jazyku. Pokud ukazatel nabývá hodnoty 0, WWW stránka je v souladu s příslušnou normou. V opačném případě nenulová hodnota koresponduje s počtem chyb a prohřešku proti normě konsorcia W3C (XHTML dokument).
- Zpětné odkazy >= 4 (l.p.4) – nabývá hodnot přirozeného čísla a informuje o počtu zpětných odkazů s Google Toolbar PageRankem vyšším než 3.
- Zpětné odkazy <= 3 (l.p.3) – nabývá hodnot přirozeného čísla a informuje o počtu zpětných odkazů s Google Toolbar PageRankem nižším nebo rovným 3.
- Podobné stránky (s.p.) – nabývá hodnot přirozených čísel a informuje o celkovém počtu podobných dokumentů dle algoritmu Google.
- Vnitřní odkazy (l.p.w.) – nabývá hodnot přirozených čísel a informuje o celkovém počtu odkazů z dokumentu z interního website.
- Kvalita obsahu (q.c.) – nabývá hodnot 0 – 10 a jedná se o subjektivní hodnocení kvality obsahu.
- Odkazy (l.) – nabývá hodnot přirozených čísel a informuje o celkovém počtu odkazů směřujících mimo website.
- Posun (dis) – nabývá hodnot celých čísel a informuje při opakované analýze konkurence změna pořadí konkretního dokumentu oproti předchozí analýze.
- Celá URL (f.URL) – ukazatel má čistě informativní charakter o URL (WWW adresa) analyzovaného dokumentu.

- Doménové jméno (domain) – ukazatel informuje, pod jaké doménové jméno dokument spadá.

Dále je každému řádku tabulky vytvořena samostatná tabulky podávající informace o zpětných odkazech na příslušné konkurenční dokumenty. Jednotlivými ukazateli tabulky zpětných odkazů jsou:

- Google Toolbar PageRank (l.GTPR) – nabývá hodnot 0 – 10
- Text odkazu (text) – nabývá hodnot 0 nebo 1 a informuje o zastoupení klíčového slova v textu odkazu směřujícího na dokument provozovatele website.
- Tématická relevance odkazů (t.r.) – nabývá hodnot 0 – 10 a jedná se o subjektivní posouzení obsahu dokumentu, ze kterého vede odkaz na konkurenční dokument.

Veškeré informace jsou základem pro objektivní a subjektivní posuzování „úspěšného“ zobrazení konkurenčních dokumentů ve vyhledávači. Samotné celkové pořadí konkurenčních dokumentů je stanoveno seřazením výsledných hodnot jednoduchého vynásobení převrácené hodnoty pořadí dokumentu ve výsledcích vyhledávání a hodnoty klíčového slova získanou z analýzy klíčových slov. Pro získání nejkonkurenčnějších websites jsou výsledky pořadí konkurenčních dokumentů roztříděny dle doménových jmen (domain), přičemž hodnoty ze shodných domén jsou sčítány.

klíčové slovo
c_i 0,72

o	CVW	GTPR	title	h1	h2	h3	h4	h5	h6	des	kw	i.a.	B	w	d.kw.	dom	d.a.
1	0,72	7	1	1	1	1	0	0	0	1	1	40	10456	2345	0,876	1	1987
2	0,36	7	1	1	0	0	0	0	0	1	1	54	65432	5676	0,565	1	1676
3	0,24	5	1	1	0	1	0	0	0	0	1	43	45321	2123	0,66	0	1456
4	0,18	6	1	1	0	1	0	0	0	0	1	43	65456	3453	0,343	0	1345
5	0,14	4	1	0	0	0	0	0	0	0	1	47	87656	2134	0,345	0	1122
6	0,12	3	1	0	0	0	0	0	0	1	1	54	23434	2768	0,343	0	987
7	0,1	3	1	0	0	0	0	0	0	1	1	65	87954	3456	0,332	0	656
8	0,09	3	1	0	0	0	0	0	0	1	1	78	34768	4345	0,233	0	544
9	0,08	2	1	0	0	0	0	0	0	0	1	98	34987	4563	0,23	0	343
10	0,07	1	0	0	0	0	0	0	0	0	1	99	56321	7658	0,19	0	321

Tabulka 4: Ukázka tabulky pro analýzu konkurence, část 1.

o	URL	/	a.t.	b.s.	val	l.p.4	l.p.3	s.p.	l.p.w.	q.c.	l.	dis	f.URL	domain
1	1	1	4	5	1	34	78	204	76	8	12	0	url1	domena1
2	1	1	4	4	1	21	67	199	56	8	11	0	url2	domena2
3	1	2	2	3	0	21	67	204	45	6	8	0	url3	domena1
4	1	2	1	1	0	19	70	199	67	5	9	0	url4	domena3
5	0	0	1	0	0	14	43	45	43	6	12	0	url5	domena4
6	0	0	1	0	0	12	67	43	23	4	15	0	url6	domena5
7	0	0	1	0	0	12	23	28	23	3	24	0	url7	domena6
8	0	0	0	0	0	10	10	28	19	3	21	0	url8	domena6
9	0	0	0	0	0	9	6	29	17	3	8	0	url9	domena5
10	0	0	0	0	1	9	6	0	17	3	8	0	url10	domena7

Tabulka 5: Ukázka tabulky pro analýzu konkurence, část 2.

f.URL	l.GTPR	text	t.r.
url1	8	1	8
url2	8	1	8
url3	9	1	7
url4	7	1	7
url5	6	1	9
url6	8	1	6
url7	9	0	5
url8	5	1	4
url9	4	1	4
url10	4	1	4

Tabulka 6: Ukázka rozboru části zpětných odkazů k dokumentu umístěné na první pozici ve výsledcích vyhledávání.

Ukazatel konkurenční hodnoty dokumentu (*CVD*)[107] je vypočítán pro každý řádek tabulky, čili pro každý dokument z výsledku vyhledávání dle vzorce

$$CVD = c_i * \frac{1}{o}.$$

CVD – ukazatel konkurenční hodnoty dokumentu v rámci klíčového slova nebo klíčové fráze.

c_i – hodnota z vícekriteriální analýzy variant analýzy klíčových slov (např. hodnota z metody TOPSIS nebo metody váženého součtu).

o – pozice dokumentu ve výsledcích vyhledávání.

Druhým důležitým ukazatelem v rámci klíčového slova je konkurenční hodnota website (CVW_i)[108], která je součtem konkurenčních hodnot dokumentů pocházejících z jednoho doménového jména. Vzorec pro výpočet CVW_i je

[107] Competitive Value of WWW Page
[108] Competive Value of Domain

$$CVWi = c_i * \left(\frac{1}{o_1} + \frac{1}{o_2} + ... + \frac{1}{o_k} \right) = c_i * \sum_{j=1}^{k} \frac{1}{o_j} .$$

CVW_i – ukazatel konkurenční hodnoty website v rámci klíčového slova nebo fráze.

o_1 – o_k – jednotlivé pozice dokumentů ve výsledcích vyhledávání spadajících pod jedno doménové jméno.

Třetím přínosným ukazatelem je konkurenční hodnota website (*CVW*) za všechna klíčová slova, tedy součet konkurenčních hodnot website z jednotlivých klíčových slov (CVW_i), popsaný vzorcem

$$CVW = c_1 * \left(\frac{1}{o_1} + \frac{1}{o_2} + ... + \frac{1}{o_k} \right) + c_2 * \left(\frac{1}{o_1} + \frac{1}{o_2} + ... + \frac{1}{o_k} \right) + ...$$
$$... + c_i * \left(\frac{1}{o_1} + \frac{1}{o_2} + ... + \frac{1}{o_k} \right)$$

nebo-li

$$CVW = \sum_{i=1}^{n} \left(c_i * \sum_{j=1}^{k} \frac{1}{o_j} \right) = \sum_{i=1}^{n} CVW_i .$$

CVW – ukazatel konkurenční hodnoty website za všechna klíčová slova nebo fráze.

c_1 – c_i – hodnoty z vícekriteriální analýzy variant analýzy klíčových slov (např. hodnota z metody TOPSIS nebo metody váženého součtu).

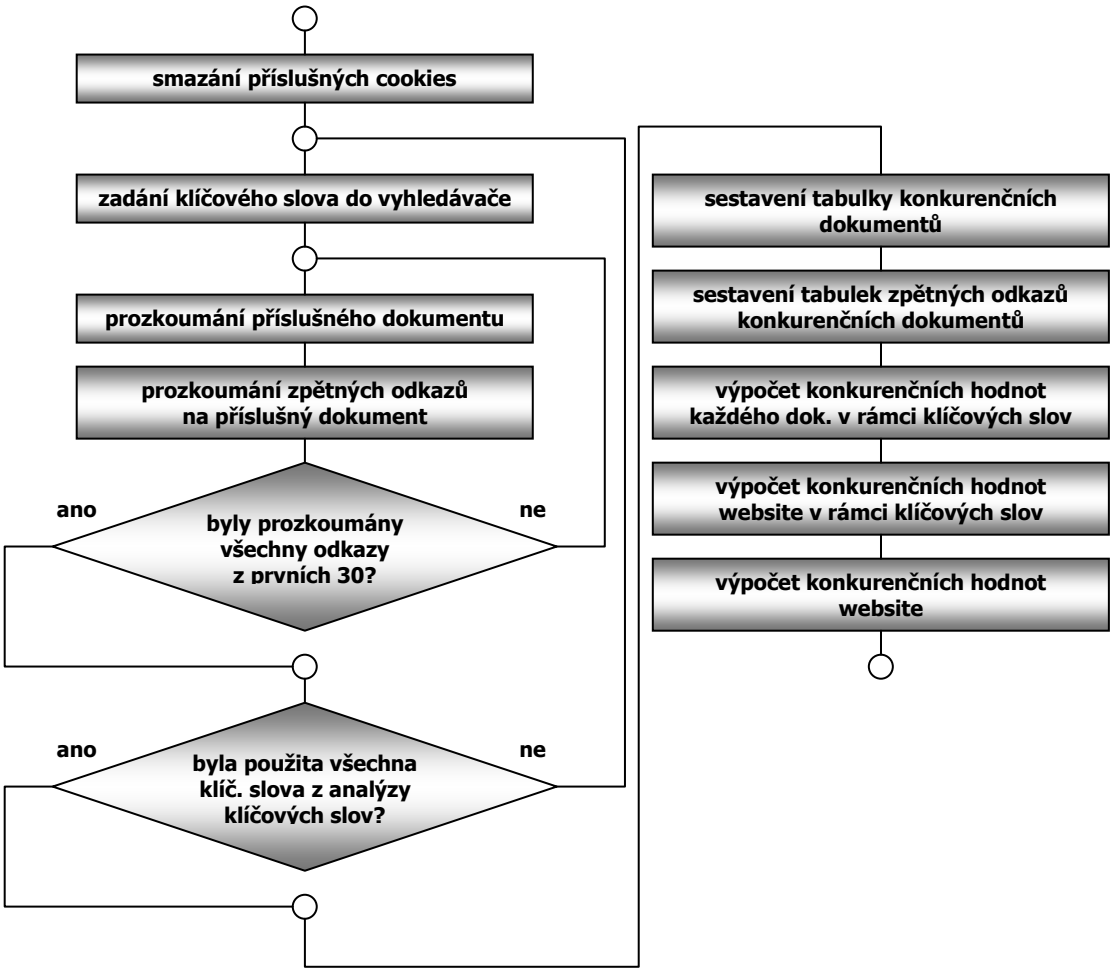


Schéma 3: Metodický postup pro analýzu konkurence.

9.3 Analýza aktuálního stavu website

Pokud provozovatel website již určitou WWW prezentaci vlastní, je nutné na ni provézt analýzu stávajícího stavu, tedy jistý audit, který částečně kopíruje postupy z analýzy konkurence. MPOD expert nebo provozovatel website vypočítá konkurenční hodnotu každého dokumentu pro každé klíčové slovo nebo frázi z analýzy klíčových slov. Pokud ve výsledcích vyhledávání není odkaz na žádný dokument provozovatele website, je konkurenční hodnota dokumentu nulová.

Dalšími velmi důležitými ukazateli každého dokumentu jsou shodné s analýzou konkurence, zejména, Google Toolbar PageRank každého dokumentu, počet zpětných odkazů, průměrná hodnota Google Toolbar PageRank dokumentů okazujících na analyzovaný dokument provozovatele website a relevance obsahů dokumentů.

Vyhledávácí robot nepřistupuje k dokumentu jako klasický uživatel. Určitě neindexuje obsah grafických formátů a má omezený přístup do Java appletů a Flash

animací, lze tedy předpokládat, že vyhledávací robot přistupuje k dokumentu v textové podobě, podobně jako například nevidomí uživatel. Pro snažší představu o zobrazení WWW stránky vyhledávacím robotem jsou využívány různé SW aplikace, např. Poodle Predictor[109].

Poodle Predictor je vlastně WWW nástroj pro získání představy, jakým způsobem bude dokument (WWW stránka) zobrazován ve výsledcích vyhledávání a jakým způsobem k ní vyhledávací robot přistupuje. Samotný spider Poodle Predictor obsahuje tři základní komponenty:

- Poodle predictor's Spider view
- Poodle predictor's Diagnostic view
- Poodle predictor's Source-code view

V první částí „Spider view" po zadání URL dokumentu vrátí, jak by mohl vypadat odkaz na dokument ve výsledcích vyhledávaní po zadání URL v samotném vyhledávači. Služba „Diagnostic view" nástroje Poodle Predictor zobrazí barevně odlišený výstup, který je základním vstupním prvkem pro roboty vyhledávače. Barevné odlišení je použito z důvodu snadné orientace MPOD experta nebo provozovatele website ve výstupu služby „Diagnostic view". Služba „Source-code view" zobrazuje barevné rozlišení samotného zdrojového kódu dokumentu (WWW stránky).

Pro komplexnější získání informací o hustotě klíčových slov existujícího dokumentu lze využít například nástroj Keyword Density & Prominence[110], kde je získána nejen hustota klíčových slov na WWW stránce, ale i jejich samotné umístění v XHTML značkách.

Posledním krokem analýzy stávajícího stavu dokumentu (popř. celé webové prezentace) je rozbor chování samotných uživatelů, tedy, jakým způsobem přistupují na dokument provozovatele (přes vyhledávače, katalogy či přímým přístupem), jakými klíčovými slovy a jak se na stránkách chovají, např. Google Analytics[111]. Část provozovatelů website však nechce poskytovat vyhledávači Google svá data. V takovém případě je nutné využít jiných služeb pro mapování uživatelů, popř. zakomponovat do website své vlastní řešení.

[109] http://www.gritechnologies.com/tools/spider.go
[110] http://www.ranks.nl/cgi-bin/ranksnl/spider/spider.cgi?lang=
[111] http://www.google.com/analytics/

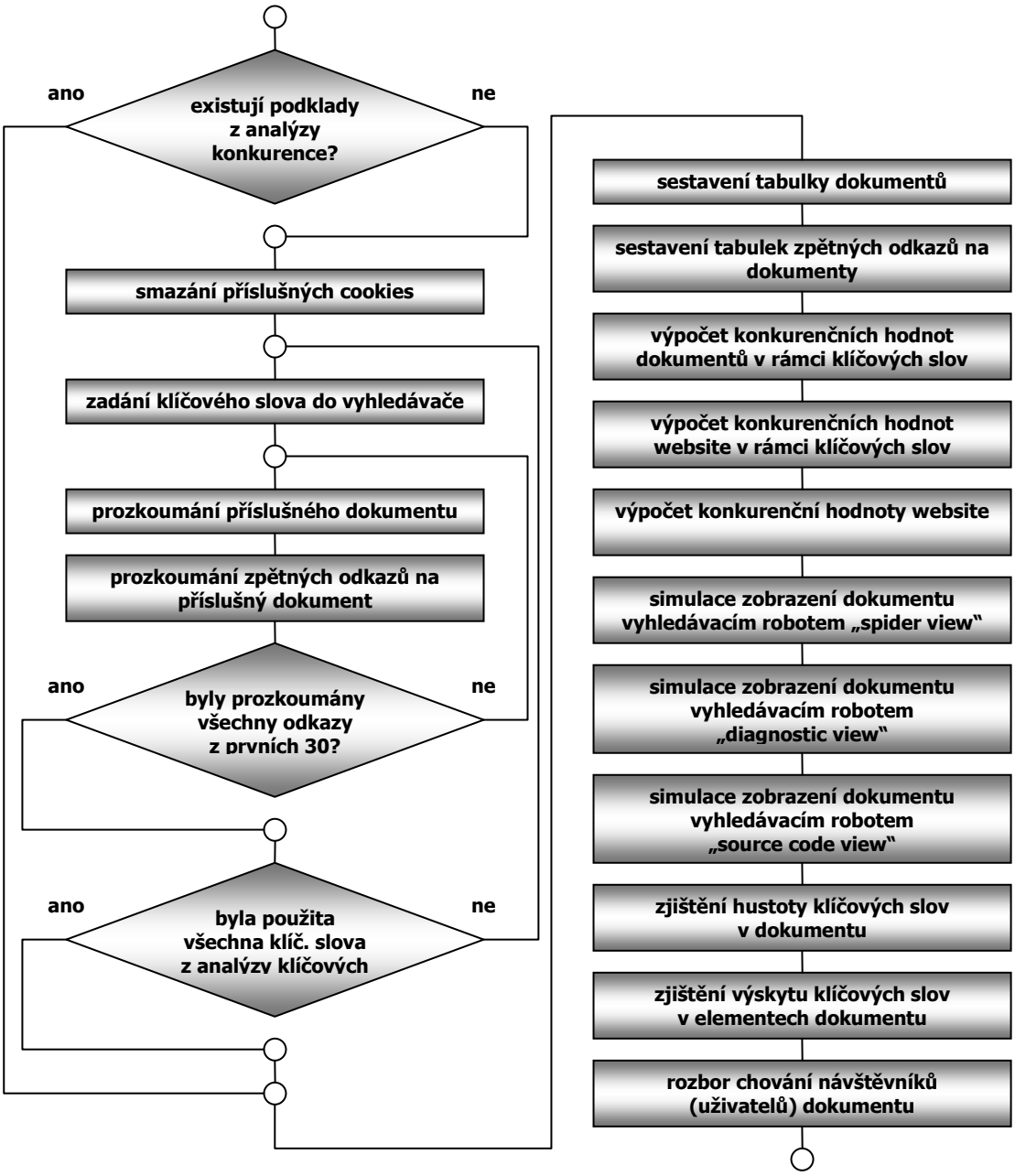


Schéma 4: Metodický postup pro analýzu aktuálního stavu website.

9.4 MPOD copywriting

Copywriting pro web se od klasického copywritingu liší tzv. „těkavostí uživatele“. Na papíře je uživatel schopen přečíst dlouhý text v celku ovšem na webu se čte jinak. Člověk většinou těká očima po stránkách. Hledá záchytné body, skáče z jednoho místa na druhé. Text pro web musí být krátký, s krátkými odstavci. Jinak se čtenář unaví a odejde jinam. Psaní pro web se také vyznačuje častějším používáním tučného písma a kurzívy, čili zvýrazněnými částmi textu. Samotné sdílení informací je

tedy disciplína, která vyžaduje náležitý trénink, jasné cíle a odpovídající prostředky [26].

Tradiční copywriting rozeznává čtyři základní typy reklamy podle jejich účelu:

- přímý prodej zboží a služeb
- vzbuzení zájmu a poptávky
- budování povědomí o zboží a službách
- budování firemní image a značky

V tradičních médiích jsou typickými příklady přímého prodeje inzeráty na některé publikace (například encyklopedie) doplněné odstřihávací objednávkou, direct maily nebo třeba teleshopping. Na webu se takto prodávají knížky, krabicový software, nejrůznější spotřební zboží a jednoznačně definovatelné služby (například webhosting). Na jednu stranu se přímý prodej hodí pro web velmi dobře, na druhou stranu je však pro textaře pořádnou výzvou a prubířským kamenem jeho schopností. Jediným cílem textu je v tomto případě závazná objednávka a především copywriter zodpovídá za získání zákazníka. Pokud zájemci neposkytne dost informací a pokud ho nepřesvědčí o výhodách nabízeného produktu, neprodá. To je důvod, proč bývají tyto typy textů hodně dlouhé (i přes tisíc slov), plné informací, technických detailů, příkladů použití, citací nezávislých hodnocení, odpovědí na časté otázky a podobně. Na webu mívají buď podobu jednoho dlouhého dokumentu členěného četnými podnadpisy, nebo bývají rozděleny na více menších dokumentů.

Účelem mnoha obchodních webových textů je vzbudit zájem o zboží či služby. Od čtenáře se v tomto případě nemusí očekávat závazná objednávka, nýbrž jen žádost o konkrétnější nabídku, dotaz nebo jiný projev zájmu. Úspěšné uzavření obchodu už je pak na obchodníkovi a jeho dialogu se zájemcem. Proto je tento typ důležitý pro B2B sektor, drahé zboží (včetně realit) a komplexní služby (vývoj software, webdesign). Zatímco v tradičních, neinteraktivních médiích je prakticky vždy dominantním prvkem reklam tohoto typu výrazný kontakt (telefon, adresa), na webu se může navíc jednat i o odkaz. Cílem textu pak je, aby se zájemce proklikl na jinou stránku. Patří sem tedy i úvodní či další navigační stránky webů, různé upoutávky na okrajích stránek, nebo textová reklama ve vyhledávačích.

Mnoho druhů zboží nelze prodávat na dálku a nelze ani očekávat, že je někdo bude přes web poptávat. Mnozí výrobci dokonce neprodávají koncovým zákazníkům

vůbec. Pak může být cílem textu jen budování a posilování všeobecného povědomí o výrobku či službě s tím, že zájemce později nakoupí v maloobchodní síti. MPOD copywriter tu má zdánlivě snadný úkol, protože výsledky jeho práce jsou tentokrát těžko měřitelné. I neprodejní stránky však lze napsat dobře. Stačí si uvědomit, že na webu se značka buduje především detailními a aktuálními informacemi, které ocení jak zájemci během nákupního rozhodování, tak uživatelé, kteří si výrobek již koupili.

V určitých případech je marketingová strategie založena více na image firmy a značky, než na povědomí o jednotlivých produktech. Jindy jsou firma a její produkt prakticky totéž (třeba internetový portál). V zásadě se ale jedná o typ textu dost podobný předešlému. Nejtěžší, co zde copywritera čeká, je vzbuzení a udržení zájmu čtenářů. Čistě firemní prezentace proto bývají častěji orientované spíše na novináře (tiskové zprávy), investory (obchodní výsledky) a potenciální obchodní partnery [34].

Pro samotnou tvorbu obsahu dokumentů je opět základem analýza klíčových slov. Moderní fulltextové vyhledavače sice neurčují relevanci stránky podle absolutní četnosti klíčového slova v textu, ovšem počtu výskytů slova v textu přiřazují vysokou váhu. Činí tak ve vztahu k celkové délce textu, a to buď lineárně (tzv. hustota klíčových slov) nebo nelineárně. Pro významná klíčová slova se za optimální většinou považuje hustota mezi 2 až 10 procenty. Od určité hranice již větší počet výskytů klíčového slova pozici dokumentu ve výsledcích vyhledávání neovlivňuje.

Klíčová slova nelze dle jejich důležitosti psát za sebou a oddělovat je čárkou, neboť vyhledávací robot by mohl takový dokument vyhodnotit jako spam. Časté nelogické opakování klíčových slov rovněž není pro samotný dokument či celý website přínosem. Celý text by měl být koncipován pro uživatele zajímavou formou, přičemž důležitější klíčová slova jsou použita v přední části textu, kde je i vyšší samotná četnost všech klíčových slov.

Četnost klíčových slov by měla korespondovat s výsledky analýzy klíčových slov, tzn. že v textu by měla být nejvíce zastoupena klíčová slova, která se umístila na předních místech. Hustota klíčových slov by se však neměla přeceňovat. Jednak to není zdaleka jediné kritérium, podle kterého vyhledavače stránky hodnotí, a jednak je většinou důležitější stylistická kvalita, informační hodnota a jazyková pestrost textu. Ideální text by tedy měl obsahovat co nejvíce různých klíčových slov s podobným

významem, zároveň však musí být dobře napsaný, srozumitelný, čtivý, přesvědčivý a přirozený.

Psaní textů pro web je komplikovaná disciplína, která vyžaduje stručnost, výborné členění textu, přirozenost a orientaci na potřeby uživatele. Čitelný, přehledně strukturovaný a srozumitelný obsah je jedním ze základních předpokladů dobře použitelného webu. Návštěvník se musí v textu orientovat. Důležité je jasné a intuitivní rozdělení textu neboť správně napsaný text je přístupný úplně všem. Tento předpoklad ovlivňuje nejen zvolená forma a technické prostředky, ale také autorský styl. Kvalitní text na Internetu je napsaný s ohledem na všechny čtenáře a díky tomu je srozumitelný a jeho smysl jasně pochopitelný.

Správně napsaný internetový text je kvalitně zpracovaný i po technické stránce. Obsah je uzpůsoben tak, aby se dal snadno a efektivně upravovat a aktualizovat. Čím složitější je práce s textem, tím větší jsou náklady na jeho správu a tím i na správu celého website.

Obsah je tedy alfou a omegou úspěšného textu na Internetu. Špatný obsah, byť sebelépe zpracovaný, nikoho nezaujme. Při vytváření obsahu musí tvůrce myslet především na to, aby byl dostatečně srozumitelný a pochopitelný pro všechny. Text dokumentu musí být dobře zařazen v rámci celého website a čtenář by měl mít o jeho zařazení přehled a snadno se orientovat.

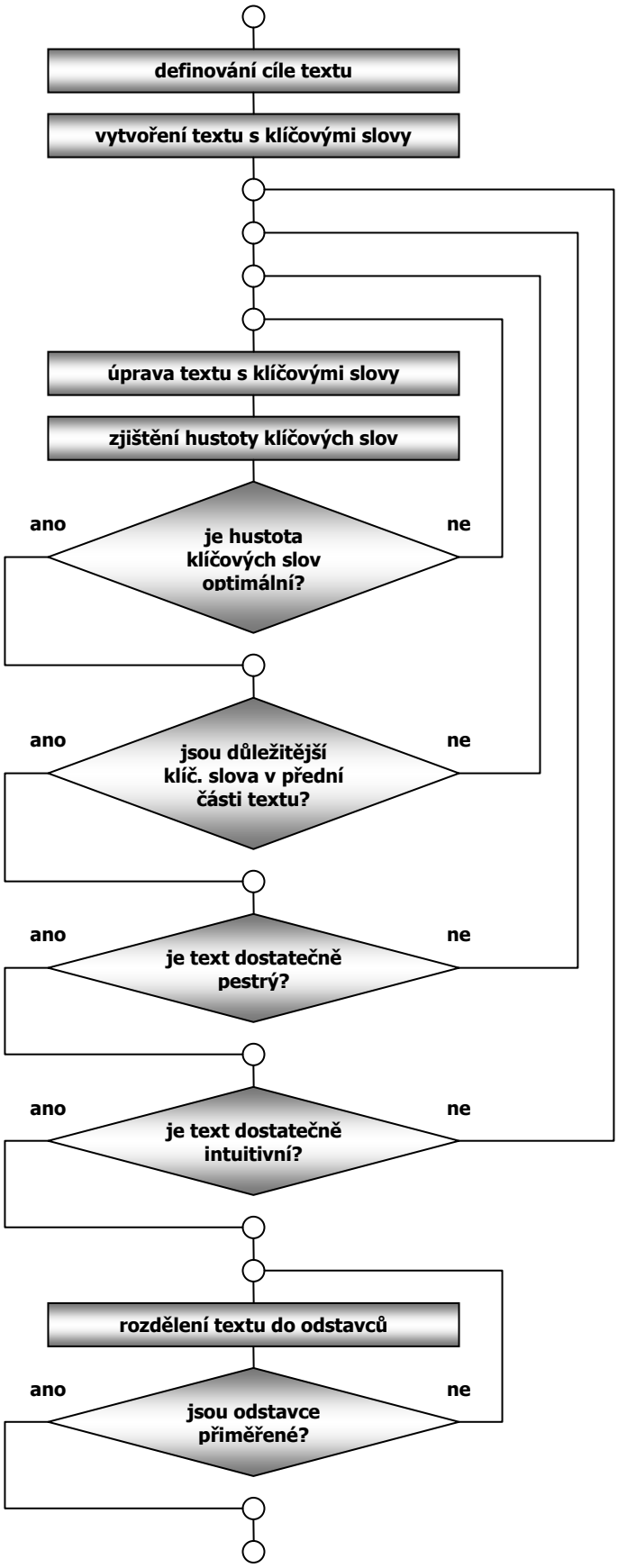


Schéma 5: Metodický postup pro copywriting.

9.5 Optimalizace dokumentu a serveru

MPOD musí projít, musí projít každý dokument z celého komplexu provozovatelova website. Základem pro MPOD je primárně analýza klíčových slov, analýza konkurence a copywriting, popř. i analýza současného stavu již existujícího website stránek. Při vlastní optimalizaci musí MPOD expert brát v potaz:

- důležitost klíčového slova, která je výsledkem analýzy klíčových slov
- vliv XHTML značek a ostatních faktorů na vyhledávací roboty
- možný počet výskytů XHTML značky v dokumentu a vliv ostatních faktorů na indexaci vyhledávacími roboty
- možný počet výskytů klíčového slova v dokumentu, který je výsledkem copywritingu
- konkurenční hodnoty dokumentů a ostatní faktory z analýzy konkurence

Z MPOD copywritingu vyplívá obsah každého dokumentu, který je určující pro samotný postup optimalizace dokumentu. Pokud např. z analýzy klíčových slov je určité klíčové slovo nebo fráze velmi důležité a budoucí obsah dokumentu má vysokou hustotu tohoto klíčového slova, určitě by takové klíčové slovo mělo figurovat u faktorů spojených s užitím klíčových slov.

9.5.1 Klíčová slova v dokumentu

Každý dokument, pokud se jedná o XHTML dokument, musí mít charakteristický titulek vystihující obsah dokumentu, který by měl obsahovat klíčové slovo, popř. slova nebo frázi, s nejvyšší důležitostí z analýzy klíčových slov. Element title dokumentu je považován za faktor se zásadním vlivem na vyhledávací roboty a lze ho v každém XHTML dokumentu použít pouze jednou. Lepší než obyčejné vložení klíčových slov do titulku dokumentu je jejich zakomponování do určité fráze.

Klíčová slova či fráze použitá v těle dokumentu, tedy ve vlastním textu dokumentu, je faktor se silným vlivem na relevanci výsledků vyhledávání a celý text vychází právě z MPOD copywritingu. Je sice možné získat vysoký ranking i pro stránky, které neobsahují klíčová slova v textu, vyžaduje si to však od správce website mnohem vyšší úsilí a množství odkazů s klíčovým slovem zevnitř i zvně website. Z MPOD copywritingu a z analýzy klíčových slov vychází rovněž samotná četnost klíčových slov a frází v textu. Jak roste sofistikovanost vyhledávačů, stává se

pro určení významu dokumentu stále důležitějším nejen umístění hledané fráze v dokumentu, ale také celkový mix všech použitých slov a frází.

Jak již bylo zmíněno v předchozích částech textu, klíčová slova musí mít určitý vztah k celkovému obsahu dokumentu, tzn. měla by být v tématickém souladu. I když vztah obsahu dokumentu a klíčových slov je považován za faktor se silným vlivem na hodnocení relevance dokumentu vyhledávacími roboty, v budoucí době se stane ještě mnohem důležitějším.

Další zásadou optimalizace XHTML dokumentu je zakomponování nejdůležitějších klíčových slov či frází do značky pro nadpis nejvyšší úrovně. XHTML značka `h1` je považována za faktor se silným vlivem na ranking XHTML stránky, nicméně by text nadpisu měl být alespoň částečně odlišný od titulku dokumentu. Samotnému rankingu dokumentu může uškodit, pokud by byla značka pro nadpis nejvyšší úrovně použita v těle dokumentu víckrát, je tedy nutné dodržet logickou strukturu obsahu XHTML dokumentu a použít nadpis nejvyšší úrovně pouze jednou v horních partiích dokumentu.

Klíčové slovo nebo slova v názvu doménového jména hrají rovněž vysokou roli v samotném procesu MPOD. Celý website, je uložen pod jedním doménovým jménem. Z mnoha experimentů vyplývá, že nejlépe jsou vyhledávacími roboty hodnoceny doménová jména obsahující výstižná klíčová slova. U víceslovních doménových jmen jsou jednotlivá slova oddělena pomlčkami, neboť znak pomlčky je vyhledávacími roboty považován za oddělení jednotlivých slov. Samozřejmě, lze využít i víceslovních doménových jmen, u kterých slova nejsou oddělena pomlčkou a příslušný ranking dokumentu získat pomocí ostatních faktorů ovlivňujících pozici dokumentu ve výsledcích vyhledávání, z uživatelského hlediska jsou však mnohem lepší kratší doménová jména, přičemž nejvýznamnější klíčová slova by měla být obsahem doménového jména a u víceslovních doménových jmen by významnější klíčové slovo mělo být na předním místě. Problémem může být, pokud je požadované doménové jméno již registrováno konkurencí. V takovém případě je nutné zvolit jiné, podobné doménové jméno, nebo požadovaný ranking dokumentu získat kvalitním obsahem a vhodným URL.

Každý dokument celého website má jedinečný URL, tedy WWW adresu, přes kterou lze dokument ve klientském WWW prohlížeči získat. „Přívětivé“ URL je takové, které neobsahuje otazník, za kterým jsou předávány tzv. parametry mezi

klientem a serverem. URL je rovněž významné, pokud obsahuje určitá klíčová slova z obsahu dokumentu (opět dle jejich významu) a zároveň nepoukazuje a silnou míru „vnoření“ dokumentu v adresářové struktuře website, které je určeno počtem znaků lomeno „`/`“. Jedním z přínosných přístupů je, pokud URL za doménovým jménem obsahuje pouze jeden znak lomítka a následná část s klíčovými slovy vystihuje obsah dokumentu např. v korespondenci s nadpisem nejvyšší úrovně. Celá výsledná podoba URL je výsledkem korektní optimalizace WWW serveru.

Dalšími XHTML značkami, které mají vliv na výslednou pozici XHTML dokumentu jsou značky pro nadpisy druhých a nižších úrovní. Standard jazyka XHTML popisuje nadpisové značky do šesté úrovně, přičemž vyhledávací roboty ovlivňuje zejména již zmíněný nadpis první úrovně a následně nadpisy druhé a třetí úrovně. Nadpisy druhé a třetí úrovně se mohou v dokumentu vyskytovat několikrát, ale pokaždé by měly obsahovat jiný text a alespoň z určité části tvořený klíčovými slovy, které mají vysoký stupeň důležitosti (ne nejvyšší[112]) z analýzy klíčových slov. Nadpisy nižších úrovní sice nemají oproti běžnému textu v dokumentu o mnoho vyšší vliv (určitý však ano) na umístění dokumentu ve výsledcích vyhledávání, nicméně je vhodné je používat z důvodu přehlednosti textu a logické struktury textu.

Klíčová slova použitá v atributech `alt` a `title` značky pro umístění obrázku do XHTML dokumentu mají také určitý vliv na celkový ranking dokumentu. Výhodou je, že XHTML značka pro obrázek může být v dokumentu použita vícekrát a proto je v tomto případě vhodné použít širší spektrum klíčových slov, nikoli pouze jen ta nejdůležitější. Ovšem, tvůrce dokumentu (MPOD expert) by měl důležitější slova použít v předních partiích použitého textu a měla by být v určitém souladu s obsahem obrázku.

Částečný vliv na celkovém hodnocení relevance dokumentu vyhledávacím strojem mají rovněž značky `b`, `strong`, `i` a `em`, čili značky pro určité zvýraznění textu. Tím, že je text zvýrazněn, má vyšší důležitost od ostatního textu dokumentu, proto je vhodné využívat pro důležitější výrazy (klíčová slova či fráze) právě XHTML značek oproti zvýraznění textu kaskádními styly, které jsou rozpoznatelné pouze ve WWW prohlížeči se zapnutou podporou CSS.

[112] Slova s nejvyšším stupněm důležitosti z analýzy klíčových slov by měla být použita v nadpise první úrovně.

Klíčová slova použitá v metaznačce typu `description` v hlavičce XHTML dokumentu by měla být formulována do určitého srozumitelného textu, např. anotace celého obsahu dokumentu a pro každý dokument by měl být tento popisek jedinečný, neboť jej lze použít pouze jednou. Měl by tedy obsahovat nejdůležitější klíčová slova.

Metaznačka `keywords` v hlavičce XHTML dokumentu má minimální vliv na celkové hodnocení relevance dokumentu a např. vyhledávač Google ji zcela ignoruje. Existují však vyhledávače, které této metaznačce ještě přikládají určitý význam a potenciální zákazník může na dokument (tedy website) přijít právě přes tento vyhledávač. Metaznačka může být v každém XHTML dokumentu použita pouze jednou, ale může obsahovat všechna klíčová slova z analýzy klíčových slov.

9.5.2 Vlastnosti dokumentu

Link popularity v rámci vnitřní struktury website se vztahuje k počtu a důležitosti odkazů, které vedou na cílový dokument zevnitř website. Bez odpovídající struktury odkazů nezískají některé stránky patřičnou váhu. Je vhodné použít klíčový text v odkazech, protože takové odkazy podstatně zvýší hodnotu navigačních odkazů.

Tvůrce dokumentu (MPOD expert) by měl věnovat vysokou pozornost kvalitě a relevanci odkazů na externí dokumenty a websites, protože se jedná o faktor se silným vlivem na hodnocení relevance dokumentu ve výsledcích vyhledávání vyhledávacím strojem. Text odkazu by měl obsahovat klíčová slova, která vyšla z analýzy klíčových slov a zároveň by měl korespondovat s tématickým obsahem obou dokumentů (ze kterého je odkaz i na který link odkazuje).

Stáří samotného dokumentu[113] má rovněž silnou roli na umístění dokumentu ve výsledcích vyhledávání. Starší dokumenty mohou být vnímány jako autoritativnější, zatímco novější mohou být vůči tématu dočasně relevantnější. Starší dokumenty mohou být také důvěryhodnější, zvláště pokud jsou často citovány a neobsahují příliš mnoho nefunkčních odkazů. V případě blogů a zpravodajských serverů mívají novější stránky obvykle vyšší PageRank, vlivem interní struktury webu. Nové dokumenty mohou být také podpořeny svou aktuálností. Stáří domény[114] je důležitější než stáří dokumentu. Tudíž stejný dokument na starší důvěryhodné

[113] Stáří dokumentu je počítáno od první indexace dokumentu vyhledávacím strojem, nikoli od datumu vzniku dokumentu.
[114] Opět je myšleno datum první indexace jakéhokoli dokumentu vyhledávacím strojem z doménového jména, nikoli datum registrace doménového jména.

doméně bude hodnocen lépe než tentýž dokument na doméně nové. Záleží, zda informace, které jsou poskytovány samotným dokumentem, podléhají aktuálnosti. Pokud ano, je nutné dokument v určitých časových intervalech obměňovat.

Množství indexovatelného textového obsahu dokumentu se vztahuje přímo k množství viditelného textu a vychází z MPOD copywritingu. Pokud se velikost příliš mnoha stránek blíží nebo rovná velikosti mnoha dokumentů považovaných za spam, může být konkrétní dokument s vysokou pravděpodobností také za spam považován. Google také neustále zpřísňuje svá kritéria pro dokumenty v primárním indexu požadavkem na určitou minimální hodnotu jejich PageRanku. Delší a autoritativnější dokumenty jsou lepší než články rozsekané na mnoho kratších stránek, přičemž nejdůležitější klíčová slova by měla být umístěna v horních partiích textu.

Kvalita obsahu dokumentu může mít rovněž jistý význam, vzhledem k předpokladu, že vyhledávače mohou používat textové, vizuální a další metody analýzy k určení platnosti (validity) a hodnoty obsahu. Proto je nutné, aby výsledkem MPOD copywritingu byly kvalitní a zajímavé texty.

Částečný vliv na výsledné umístění dokumentu ve vyhledávači má i uspořádání toku textu v dokumentu, například v publicistických článcích se vyžaduje úvod zaměřený na konkrétní detail, následovaný co nejširším přehledem informací o problematice s postupným zvyšováním detailnosti textu ke konci dokumentu. Takováto organizace toku dokumentu má velký význam u specializovaných dokumentů, tedy zaměřených na určité specifické téma.

Množství změn učiněných v dokumentu a časový interval, v němž k těmto změnám dochází rovněž částečně ovlivňuje výsledné umístění dokumentu ve výsledcích vyhledávání. Pokud je dokument nebo celý website často aktualizován, je i častěji navštěvován a indexován vyhledávači.

Validita dokumentu podle standardu W3C sice v současné době ovlivňuje vyhledávací roboty minimálně, nicméně v budoucnosti se bude význam validity zvyšovat. Tabulkový layout, formátování textu značkou font, či tzv. „špagetový kód“ zdvojnásobují až ztrojnásobují velikost i té nejjednodušší stránky. Návštěvníci (uživatelé) pak platí zbytečnou cenu za čekání na zobrazení celé stránky. Mnoho uživatelů se setkalo s problémem, kdy měli nainstalovaný nejnovější WWW prohlížeč a při návštěvě website jim byla zobrazena informace, že je vyžadován „moderní“

WWW prohlížeč, který byl ve skutečnosti mnohem starší, než který uživatelé používal. V takovém případě ztrácí celý metodický postup optimalizace dokumentu smysl. Technologie vytvořené sdružením World Wide Web Consorcium a další standardní techniky jsou podporovány většinou moderních WWW prohlížečů a umožňují vytvářet takové dokumenty, které budou dostupné i na nových verzích WWW prohlížečů i v případě, že dojde k dalšímu vývoji samotných standardů.

9.5.3 Doména dokumentu

Stáří website má také významnou roli při optimalizaci dokumentu pro vyhledávací roboty. Nejde ani tak o datum registrace domény, jako spíše o zpřístupnění indexovatelného obsahu čitelného pro vyhledávače (může se měnit se změnami vlastníka domény). Velmi nové weby mohou pomaleji získávat v hodně konkurenčních oblastech, protože stáří doménového jména (zpřístupnění indexovatelného obsahu vyhledávačům) má do jisté míry podíl na relevanci webu k danému tématu.

Částečný vliv na vyhledávací stroje, tedy na hodnocení relevančnosti dokumentu na daná klíčová slova má i četnost přidávání nových dokumentů do komplexu website, ovšem v závislosti na kontextu. Zpravodajský server může snadno vyprodukovat 1 000 nových dokumentů denně, zatímco web o domácím mazlíčkovi může mít jednu týdně, přičemž je to normální pro tento typ webu. Přidávání dokumentů tedy musí být v souladu s „normálem“ k danému tématu a povaze website.

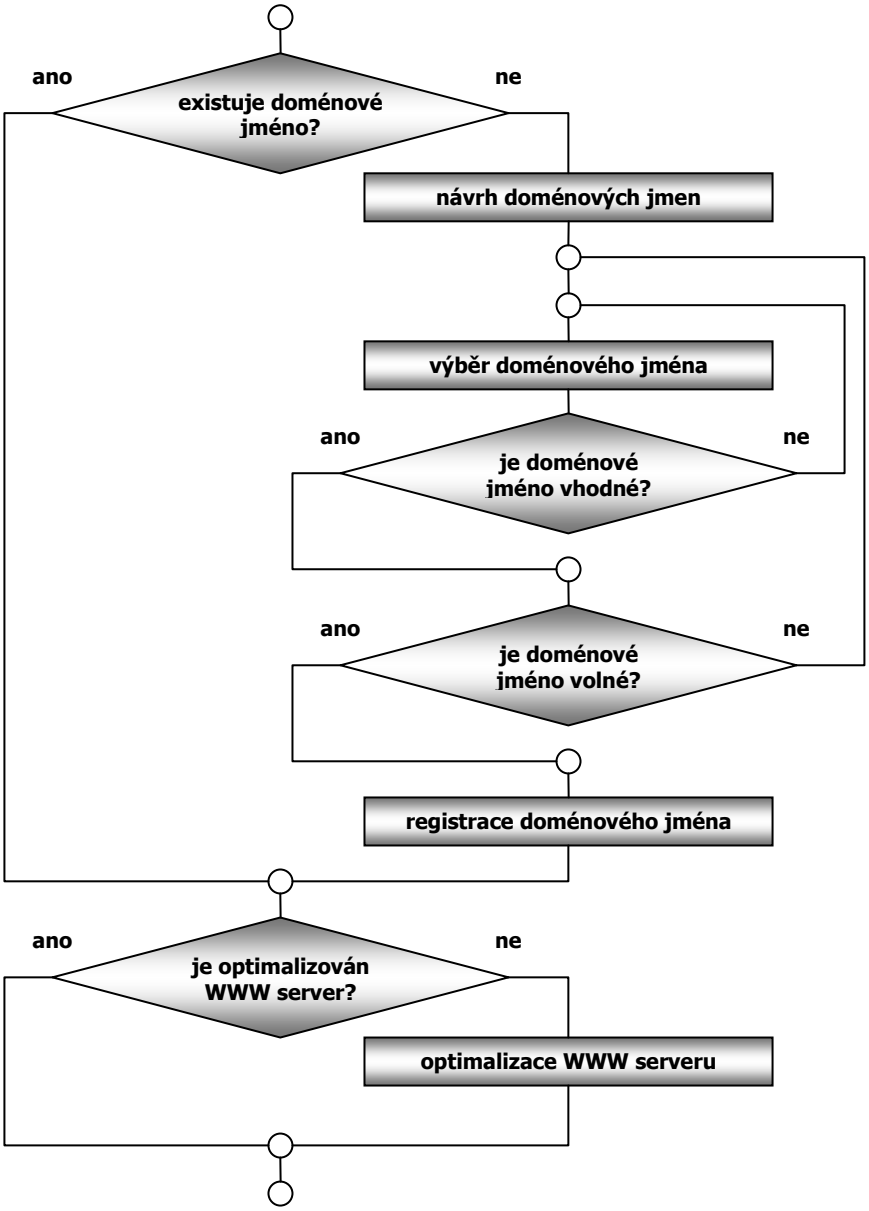


Schéma 6: Metodický postup pro optimalizaci dokumentu a serveru, část 1.

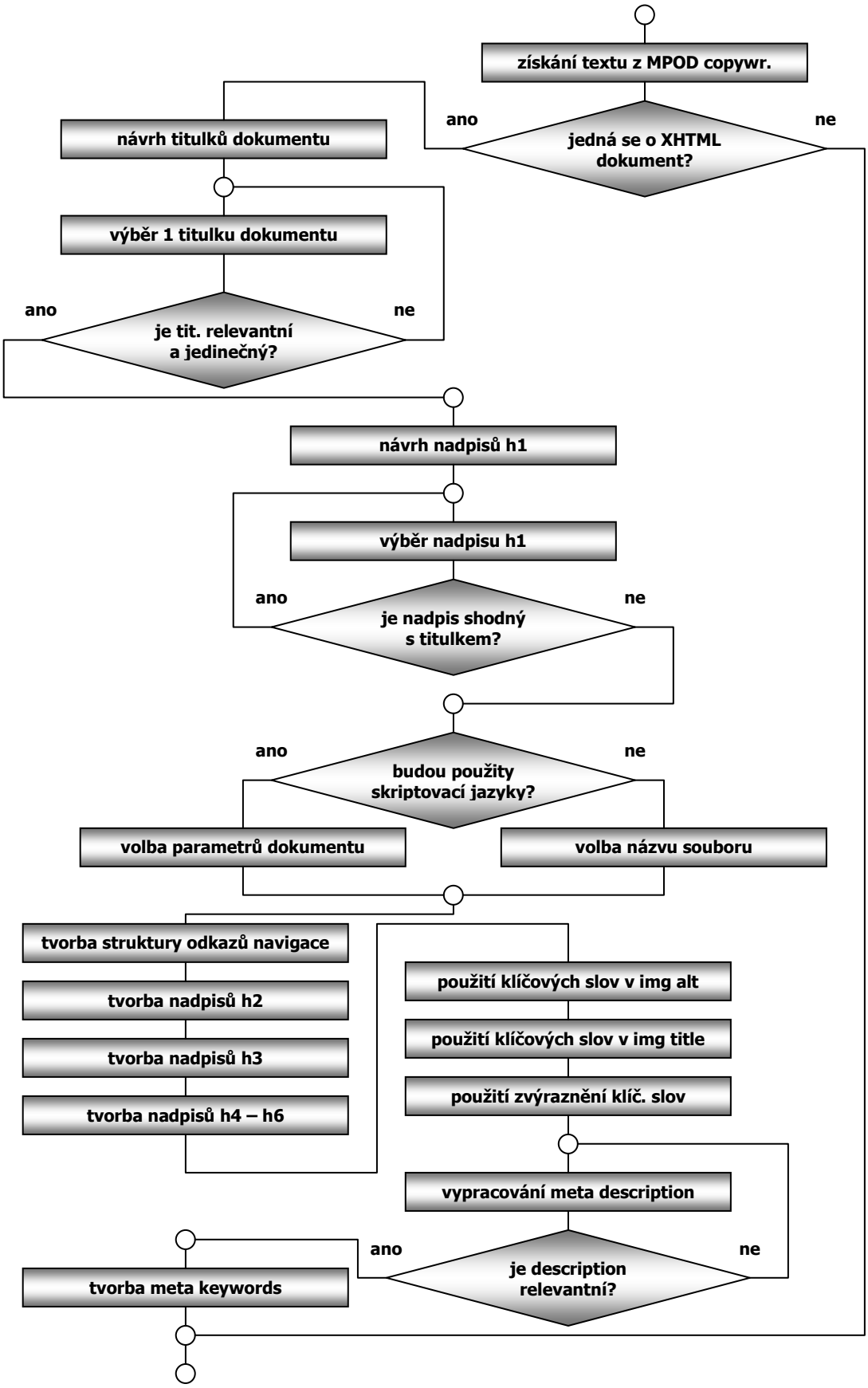


Schéma 7: Metodický postup pro optimalizaci dokumentu a serveru, část 2.

9.6 Tvorba zpětných odkazů a registrace do katalogů

Zpětné odkazy jsou odkazy, které odkazují na provozovatelův website nebo dokument a získávají se většinou výměnou. Při budování zpětných odkazů je tématická relevance odkazů směřujících na konkrétní dokument velmi důležitá. Vyjadřuje tématickou příbuznost mezi odkazujícím dokumentem či website, cílovým dokumentem a klíčovým slovem.

Rovněž zásadní vliv na posouzení relevance dokumentu vyhledávacími stroji má link popularity webu v rámci skupiny tématicky souvisejících webů, který se vztahuje k počtu a důležitosti odkazů, které vedou na cílový website ze všech tématicky příbuzných webů na Internetu. Vysoce specializovaný website nemusí mít velké množství odkazů, ale několik odkazů od tzv. „autorit“ obvykle dostačuje k tomu, aby získal vysokou pozici na klíčová slova ze svého specializovaného oboru.

Při tvorbě zpětných odkazů musí být vhodně vybrán text odkazů, který je obsažen v elementu (značce) a, neboť text odkazu, přes přesunutí důrazu Googlem od textů tvořících odkazy k autoritě domény a počtu citací, má na umístění dokumentu ve výsledcích vyhledávání stále zásadní vliv. Text tvořící odkaz je tedy tím nejjednodušším ukazatelem toho, o čem si odkazující myslí, že je obsahem dokumentu či website. Je však důležité zachovat kontext s klíčovými slovy i okolního textu odkazu neboť i ten hraje významnou roli. Text okolo odkazu bude pravděpodobně mnohem důležitější v budoucnu. Například sady odkazů v patičkách stránek jsou pro Google jasným znamením, že tyto linky nemají takovou hodnotu (může se jednat o copywrite, navigaci, odkaz na standardy, apod.).

Globální link popularity odkazujícího webu má na posouzení relevance dokumentu ke klíčovému slovu vyhledávacím strojem rovněž silný vliv a vztahuje se k počtu a důležitosti odkazů, které vedou na odkazující web ze všech ostatních webů na Internetu. Podle globálního linku popularity odkazujícího webu musí MPOD expert ve spolupráci s provozovatelem website posuzovat samotné websites, neboť link popularity odkazujícího dokumentu a webu je pravděpodobně nejdůležitějším faktorem hodnocení kvality odkazu.

Doba, po kterou odkaz na cílový web existuje, tedy stáří odkazu, se zřejmě v budoucnu stane základem vyhledávacích algoritmů. Z testů vyplývá, že pro Google je zapotřebí minimální stáří odkazu 3 až 4 měsíce – čím starší, tím lepší.

Pro umisťování zpětných odkazů je vhodné primárně tyto odkazy umisťovat do dokumentů popř. website s vyšší hodnotou PageRank nebo Google Toolbar PageRank. Samozřejmě, že ve finále je přínosem každý zpětný odkaz z relevantního website, nicméně výše Google Toolbar PageRanku má na hodnocení dokumentu vyhledávacím strojem také částečný vliv. Vyhledávací stroje rovněž částečně ovlivňují vlastnosti spojené s časem odkazu, tedy kdy byl odkaz vytvořen a kdy byl měněn, podobně jako samotné stáří odkazu.

Při tvorbě zpětných odkazů musí MPOD expert nebo provozovatel website vybrat vhodný titulek a popisek. Jak již bylo popsáno v předchozích částech textu, velkou váhu má text, který na dokument odkazuje. Pokud bude použit text ve tvaru „*Přijďte k nám nakoupit*“, dostane dokument sice zpětný odkaz, ale zároveň se přiřadí jistá váha při vyhledávání na dotaz „*Přijďte k nám nakoupit*“ což z hlediska metodického postupu optimalizace dokumentu pro WWW vyhledávače není přínosné, protože výraz „*Přijďte k nám nakoupit*“ nikdo nehledá.

Proto je mnohem účelnější použít titulek XHTML dokumentu (např. „*Kladiva a šroubováky ze všech koutů světa*“), který obsahuje klíčová slova kladiva a šroubováky. Tím získají na váze slova kladiva a šroubováky. Pokud provozovatel website žádá provozovatele či administrátora příslušného relevantního webu o vložení odkazu na svůj dokument, musí dodržovat určitá pravidla, zejména:

- Odkazovat jako první. Pokud se někomu ozve, je pravděpodobné, že právě provozovatel website potřebujete odkaz, ne správce dotyčného serveru. Proto odkazuje jako první.
- Neměl by zakazovat indexaci stránek, ze kterých odkazuje.
- Neměl by skrývat odkazy pomocí CSS, tedy nedávat jim barvu stejnou nebo podobnou jako pozadí apod.
- Pokud má speciální stránku pro odkazy, měl by se snažit, aby odkaz na ni byl dobře viditelný z hlavní stránky.

Registrace do katalogů má rovněž stále smysl. Převážná většina katalogů sice nepředává extrémní hodnotu PR (na stránce se vyskytuje velké množství odkazů, mezi které se PR dělí), nicméně v žádném případě se taková registrace neprojevuje na website provozovatele negativně. Ze zkušeností s registracemi do katalogů vyplývá, že když byl spouštěn nový web a je přidán do seznamu nejpoužívanějších

katalogů, bude Google Toolbar PageRank mezi hodnotami 2 a 4, což je dobrý základ pro další formu budování zpětných odkazů, např. výměnu odkazů mezi spřátelenými weby.

Samotné katalogy umožňují všem uživatelům:

- procházet stromovou strukturu sekcí podle oborů
- prohlížení odkazů
- vyhledávání odkazů podle určitého slova
- přidávání odkazů do databáze

Základ přidávání odkazu do katalogu je na většině katalogů stejný nebo velmi podobný a i zde je velmi důležitý titulek a popis odkazu. Titulek nesmí obsahovat žádná tzv. „balastní“ slova typu „internetové stránky“, „největší výběr“, apod. Důležitá, tedy klíčová slova, patří na začátek titulku a je vhodné je použít vícekrát v různých mluvnických tvarech, v prvním pádu nebo v pádech častých vazeb. Jednoduchý postup pro zadání odkazu na dokument (website) do katalogu je:

- Nalezení postupným procházením katalogu tématické sekce.
- Nalezení odkazu „Přidat odkaz“ nebo podobného.
- Vyplnění formuláře, minimálně:
    - url optimalizovaného dokumentu (website) provozovatele
    - titulek odkazu
    - popis odkazu
    - kontaktní email
    - kontaktní osoba
- Po určitém čase kontrola, zda byl příslušný odkaz přidán

Řada uživatelů provozuje velmi kvalitní website se spoustou zajímavých podstránek a dokumentů různých oborů, ale mylně se domnívají, že katalog automaticky zařadí všechny stránky jejich website. Takový přístup platí pro fulltextové vyhledávací roboty, nikoli pro katalogy. Jediným možným postupem pro zařazení tématických podstránek website je jejich ruční přidání do katalogu, které však může být degradováno administrátory katalogu obávajících se podvodů.

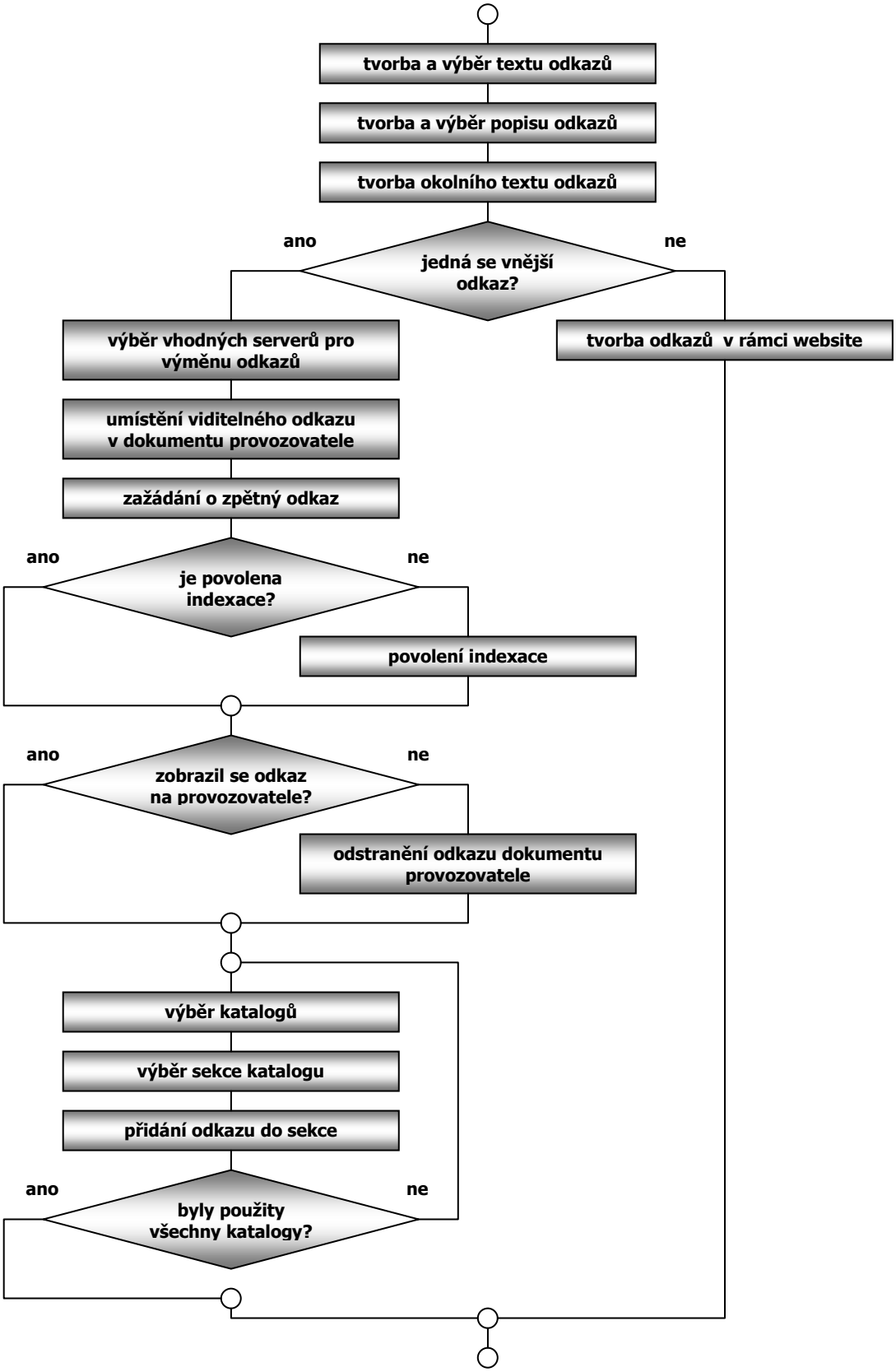


Schéma 8: Metodický postup pro tvorbu zpětných odkazů (zajištění link popularity).

9.7 Analýza efektivity

Každý provozovatel internetové prezentace by se měl zajímat o to, zda a jak dobře plní svůj účel. Potřebné informace lze nejlépe získat analýzou dat o počtu, cílech a činnostech návštěvníků website. Existují obvykle tři možnosti, jak toho dosáhnout. Zpracováním serverových logů, nebo nasazením některého z měřících systémů, např. Google Analytics, nebo využitím svého vlastního měřicího systému.

Pro vyhodnocení efektivity vynaložených prostředků na optimalizaci dokumentu (website) je důležité sledovat následující klíčové údaje:

- počet návštěv a zhlédnutých dokumentů
- poměr nových návštěvníků vůči těm, kteří se na website vracejí
- geografické rozdělení návštěvníků
- odkazující websites (zdroje), odkud návštěvníci přicházejí
- počet konverzí[115]
- klíčová slova, přes která návštěvníci přišli z vyhledávačů
- které dokumenty v rámci website jsou nejnavštěvovanější
- kterým dokumentem návštěvníci website přicházejí
- kterým dokumentem návštěvníci website opouštějí
- chování návštěvníků na celém website (způsob, jakým website procházejí)

Samotný počet návštěvníků informuje detailněji o unikátních přístupech na website za dané období, o průměrných přístupech na website za jednotlivá kratší období, např. dny a kolik návštěvníků přišlo na website poprvé.

Vedle zjišťování běžných údajů o pohybu a charakteru návštěvníků je měření konverzí nejdůležitějším údajem, neboť provozovatelovi website napoví, jak je na tom jeho prezentace z hlediska obchodní úspěšnosti. V případě komerčních website konverze nastane, když se z návštěvníka website provozovatele stane zákazník, neboli náklady, které provozovatel website do získání návštěvníka investoval, se promění v zisk.

[115] Konverze určuje poměr celkového počtu návštěvníků a těch, kteří splnili určitý předem definovaný cíl.

Aby bylo možno poznat, že konverze nastala, je nutné určit akci, která znamená uskutečnění konverze. Často jde o zaslání objednávky zboží, může jí být i vyplnění kontaktního formuláře či registrace uživatele. Takováto akce je pak definována jako jeden z cílů. Cesta, která vede k uskutečnění konverze je nazývána pojmem „trychtýř“. K uskutečnění konverze, třeba objednání zboží, se na webu provozovatele zákazníci dostávají různými cestami. Na začátku cesty jich stojí velký počet, ovšem skrz jednotlivé kroky se k objednávce dostane množství daleko menší. Toto zužování počtu klientů během postupu k cíli svým principem připomíná právě již zmiňovaný trychtýř.

Samotné cesty umožní určit jednotlivé kroky vedoucí k cíli a sledovat, během kterých odpadne nejvíce potenciálních zákazníků. Na základě těchto informací pak lze zlepšovat navigační strukturu a použitelnost webu.

Měření konverzí je velmi přínosné detailněji sledovat, tedy:

- počet konverzí za dané časové období
- konverzní poměr za vybraný časový úsek
- statistika „trychtýřů“
- procento návštěvníků, kteří opustí definovanou cestu „trychtýř“ ještě před dosažením cíle
- zpětné sledovaní cesty k danému cíli

Při měření hodnot konverzí existují dva přístupy. První, jednodušší, je v nastavení cíle zadáním zisku, který jeho dosažení přinese. To se hodí, pokud existuje cílů jen několik a jejich hodnota se příliš nemění, například si zákazník objedná čištění oken, čímž provozovatel website vydělá 2 000 Kč. Druhý případ se týká například internetových obchodů, kde je cena zboží proměnlivá a sortiment rozsáhlý, tím pádem se liší i profit, který každá transakce přináší. Řešením je dynamické zadávání hodnoty konverze podle ceny produktu.

Jedním z nejdůležitějších údajů jsou zdroje návštěvníků provozovatelova website, tedy stránky, vyhledávače nebo reklamní kampaně, ze kterých návštěvníci přicházejí včetně kolik stránek uživatel během návštěvy zhlédl, případně které z definovaných cílů splnil.

Správná volba klíčových slov v analýze klíčových slov zajistí dobrou pozici ve vyhledávačích a samotný přísun návštěvníků. Jaká slova přivádějí uživatele vyhledávačů, kolik stránek si prohlédnou, případně které cíle splní je snadno zjistitelné z podrobné analýzy samotných přístupů na website. Díky těmto údajům lze identifikovat klíčová slova, která lze porovnat s důležitosti klíčových slov z analýzy klíčových slov, a na která je vhodné více zaměřit další optimalizaci.

Výsledkem analýzy efektivity je určit, odkud uživatelé na website nebo jeden konkrétní dokument přicházejí, jaké mají preference, na jaká klíčová slova, která klíčová slova jsou nejvyhledávanější a hlavně, která klíčová slova jsou nejziskovější. Pomocí analýzy efektivity lze vyčíslit poměr přínosů provozovatelovi website v podobě příjmů z prodeje produktů a služeb a nákladů na optimalizaci dokumentu a celého website. Zjištěním nejziskovějších a nejvyhledávanějších klíčových slov lze snadno výsledky porovnávat s analýzou klíčových slov, která, v případě jejího kvalitního zpracování, dává podobné výsledky s analýzou efektivity.

V analýze efektivity je rovněž vhodné sledovat pro technické řešení website určité technické údaje v podobě parametrů zobrazovacích zařízení uživatelů (návštěvníků), typů a verzí jejich WWW prohlížečů, typů operačních systémů, preferované jazykové mutaci apod.

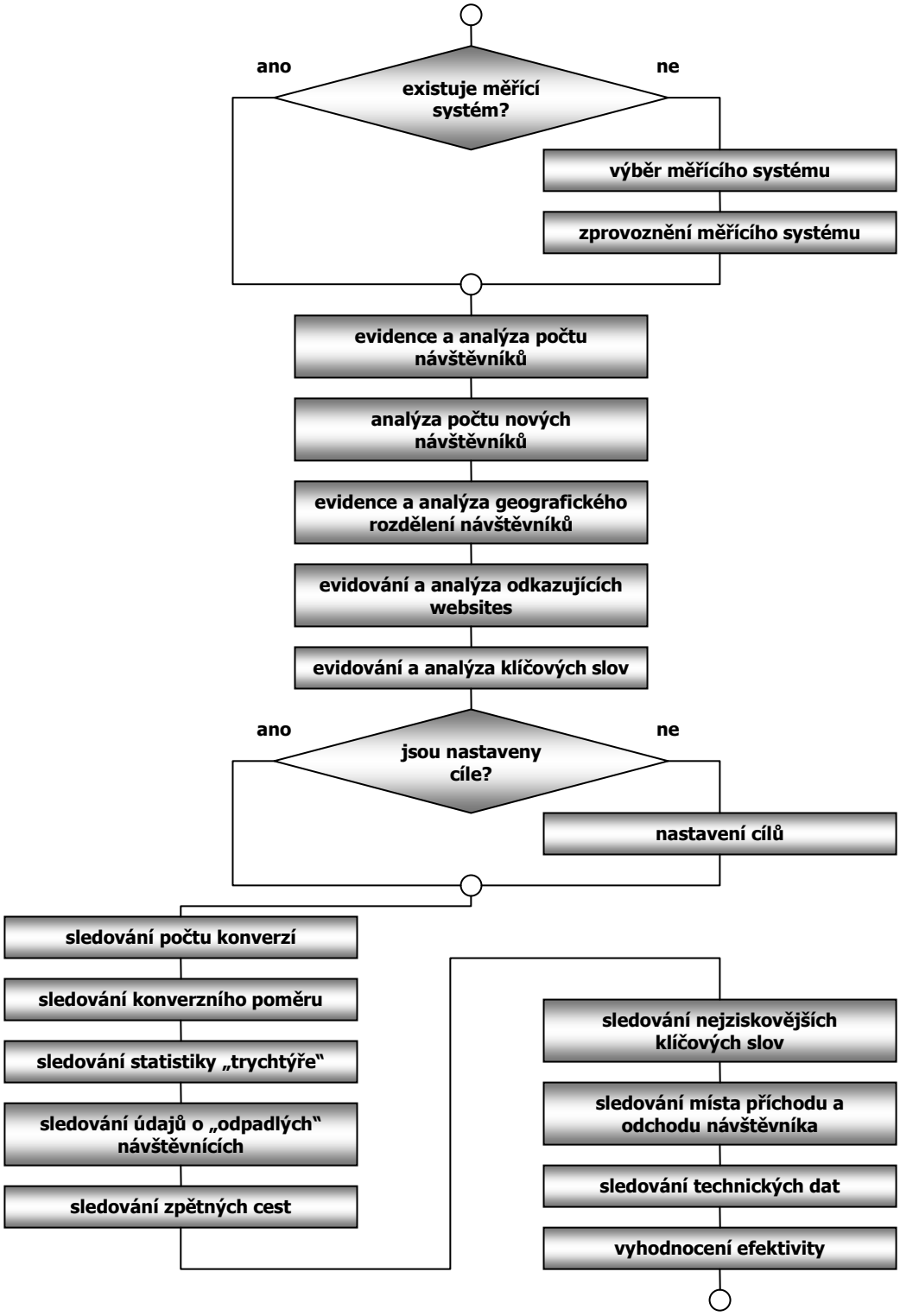


Schéma 9: Metodický postup pro analýzu efektivity.

10 Ověření metodického postupu optimalizace dokumentu pro WWW vyhledávače

Metodický postup optimalizace dokumentu pro WWW vyhledávače byl průběžně aplikován na několika WWW projektech, zejména na:

- WWW prezentaci Veteran Car Clubu Dvůr Králové nad Labem (`www.vcc.cz`)
- WWW prezentaci stěhovací firmy Čtvrtečka (`www.ctvrtecka.cz`)
- WWW prezentaci stavařské firmy Holiday-Pacific Homes-Bohemia (`www.holiday-pacific.cz`)
- WWW prezentaci projektu Žirafy a antilopy (`www.zirafy-antilopy.cz`)
- WWW prezentaci Města Dvůr Králové nad Labem (`www.mudk.cz`)
- WWW prezentaci firmy La Linea (`www.lalinea.cz`)

Výsledky jednotlivých optimalizací byly v zásadě srovnatelné. Z důvodu rozsahu disertační práce jsou podrobněji uvedeny výsledky metodického postupu optimalizace dokumentu prezentace Veteran Car Clubu Dvůr Králové nad Labem a stěhovací firmy Čtvrtečka.

10.1 Aplikace MPOD na WWW prezentaci VCC Dvůr Králové nad Labem

Veteran Car Club Dvůr Králové nad Labem je samostatným a dobrovolným občanským sdružením zájemců v oblasti historie motorismu v Asociaci Veteran Car Clubů AČR[116]. Předmětem a cílem činnosti VCC Dvůr Králové nad Labem je zejména:

- rozvoj veteránského hnutí jako celku se zvláštním zřetelem na zastupování svých členů při prosazování jejich společných zájmů a potřeb
- uchování památek dokládajících vývoj motorismu tj. sběratelství historických vozidel, literatury a s motorismem spojených předmětů a doplňků
- testování vozidel na historickou původnost dle příslušných předpisů
- pořádání veteránských sportovních a společenských akcí

[116] Autoklub České republiky

- reprezentace klubu doma i v zahraničí

Hlavním zdrojem příjmů občanského sdružení jsou vedle členských příspěvků poplatky za realizované testování historických vozidel. Oficielní WWW prezentace Veteran Car Clubu Dvůr Králové nad Labem byla realizována na adrese `http://vcc.nanamesti.cz` v lednu roku 2005, ovšem nepřinášela téměř žádný efekt. V červnu 2006 byla celá WWW prezentace přepracována dle metodického postupu optimalizace dokumentu uveřejněného v disertační práci ze zaměřením na fulltextový vyhledávací stroj Google a přesunuta na novou adresu.

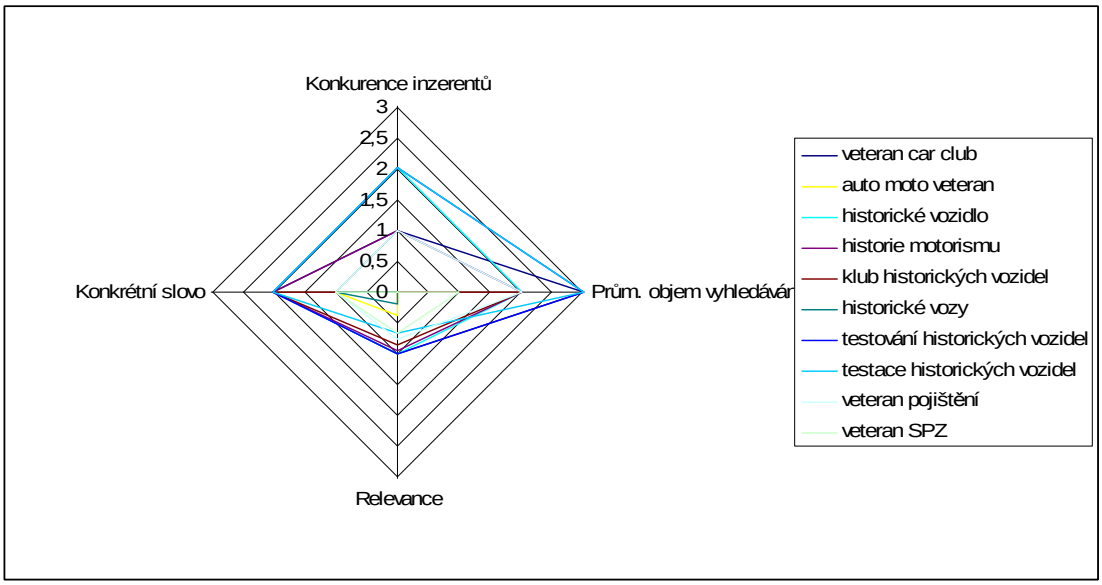
10.1.1 Analýza klíčových slov

Primární klíčová slova a fráze po dlouhé diskuzi byly zvoleny *testování historických vozidel*, *historické vozidlo* a *veteran car club*, ke kterým byla vybrána některá další vhodná slova tak, že výchozí tabulka pro analýzu klíčových slov vypadala následovně:

Klíčové slovo	Konkurence inzerentů	Prům. objem vyhledávání	Relevance	Konkrétní slovo
veteran car club	1	3	1	2
auto moto veteran	0	0	0,372093023	1
historické vozidlo	2	2	1	2
historie motorismu	1	2	0,95	2
klub historických vozidel	0	2	0,85	2
historické vozy	0	0	0,2	1
testování historických vozidel	2	3	1	2
testace historických vozidel	2	3	0,666666667	2
veteran pojištění	1	2	0,75	1
veteran SPZ	0	1	0,65	1

Tabulka 7: Výchozí tabulka pro analýzu klíčových slov Veteran Car Clubu.

Jelikož nebylo možné seřadit jednotlivá klíčová slova dle kritérií, byly použity pro seřazení důležitosti klíčových slov metody pro vícekriteriální analýzy variant.



Graf 14: Grafické znázornění výchozích variant pro analýzu klíčových slov (Veteran Car Club Dvůr Králové nad Labem.

Metodou váženého součtu a metodou TOPSIS bylo dosaženo shodného pořadí a důležitosti jednotlivých klíčových slov.

Metoda váženého součtu		Metoda TOPSIS	
Klíčové slovo	**užitek**	**Klíčové slovo**	**c_i**
testování historických vozidel	1	testování historických vozidel	1
testace historických vozidel	0,953636755	testace historických vozidel	0,95822136
veteran car club	0,86752736	veteran car club	0,808675565
historické vozidlo	0,802521448	historické vozidlo	0,714964637
historie motorismu	0,663094321	historie motorismu	0,634684353
veteran pojištění	0,603929099	veteran pojištění	0,632141204
klub historických vozidel	0,516712707	klub historických vozidel	0,531704849
veteran SPZ	0,260068933	veteran SPZ	0,290570142
auto moto veteran	0,023936373	auto moto veteran	0,021981897
historické vozy	0	historické vozy	0

Tabulka 8: Výsledná pořadí klíčových slov (Veteran Car Club Dvůr Králové nad Labem).

10.1.2 Analýza konkurence

Před samotným užitím metodického postupu optimalizace dokumentu pro WWW vyhledávače nefiguroval žádný odkaz na Veteran Car Club Dvůr Králové nad Labem v prvních třiceti výsledcích vyhledávání na zadaná klíčová slova. Z důvodu omezeného rozsahu disertační práce jsou uvedeny výsledky analýzy konkurence v podobě konkurenčních hodnot dokumentů v rámci klíčových slov a doménového jména.

testovaní historických vozidel

c_i 1

o	CVW	URL
1	1	www.akhv.cz/download/Zmena_zakona.doc
2	0,5	www.sweb.cz/podjestedsky.klub/testace.htm
3	0,333333333	www.e-law.cz/zakony/240_02.html
4	0,25	www.mdcr.cz/NR/rdonlyres/D220782B-C0D1-4D0E-8F73-55139EB4B8AA/0/v24002.rtf
5	0,2	info.opelclassiccars.cz/79b.pdf
6	0,166666667	http://www.brno.cz/index.php?nav01=1248&nav02=2227&nav03=70&nav04=1007&nav05=1018&nav06=2551&nav07=2553&nav08=10087
7	0,142857143	http://www.podnikame.cz/zakony02/index.php3?co=Z2002240
8	0,125	www.mmkv.cz/dokumenty/CoDelat/poskytovani_informaci/OD/13.OD-DSA.doc
9	0,111111111	www.mmkv.cz/dokumenty/CoDelat/poskytovani_informaci/OD/15.OD-DSA.doc
10	0,1	http://www.vccpraha.cz/main.php?p=YWN0PWRldGFpbCZpZD0xNCZ0PTExMzU4ODY0ODI

Tabulka 9: CVD za klíčová slova testování historických vozidel.

CWD_i	Doménové jméno
1	www.akhv.cz
0,5	www.sweb.cz/podjestedsky.klub
0,3333333	www.e-law.cz
0,25	www.mdrcr.cy
0,2	info.opelclassiccars.cz
0,1666667	www.brno.cz
0,1428571	www.podnikame.cz
0,2361111	www.mmkv.cz
0,1	www.vccpraha.cz

Tabulka 10: CVWi za klíčová slova testování historických vozidel.

testace historických vozidel

c_i 0,95822136

o	CVW	URL
1	0,95822136	www.eurooldtimers.com/
2	0,47911068	www.kropi.webz.cz/
3	0,31940712	www.kropi.webz.cz/index_testace.htm
4	0,23955534	www.sweb.cz/podjestedsky.klub/testace.htm
5	0,191644272	www.sweb.cz/podjestedsky.klub/historie.htm
6	0,15970356	http://forum.predseda.com/viewtopic.php?p=1638&sid=17ec767922132c6a1a190bfde86928c9
7	0,136888766	www.autocar.cz/moravsky/testace/testace.htm
8	0,11977767	www.hicos.cz/vkv/page4.html
9	0,10646904	www.militarycarclub.cz/index.php?page=testace_hv
10	0,095822136	www.veteranklub.wz.cz/index_klub_historie.html

Tabulka 11: CVD za klíčová slova testace historických vozidel.

CWD_i	Doménové jméno
0,95822136	www.eurooldtimers.com
0,7985178	www.kropi.webz.cz
0,431199612	www.sweb.cz/podjestedsky.klub
0,15970356	forum.predseda.com
0,136888766	www.autocar.cz
0,11977767	www.hicos.cz
0,10646904	www.militarycarclub.cz
0,095822136	www.veteranklub.wz.cz

Tabulka 12: CVWi za klíčová slova testace historických vozidel.

veteran car club

c_i 0,808675565

o	CVW	URL
1	0,808675565	www.vccofgb.co.uk/
2	0,404337782	www.vccpraha.cz/
3	0,269558522	vcctrebic.info/
4	0,202168891	www.vccsturm.com/
5	0,161735113	www.veterancarclubcb.cz/
6	0,134779261	www.veteran-ova.cz/
7	0,115525081	www.vcc-strakonice.cz/
8	0,101084446	www.vcckaplice.cz/
9	0,089852841	www.vccansw.org/
10	0,080867556	www.volny.cz/veteran-pi/

Tabulka 13: CVD za klíčová slova veteran car club.

CWD_i	Doménové jméno
0,808675565	www.vccofgb.co.uk
0,404337782	www.vccpraha.cz
0,269558522	www.vcctrebic.info
0,202168891	www.vccsturm.com/
0,161735113	www.veterancarclubcb.cz/
0,134779261	www.veteran-ova.cz/
0,115525081	www.vcc-strakonice.cz/
0,101084446	www.vcckaplice.cz/
0,089852841	www.vccansw.org/
0,080867556	www.volny.cz/veteran-pi/

Tabulka 14: CVWi za klíčová slova veteran car club.

historické vozidlo

c_i 0,714964637

o	CVW	URL
1	0,714964637	www.rozhlas.cz/leonardo/technologie/_zprava/375694
2	0,357482319	muzeum.wz.cz/
3	0,238321546	www.brno.cz/index.php?nav01=1248&nav02=2227&nav03=70&nav04=1007&nav05=1018&nav06=2551
4	0,178741159	www.fkhv.cz/zobraz.php?co=aktuality
5	0,142992927	www.sweb.cz/podjestedsky.klub/
6	0,119160773	www.oldtimerauto.cz/
7	0,102137805	www.e-law.cz/zakony/240_02.html
8	0,08937058	www.akhv.cz/download/Zmena_zakona.doc
9	0,079440515	www.autoklub.cz/acr/veterancc/2006/poplatky/pdf/novezpoplatnenespravniukonypriregistraci.pdf
10	0,071496464	www.akhv.cz/download/Zmena_zakona.doc

Tabulka 15: CVD za klíčová slova historické vozidlo.

CWD_i	Doménové jméno
0,714964637	www.rozhlas.cz
0,357482319	muzeum.wz.cz
0,238321546	www.brno.cz
0,178741159	www.fkhv.cz
0,142992927	www.sweb.cz/podjestedsky.klub
0,119160773	www.oldtimerauto.cz
0,102137805	www.e-law.cz
0,160867043	www.akhv.cz
0,079440515	www.autoklub.cz

Tabulka 16: CVWi za klíčová slova historické vozidlo.

historie motorismu

c_i 0,634684353

o	CVW	URL
1	0,634684353	www.infoway.cz/artc.asp?id=19
2	0,317342177	www.abchistory.cz/dejiny-auto-moto.htm
3	0,211561451	www.f1-info.cz/?document=sezony/sezona0000/sezona0000&index=02
4	0,158671088	www.rozhlas.cz/strednicechy/aktualne/_zprava/359943
5	0,126936871	www.porsche-club.cz/report_archiv.php3?sc_id=43&range=0&show=10 - 23k
6	0,105780726	www.motomagazin.cz/
7	0,090669193	www.automuzeum.cz/
8	0,079335544	auto-moto.ihned.cz/c1-21611980-pro-milovniky-historickych-vozidel-se-kona-retro-prague-2007
9	0,070520484	www.motorjournal.cz/2001/12/mj12au.htm
10	0,063468435	www.jetotam.cz/?treeid=293

Tabulka 17: CVD za klíčová slova historie motorismu.

CWD_i	Doménové jméno
0,634684353	www.infoway.cz
0,317342177	www.abchistory.cz
0,211561451	www.f1-info.cz
0,158671088	www.rozhlas.cz
0,126936871	www.porsche-club.cz
0,105780726	www.motomagazin.cz
0,090669193	www.automuzeum.cz
0,079335544	auto-moto.ihned.cz
0,070520484	www.motorjournal.cz
0,063468435	www.jetotam.cz

Tabulka 18: CVWi za klíčová slova historie motorismu.

veteran pojištění

c_i 0,632141204

o	CVW	URL
1	0,634684353	www.svetveteranu.cz/pojisteni.html
2	0,317342177	www.pojistenicz.com/page.php?id=kradez
3	0,211561451	www.vccpraha.cz/main.php?p=YWN0PWRldGFpbCZpZD0xNCZ0PTExMzU4ODY0ODI
4	0,158671088	fincentrum.idnes.cz/ruceni.asp?y=ruceni/ruceni03.htm
5	0,126936871	www.autoklub.cz/show.php?page=acr/infoprokluby/2006/pojisteni/index.htm&asoc=7
6	0,105780726	www.autoklub.cz/show.php?page=acr/infoprokluby/2003/pojisteni/postup.htm&asoc=7
7	0,090669193	www.veteranbrno.cz/sobesice/prop.html
8	0,079335544	autobazar.sbazar.cz/auta-veterani/praha/prodam-koupim-aukce.html
9	0,070520484	doprava.impuls.cz/osmnact-let-stare-auto-neni-veteran-proto--/article.html?id=515
10	0,063468435	www.army.cz/images/id_7001_8000/7109/rozkaz.doc

Tabulka 19: CVD za klíčová slova veteran pojištění.

CWD_i	Doménové jméno
0,634684353	www.svetveteranu.cz
0,317342177	www.pojistenicz.com
0,211561451	www.vccpraha.cz
0,158671088	fincentrum.idnes.cz
0,232717596	www.autoklub.cz
0,105780726	www.veteranbrno.cz
0,090669193	autobazar.sbazar.cz
0,079335544	doprava.impuls.cz
0,070520484	www.army.cz

Tabulka 20: CVWi za klíčová slova veteran pojištění.

klub historických vozidel

c_i 0,531704849

o	CVW	URL
1	0,531704849	www.akhv.cz/index.php?p=kluby
2	0,265852425	www.akhv.cz/
3	0,17723495	www.sweb.cz/podjestedsky.klub/
4	0,132926212	www.sweb.cz/podjestedsky.klub/testace.htm
5	0,10634097	www.khvopava.cz/
6	0,088617475	www.fkhv.cz/zobraz.php?co=aktuality
7	0,075957836	www.firmy.cz/detail/406058-podjestedsky-klub-historickych-vozidel-v-acr-chrastava.html
8	0,066463106	www.kphv.cz
9	0,059078317	www.eurooldtimers.com/
10	0,053170485	www.automuzeum.cz/

Tabulka 21: CVD za klíčová slova klub historických vozidel.

CWD_i	Doménové jméno
0,797557274	www.akhv.cz
0,386118998	www.sweb.cz/podjestedsky.klub
0,10634097	www.khvopava.cz
0,088617475	www.fkhv.cz
0,066463106	www.kphv.cz
0,059078317	www.eurooldtimers.com
0,053170485	www.automuzeum.cz

Tabulka 22: CVWi za klíčová slova klub historických vozidel.

CWD	Doménové jméno	CWD	Doménové jméno
1,958424317	www.akhv.cz	0,158671088	fincentrum.idnes.cz
1,460311537	www.sweb.cz/podjestedsky.klub	0,143839678	www.automuzeum.cz
1,017299677	www.eurooldtimers.com	0,142857143	www.podnikame.cz
0,873635725	www.rozhlas.cz	0,136888766	www.autocar.cz
0,808675565	www.vccofgb.co.uk	0,134779261	www.veteran-ova.cz/
0,7985178	www.kropi.webz.cz	0,126936871	www.porsche-club.cz
0,715899233	www.vccpraha.cz	0,11977767	www.hicos.cz
0,634684353	www.infoway.cz	0,119160773	www.oldtimerauto.cz
0,634684353	www.svetveteranu.cz	0,115525081	www.vcc-strakonice.cz/
0,435471139	www.e-law.cz	0,10646904	www.militarycarclub.cz
0,404988212	www.brno.cz	0,10634097	www.khvopava.cz
0,357482319	muzeum.wz.cz	0,105780726	www.motomagazin.cz
0,317342177	www.abchistory.cz	0,105780726	www.veteranbrno.cz
0,317342177	www.pojistenicz.com	0,101084446	www.vcckaplice.cz/
0,312158111	www.autoklub.cz	0,095822136	www.veteranklub.wz.cz
0,269558522	www.vcctrebic.info	0,090669193	autobazar.sbazar.cz
0,267358634	www.fkhv.cz	0,089852841	www.vccansw.org/
0,25	www.mdrcr.cy	0,080867556	www.volny.cz/veteran-pi/
0,236111111	www.mmkv.cz	0,079335544	auto-moto.ihned.cz
0,211561451	www.f1-info.cz	0,079335544	doprava.impuls.cz
0,202168891	www.vccsturm.com/	0,070520484	www.army.cz
0,2	info.opelclassiccars.cz	0,070520484	www.motorjournal.cz
0,161735113	www.veterancarclubcb.cz/	0,066463106	www.kphv.cz
0,15970356	forum.predseda.com	0,063468435	www.jetotam.cz

Tabulka 23: CVW za všechna klíčová slova.

10.1.3 Analýza stávajícího stavu

Původní WWW prezentace nekorespondovala se žádnými již dříve publikovanými doporučeními optimalizace WWW stránek pro vyhledávací roboty. Jejich konkurenční hodnota na klíčová slova získaná z analýzy klíčových slov byla téměř nulová (jak jednotlivých dokumentech, tak celého website). Pokud se nějaký dokument z website Veteran Car Clubu Dvůr Králové nad Labem objevil na prvních třech stránkách výsledků vyhledávání, bylo to vždy na specifické, přesně definované, fráze.

Podrobnější analýza stávajícího stavu byla tedy vynechána a byl vydána požadavek na copywriting nových textů, které budou zveřejněny na nové adrese.

10.1.4 MPOD copywriting

Základním vodítkem pro samotnou tvorbu textů, které měly být zveřejněny v podobě oficielní WWW prezentace Veteran Car Club Dvůr Králové nad Labem, byly výsledky z analýzy klíčových slov. Samotná analýza klíčových slov určila, jaká bude pozice klíčových slov v textu a při jaké četnosti (hustotě).

Je známo, že text musí být souvislý a členěn do nadpisů a odstavců, nepřipadá tedy v úvahu pouhý výčet klíčových slov v určitém pořadí. Tvorbou textů byly pověřeni samotní členové Veteran Car Clubu, kterým byly přesně definovány požadavky pro tvorbu textů, zejména dodržení pořadí a hustoty klíčových slov.

10.1.5 Optimalizace dokumentu a serveru

Jelikož se jednalo o menší projekt a samotná WWW prezentace byla komplexně přepracována, bylo k dokumentům přistupováno globálním pohledem, tedy ne optimalizací každého dokumentu zvlášť, ale optimalizací celého website.

Prvním krokem byla volba nových doménových jmen `veteran-car-club.cz` a `vcc.cz` a optimalizace WWW serveru na `mod_rewrite`. V úvahu připadla i registrace doménového jména `testovani-historickych-vozidel.cz`, ale z důvodů penalizace duplicitního obsahu vyhledávacími stroji bylo od ní upuštěno.

Samotná tvorba WWW prezentace byla v souladu s navrženým metodickým postupem optimalizace dokumentu pro WWW vyhledávače a vycházela z analýzy klíčových slov. Mimo jiné se jednalo zejména o uvedení klíčových slov v titulcích dokumentů, v nadpisech textu, zvýrazněném textu, apod. Byl také brán zřetel na zamezení shodného obsahu titulku dokumentu a nadpisu nejvyšší úrovně.

Bylo upuštěno od formátování vzhledu HTML značkami a veškerý vzhled WWW prezentace byl zpracován kaskádovými styly. Takovýto přístup umožnil zoptimalizování zdrojového kódu XHTML dokumentů. Jelikož původní dokumenty měly mizivou hodnotu GTPR, nebylo na původní adrese umístěno trvalé přesměrování na novou WWW prezentaci Veteran Car Clubu Dvůr Králové nad Labem.

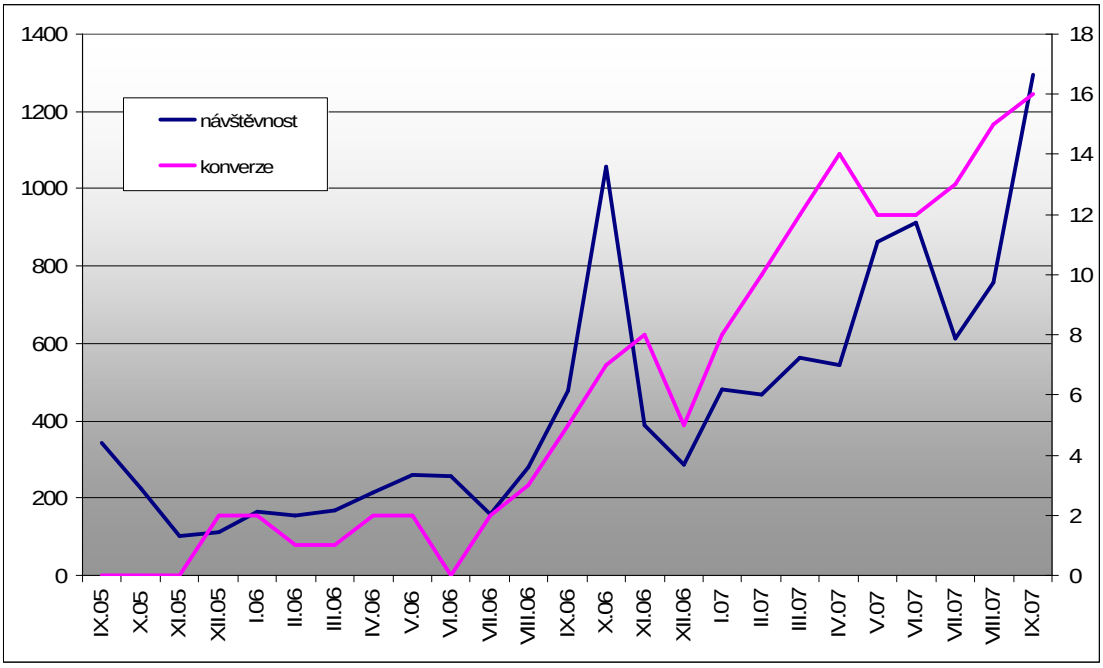
10.1.6 Registrace do katalogů a tvorba zpětných odkazů

Posledním krokem optimalizace dokumentů bylo komplexní zrevidování odkazů vedoucích na původní website Veteran Car Clubu ve Dvoře Králové nad Labem. U stávajících odkazů v sekcích katalogů došlo ke sjednocení struktury odkazů a k registraci nových. Jednalo se zejména o katalogové sekce portálů `Seznam.cz`, `Centrum.cz`, `Atlas.cz`, `Tiscali.cz` a `Caramba.cz`.

Stejný postup byl uskutečněn i u tvorby zpětných odkazů na serverech s podobnou tématikou, např. `i-veteran.cz` a naopak, odkazy z obsahově odlišných serverů byly smazány.

10.1.7 Analýza efektivity

Od srpna 2006, krátce po spuštění nové WWW prezentace občanského sdružení Veteran Car Club byl zaznamenán prudký nárůst návštěvnosti, který byl zapříčiněn téměř žádným stářím samotných dokumentů a doménového jména a pravděpodobně chystanou zářijovou akcí orientační soutěže historických vozidel. Po uklidnění situace v říjnu 2006 však docházelo k postupnému nárůstu návštěvnosti a hlavně k postupnému růstu žádostí o testování historického vozidla.



Graf 15: Vývoj návštěvnosti a žádostí o testování historického vozidla.

O rok později (srpen 2007) bylo posuzováno hodnoceni relevance WWW prezentace na klíčová slova z analýzy klíčových slov vyhledávacím strojem Google. Odkaz na WWW prezentace se vždy objevil mezi prvními třemi desítkami výsledků.

slovo	výsledků	pořadí	úspěšnost	úspěšnost v top 30
testování historických vozidel	596000	1	100,00000000000000%	100%
testace historických vozidel	2070	2	99,95169082125600%	97%
veteran car club	2080000	7	99,99971153846150%	80%
historické vozidlo	1860000	6	99,99973118279570%	83%
historie motorismu	925000	4	99,99967567567570%	90%
veteran pojištění	227 000	9	99,99647577092510%	73%
klub historických vozidel	492000	20	99,99613821138210%	37%

Tabulka 24: Umístění odkazu na website Veteran Car Clubu.

slovo	výsledků	pořadí	úspěšnost	úspěšnost v top 30
"testování historických vozidel"	322	1	100,00000000000000%	100%
"testace historických vozidel"	201	0	0,00000000000000%	0%
"veteran car club"	45200	6	99,98893805309730%	83%
"historické vozidlo"	573	11	98,25479930191970%	67%
"historie motorismu"	989	18	98,28109201213350%	43%
"veteran pojištění"	0	0	0,00000000000000%	0%
"klub historických vozidel"	3890	0	0,00000000000000%	0%

Tabulka 25: Umístění odkazu na website Veteran Car Clubu II.

slovo	výsledků	pořadí	úspěšnost	úspěšnost v top 30
testování historických vozidel testování historických vozidel	506000	1	100,00000000000000%	100%
testace historických vozidel testace historických vozidel	2030	7	99,70443349753690%	80%
veteran car club veteran car club	2300000	3	99,99991304347830%	93%
historické vozidlo historické vozidlo	1810000	12	99,99939226519340%	63%
historie motorismu historie motorismu	812000	19	99,99778325123150%	40%
veteran pojištění veteran pojištění	269000	6	99,99814126394050%	83%
klub historických vozidel klub historických vozidel	469000	18	99,99637526652450%	43%

Tabulka 26: Umístění odkazu na website Veteran Car Clubu III.

CVW	Doménové jméno	CVW	Doménové jméno
1,942988106	www.vcc.cz	0,166666667	info.opelclassiccars.cz
0,95822136	www.eurooldtimers.com	0,161735113	www.veterancarclubcb.cz/
0,841901508	www.rozhlas.cz	0,158671088	fincentrum.idnes.cz
0,827674093	www.sweb.cz/podjestedsky.klub	0,136888766	forum.predseda.com
0,808675565	www.vccofgb.co.uk	0,134779261	www.veteran-ova.cz/
0,634684353	www.infoway.cz	0,125	www.podnikame.cz
0,634684353	www.svetveteranu.cz	0,11977767	www.autocar.cz
0,615899233	www.vccpraha.cz	0,115525081	www.vcc-strakonice.cz/
0,579440515	www.akhv.cz	0,10646904	www.hicos.cz
0,55896246	www.kropi.webz.cz	0,105780726	www.porsche-club.cz
0,381178689	www.brno.cz	0,102137805	www.oldtimerauto.cz
0,357482319	muzeum.wz.cz	0,101084446	www.vcckaplice.cz/
0,33937058	www.e-law.cz	0,095822136	www.militarycarclub.cz
0,317342177	www.abchistory.cz	0,090669193	www.motomagazin.cz
0,317342177	www.pojistenicz.com	0,090669193	www.veteranbrno.cz
0,30421406	www.autoklub.cz	0,089852841	www.vccansw.org/
0,269558522	www.vcctrebic.info	0,079335544	autobazar.sbazar.cz
0,211561451	www.f1-info.cz	0,079335544	www.automuzeum.cz
0,211111111	www.mmkv.cz	0,070520484	auto-moto.ihned.cz
0,202168891	www.vccsturm.com/	0,063468435	doprava.impuls.cz
0,2	www.mdrcr.cy	0,063468435	www.motorjournal.cz
0,178741159	www.fkhv.cz		

Tabulka 27: CVW za všechna klíčová slova po optimalizaci (mimo slovního spojení *klub historických vozidel*).

10.2 Stěhovací firma Čtvrtečka

Stránky stěhovací firmy Jiřího Čtvrtečky (`http://www.ctvrtecka.cz/`) byly vytvořeny již v roce 1998. Ve chvíli, kdy začala jejich návštěvnost poprvé klesat (listopad 2001), přišel od majitele požadavek pokusit se tento propad zastavit. Byla provedena analýza klíčových slov (ještě bez užití technik z metodického postupu optimalizace dokumentu pro WWW vyhledávače), tato slova přidána do kódu stránek a stránky zaregistrovány do několika katalogů a vyhledávačů.

Prvním krokem optimalizace, v té době nekonzistentní, byla analýza vývoje návštěvnosti a klíčových slov. Data byla získána ze statistik získaných z logů webu pomocí programu awstats (Advanced Web Statistics[117]). Identifikovaná klíčová slova (*stěhování*, *malování*, *Čtvrtečka*, *úklid*, *ilona czakova*, *stěhování praha*, *úklid*) byla vložena do záhlaví stránek. Aby se zvýšilo povědomí o webu a vzrostl počet zpětných odkazů, byly stránky zaregistrovány do katalogů Seznamu (`http://www.seznam.cz/`), Atlasu (`http://atlas.cz/`) a Centra (`http://katalog.centrum.cz/`).

[117]) Dostupný zdarma na adrese http://awstats.sourceforge.net/

O několik let později (na konci roku 2004) začala návštěvnost opět klesat. Proto byla provedena podrobná optimalizace website dle navrhnutého metodického postupu optimalizace dokumentu pro WWW vyhledávače. Pro analýzu klíčových slov byla zvolena primární klíčová slova stěhování a malování. Po odmazání irelevantních klíčových slov jako základ pro analýzu klíčových slov bylo získáno 49 klíčových slov, která byla dále zredukována metodami vícekriteriální analýzy variant. Výsledná pořadí důležitosti klíčových slov stanovená metodou váženého součtu a metodou TOPSIS byla téměř shodná.

Klíčové slovo	**ci**	**Klíčové slovo**	**užitek**
stěhování	1	stěhování	1
malování	0,799690114	malování	0,911684907
stěhovací služby	0,336445531	malování pokojů	0,438090101
malování pokojů	0,336431606	malování bytu	0,433609358
malování bytu	0,336089894	stěhování praha	0,432941177
stěhování praha	0,33603976	stěhování brno	0,427045462
stěhování brno	0,335606733	malování bytů	0,4224075
malování bytů	0,335277975	malování stěn	0,42166071
malování stěn	0,335226027	stěhování ostrava	0,421149748
stěhování ostrava	0,335190642	stěhovací služby	0,407489617
malování brno	0,334062773	malování brno	0,403737737
malování pokoje	0,333884708	malování pokoje	0,400750575
barvy na malování	0,333378104	barvy na malování	0,391789089
malovat	0,332839988	malovat	0,381334021
malování na zeď	0,332558386	malování na zeď	0,375359697
maluje	0,331142307	maluje	0,334286218
stěhování kanceláří	0,068837229	stěhování kanceláří	0,139671206
stěhování nábytku	0,061867965	stěhování nábytku	0,127879777
levné stěhování	0,058346267	levné stěhování	0,121984062
stěhování zlín	0,056576302	stěhování zlín	0,119036205
malování interiérů	0,054586824	malování interiérů	0,115734604
stěhování pian	0,053018452	stěhování pian	0,11314049
autodoprava stěhování	0,051230728	autodoprava stěhování	0,110192633
stěhování olomouc	0,047638201	stěhování olomouc	0,104296918
stěhování hradec králové	0,045833665	stěhování hradec králové	0,101349061
stěhování bytů	0,044023836	stěhování bytů	0,098401203
stěhování liberec	0,042208922	stěhování liberec	0,095453346
malování na stěnu	0,041821105	malování na stěnu	0,09482447
stěhování plzeň	0,040389172	stěhování plzeň	0,092505489
stěhování klavíru	0,038564887	stěhování klavíru	0,089557631
stěhování klavírů	0,036736438	stěhování klavírů	0,086609774
stěhování pianin	0,033068953	stěhování pianin	0,080714059
malování válečkem	0,031623399	malování válečkem	0,078395078
potřeby pro malování	0,029759766	potřeby pro malování	0,075407916
stěhování strojů	0,029391818	stěhování strojů	0,074818345
stěhování strakonice	0,02755201	stěhování strakonice	0,071870487
stěhování české budějovice	0,025713207	stěhování české budějovice	0,06892263
stěhování pianina	0,02387733	stěhování pianina	0,065974773
váleček na malování	0,021852164	váleček na malování	0,062712477
stěhování piána	0,01841808	stěhování piána	0,057131201
malování bytu barevné kombinace	0,016820804	malování bytu barevné kombinace	0,054497781

Tabulka 28: Výsledná pořadí důležitosti klíčových slov.

Jelikož se žádný dokument z WWW prezentace stěhovací firmy Čtvrtečka nezobrazoval na žádné z první tří stránek výsledků vyhledávání, není z důvodů rozsahu disertační práce uvedena analýza konkurence a analýza stávajícího stavu. Nicméně pro výpočet konkurenční hodnoty dokumentu, konkurenční hodnoty website v rámci klíčového slova i konkurenční hodnoty website byly použity metody uvedené v kapitole „Analýza konkurence“.

Z důvodu předpokladu, že čím blíže je slovo od začátku WWW stránky, tím má pro fulltextové vyhledávače větší důležitost, byl komplexně přepracován samotný obsah každého dokumentu s klíčovými slovy a při redesignu stránek byly dodržovány všechny zásady tvorby XHTML dokumentů pomocí validního XHTML kódu a kaskádových stylů.

```
<body>
  <div id="main">
    <div id="content">
      <h1><a name="top"></a>Naše firma Vám nabízí</h1>
      <ul>
        <li>Bezplatné prohlídky a konzultace</li>
        <li>Zapůjčení stěhovacích přepravek zdarma</li>
        <li>Termínové slevy</li>
        <li>Stěhování po celé ČR</li>
        <li>Stěhování ČR + EVROPA</li>
        <li>Mezinárodní stěhování</li>
        <li>Celní služby</li>
        <li>Stěhování všeho druhu</li>
        <li>Stěhování bytů a kanceláří</li>
        <li>Stěhování archivů a knihoven</li>
        <li>Stěhování skladů</li>
        <li>Stěhování těžkých břemen</li>
        <li>Stěhování kancelářské techniky</li>
        <li>Balení stěhovaných věcí a nábytku</li>
        <li>Demontáž a montáž nábytku</li>
        <li>Prodej balícího materiálu</li>
        <li>Pronájem skladových prostor</li>
        <li><strong>Kompletní servis na klíč</strong></li>
        <li>Malířské práce</li>
        <li>Lakýrnické práce</li>
        <li>Úklidové práce</li>
        <li>Likvidaci komunálního odpadu</li>
        <li>Skartace písemností</li>
        <li>Skladování</li>
      </ul>
      <p><a id="tothetop" href="#top"> </a></p>
    </div>
    <div id="menu">
      <p><strong>Úvod</strong><span> | </span>
      <a href="sluzby.html">Přehled služeb</a><span> | </span>
      <a href="historie.html">Historie společnosti</a><span> | </span>
      <a href="park.html">Vozový park</a><span> | </span>
      <a href="obaly.html">Obalový materiál</a><span> | </span>
      <a href="cenik.html">Ceník služeb</a><span> | </span>
      <a href="pojisten.html">Pojištění a záruky</a><span> | </span>
      <a href="referenc.html">Reference</a><span> | </span>
      <a href="poptavka.html">Nezávazná poptávka</a><span> | </span>
      <a href="kontakt.html">Kontakt</a></p>
        <img src="i/menu.jpg" width="191" height="171" />
    </div>
    <p id="copyright">Copyright &copy;1999 - 2005 Čtvrtečka</p>
  </div>
</body>
```

Příklad 3: XHTML kód tagu body hlavní stránky webu ctvrtecka.cz.

Rozmístění jednotlivých prvků na stránce je zajištěno pomocí kaskádových stylů – záhlaví je řešeno grafickým pozadím stránky a hlavní menu, byť se v kódu stránky nalézá až za samotným obsahem, je umístěno do levého horního rohu pomocí absolutního pozicování.

Vzhledem k tomu, že vedení firmy Čtvrtečka odmítlo propagaci jakoukoliv formou placených odkazů, jsou dosažené úspěchy čistě výsledkem metodického postupu optimalizace dokumentu pro vyhledávače. Zatímco před registrací o stránkách nikdo nevěděl, v prosinci 2006 již bylo možné najít odkaz na stránky firmy v mnoha katalozích.

Seznam Firmy.cz – sekce Stěhovací služby `http://dir.seznam.cz/Vse-pro-firmy/Sluzby-pro-firmy/Stehovaci-sluzby/`
Katalog Atlas – sekce Malířství a natěračství `http://katalog.atlas.cz/obchody_a_sluzby/remeslne_prace_a_drobna_vyroba/malirstvi_a_nateracstvi/`
Katalog Služby – sekce Stěhovací společnosti `http://www.sluzby.kvalitne.cz/stehovani/stehovani.htm`
Firmy Centrum.cz – sekce Stěhovací služby `http://search.centrum.cz/s-5604009-stehovaci-sluzby`
České obchody – sekce Zahrada, řemesla `http://zahrada.ceskeobchody.com/`
EUkatalog.cz `http://www.eucatalog.cz/search.php?cat=51930`
Katalog klikni.cz - sekce Úklidové práce `http://klikni.idnes.cz/Sluzby/Uklidove-prace/index.htm`
Hyperinzerce, sekce Stěhování `http://sluzby.hyperinzerce.cz/stehovani/`
Katalog Začátek – sekce Doprava nákladní silniční `http://www.zacatek.cz/katalog/firmy/sluzby/cestovani/doprava/nakladni-silnicni/?abc=c`
Stavíme – bydlíme, sekce Stavební řemesla `http://www.stavime-bydlime.cz/index.php?ID=1442`
Katalog OnTheWeb – sekce Stěhovací služby `http://ontheweb.cz/katalog/obchody-sluzby/doprava-logistika/stehovaci-sluzby/`

Tabulka 29: Seznam odkazů na ctvrtecka.cz.

Zlepšení pozice ve vyhledávačích může být jedním z důvodů nárůstu návštěvnosti. Jak vyplývá ze statistik návštěvnosti zpracovaných z webových logů, v roce 1999 byl poslán na stránky 1 návštěvník (vyhledávačem Ask Jeeves), za rok 2006 poslal na stránky Seznam 6 825 návštěvníků, Google 833, Atlas.cz 227,

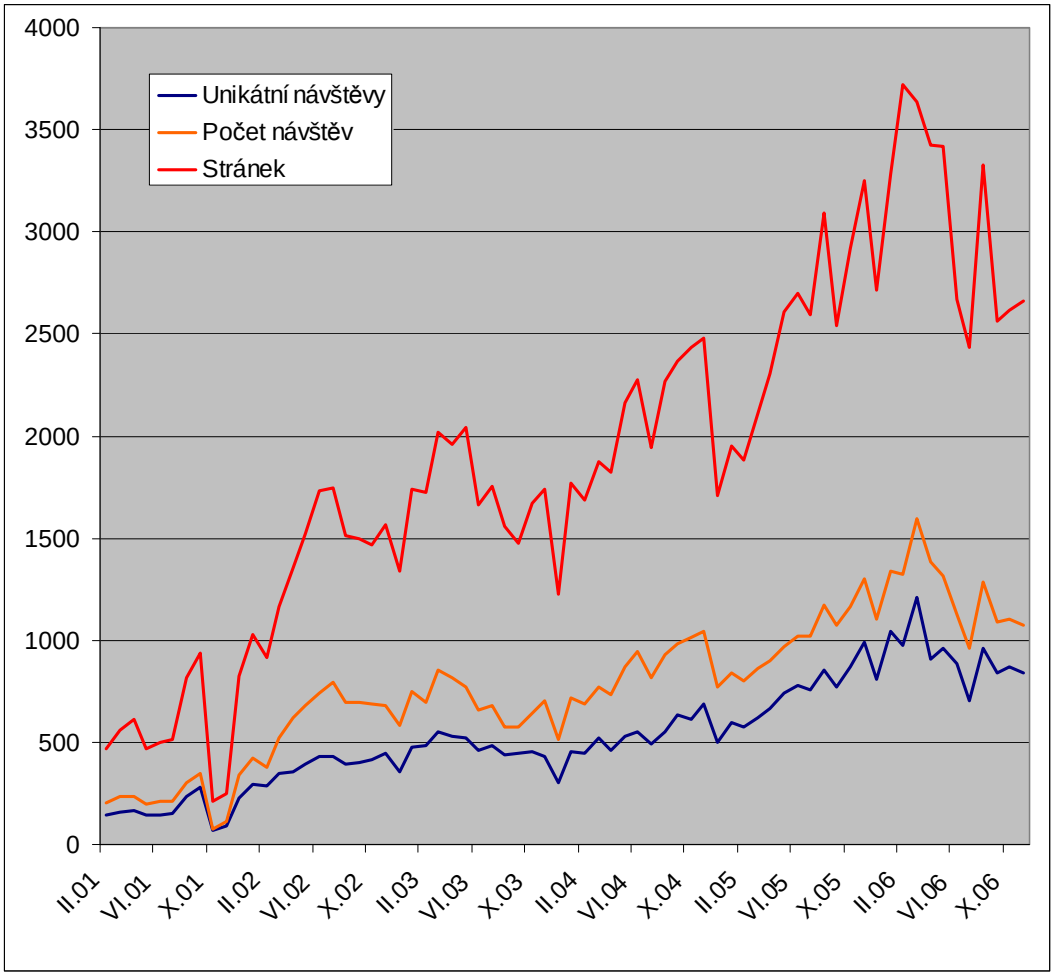
Centrum.cz 128 a ostatní vyhledávače 131 návštěvníků. Nepočítaje 14 682 návštěvníků, kteří přišli přímo a 282 z jiných webů.

Následující tabulka zobrazuje vývoj počtu návštěvníků, kteří přišli na web ctvrtecka.cz skrze nejčastější hledaná klíčová slova. Zatímco v roce 2001 přišlo pomocí klíčových slov na stránky 470 návštěvníků, v roce 2006 to bylo již 6 259 návštěvníků.

Hledaný výraz	**v r. 2001**	**v r. 2006**
stěhování	249x	1644x
malování	31x	793x
stěhovací služby	28x	743x
malování pokojů	25x	721x
malování bytu	25x	672x
stěhování praha	22x	671x
malování bytů	21x	428x
čtvrtečka	20x	363x
ceník	4x	252x
praha	4x	172x
www.ctvrtecka.cz	2x	80x
pokojů	0x	64x
obalová	0x	61x
materiál	0x	59x
jiná slova	160x	2771x

Tabulka 30: Kolik návštěvníků přivedla klíčová slova na ctvrtecka.cz

Úspěch prováděného metodického postupu optimalizace pro WWW vyhledávače je patrný na následujícím grafu návštěvnosti, který odráží obě vlny optimalizace (1. optimalizace byla provedena v lednu 2001 a 2., již dle zmiňovaného metodického postupu optimalizace dokumentu, v lednu 2005).



Graf 16: Vývoj návštěvnosti webu ctvrtecka.cz.

Díky synergickému efektu uvedených metod se podařilo odkaz na web ctvrtecka.cz vyzdvihnout na vyhledávači Google na první stranu výsledků vyhledávání na hledaný výraz stěhování a na první místo celkem na hledaná slova *stěhování malování*. Celkově lze optimalizační kampaň pro web `ctvrtecka.cz` hodnotit jako velmi úspěšnou, neboť v současné době (srpen 2007) se stále vyskytuje na prvním místě ve výsledcích vyhledávání na klíčová slova *malování* a *stěhování*.

11 Závěrečné shrnutí

Všech zvolených cílů předkládané disertační cíle bylo dosaženo a to jak ostatních cílů, tak zejména hlavního cíle v podobě navržení metodického postupu optimalizace dokumentu na určitá klíčová slova pro fulltextové WWW vyhledávače od úplného začátku, čili od předoptimalizačních příprav, až do úplného konce, tedy postoptimalizační analýzy efektivity. Celý navržený metodický postup optimalizace dokumentu je podrobně popsán v kapitole č. 9 a ověřen v kapitole č. 10.

Podklady pro navrhnutí metodického postupu optimalizace dokumentu pro fulltextové WWW vyhledávače je mimo rešeršní část disertační práce hlavně zhodnocení vlivu faktorů na hodnocení relevance dokumentu vyhledávacím strojem. Výsledná míra vlivu jednotlivých faktorů byla stanovena jednak ze cca čtyřletého pozorování různého průběžného umisťování 10 webových prezentací z různých oblastí a dokumentu „Google Search Engine Ranking Factors V2“,na kterém se podílelo 37 předních světových odborníků v oblasti vyhledávacích strojů.

Nově navrhnutý metodický postup optimalizace dokumentu pro fulltextové WWW vyhledávače lze shrnout do následujících činností:

1. *Analýza klíčových slov*

První krok se zabývá výběrem jednotlivých klíčových slov a stanovení jejich důležitosti za použití metod vícekriteriální analýzy variant. Analýza klíčových slov je tzv. odrazovým můstkem pro následující kroky navrženého metodického postupu optimalizace dokumentu pro fulltextové WWW vyhledávače a je většinou prováděna pro každý dokument zvlášť nebo ve specifických případech v rámci celého website.

2. *Analýza konkurence*

Analýza konkurence je svým způsobem druhým krokem metodického postupu optimalizace dokumentu na určitá klíčová slova pro fulltextové WWW vyhledávače a základem jsou výsledky z předchozí analýzy klíčových slov. Hlavním přínosem analýzy konkurence je, mimo přehledu celé řady důležitých faktorů, výpočet konkurenční hodnoty dokumentu pod určitým klíčovým slovem $CVD = c_i * \frac{1}{o}$, výpočet konkurenční hodnoty website pod určitým klíčovým slovem

$CVWi = c_i * \sum_{j=1}^{k} \frac{1}{o_j}$ a výpočet celkové konkurenční hodnoty website $CVW = \sum_{i=1}^{n} \left(c_i * \sum_{j=1}^{k} \frac{1}{o_j} \right)$.

3. *Analýza stávajícího stavu (pokud existuje)*

Za předpokladu existence website provozovatele je vhodné uskutečnit analýzu současného stavu, která víceméně vychází z dat analýzy konkurence. Dalšími informačními zdroji analýzy současného stavu website jsou zjištění hustoty klíčových slov, simulace zobrazení dokumentů vyhledávacím robotem v režimech „spider view“, „diagnostic view“ a „source-code view“, zjištění výskytu klíčových slov v jednotlivých částech dokumentu a website a rozbor chování návštěvníků website.

4. *MPOD copywriting*

Dalším, velmi důležitým, krokem v metodickém postupu optimalizace dokumentu pro fulltextové vyhledávací stroje je samotné psaní textů, tedy MPOD copywriting. Primárně se v tomto kroku vychází z definování cíle nebo cílů textu, čili, za jakým účelem nebo pro jaké uživatele je text vytvářen. Po definování cílů textu je text sestaven a v celém procesu MPOD copywritingu se následně sledují a upravují nejrůznější ukazatele, např. hustota klíčových slov, umístění klíčových slov, pestrost textu, zda je text intuitivní, rozložení odstavců apod.

5. *Optimalizace dokumentu a serveru*

Nejrozsáhlejší částí celého popisovaného metodického postupu je vlastní optimalizace dokumentu a serveru. Metodickým postupem optimalizace dokumentu, neboli optimalizací pro vyhledávací roboty, musí projít každý dokumentu z celého komplexu provozovatelova website. Základem pro optimalizaci je primárně analýza klíčových slov, analýza konkurence a MPOD copywriting, popř. i analýza současného stavu již existujícího website. Při vlastní optimalizaci musí MPOD expert brát v potaz:

- důležitost klíčového slova, která je výsledkem analýzy klíčových slov
- vliv XHTML značek a ostatních faktorů na fulltextové vyhledávací roboty
- možný počet výskytů XHTML značky v XHTML dokumentu a vliv ostatních faktorů na indexaci vyhledávacími stroji

- možný počet výskytů klíčového slova v dokumentu, který je výsledkem MPOD copywritingu
- konkurenční hodnoty dokumentů a ostatní faktory z analýzy konkurence

MPOD expert se musí zabývat takovými činnostmi, jako je výběr a registrace doménového jména, optimalizace WWW serveru, použitím klíčových slov dle jejich důležitosti v jednotlivých elementech dokumentu, tvorbou navigačních odkazů nebo popisky jednotlivých obrázků umístěných v XHTML dokumentu.

6. *Registrace do katalogů a tvorba zpětných odkazů*

Pozitivní hodnocení relevance dokumentu na určitá klíčová slova vyhledávacím strojem je podpořeno i vhodnými zpětnými odkazy a odkazy v katalogových vyhledávačích, čili přínosným link popularity dokumentu. Zpětné odkazy jsou odkazy, které odkazují na provozovatelův dokument a získávají se většinou výměnou. Při budování zpětných odkazů je tématická relevance odkazů směřujících na konkrétní dokument velmi důležitá. Vyjadřuje tématickou příbuznost mezi odkazující stránkou či webem, cílovým webem a klíčovým slovem. Registrace do katalogů má také stále smysl. Převážná většina katalogů sice nepředává extrémní hodnotu PR (na stránce se vyskytuje velké množství odkazů, mezi které se PR dělí), nicméně v žádném případě se taková registrace neprojevuje na website provozovatele negativně. Ze zkušeností s registracemi do katalogů vyplývá, že když byl spouštěn nový web a je přidán do seznamu nejpoužívanějších katalogů, bude Gogole Toolbar PageRank mezi hodnotami 2 a 4, což je dobrý základ pro další formu budování zpětných odkazů, např. právě výměnu odkazů.

7. *Analýza efektivity*

Každý provozovatel internetové prezentace by se měl zajímat o to, zda a jak dobře plní svůj účel. Potřebné informace lze nejlépe získat analýzou dat o počtu, cílech a činnostech návštěvníků website. Existují obvykle tři možnosti, jak toho dosáhnout. Zpracováním serverových logů, nebo nasazením některého z měřících systémů, např. Google Analytics, nebo využitím svého vlastního měřicího systému. Samotný počet návštěvníků informuje detailněji o unikátních přístupech na website za dané období, o průměrných přístupech na website za jednotlivá kratší období, např. dní a kolik návštěvníků přišlo na website poprvé. Vedle zjišťování běžných

údajů o pohybu a charakteru návštěvníků je měření konverzí nejdůležitějším údajem, neboť provozovatelovi website napoví, jak je na tom jeho prezentace z hlediska obchodní úspěšnosti. V případě komerčních website konverze nastane, když se z návštěvníka website provozovatele stane zákazník, nebo-li náklady, které provozovatel website do získání návštěvníka investoval, se promění v přínos.

Poslední dobou je kladen stále větší důraz na zvyšování relevance a zpřesňování poskytovaných výsledků vyhledávání. Ukázalo se, že dobrých výsledků lze dosáhnout zejména identifikací vhodných klíčových slov a jejich zapojením do obsahu stránek.

Účinnost MPOD se značně liší v závislosti na tematické oblasti a účelu website. Hlavními faktory, majícími vliv na účinnost, jsou konkurence a počet příjemců služby (čtenářů, potenciálních zákazníků). Nezanedbatelným aspektem je čas. SEM sice účinkuje mnohem rychleji, ale pouze po dobu, po kterou je průběžně financován. Hodnocení MPOD prokázalo, že účinky optimalizace se projeví až po několika měsících, ale patrné zůstávají i několik let.

12 Přehled literatury

1. ADAMEC, Radek. *Maximalizace konkurenční výhody pomocí ICT.* In: Systems Integration 2006. Sborník. Praha : Prague University of Economics, 2006, s. 120 – 124. ISBN 80-245-1050-2.
2. BHANGAL, Sham – JANKOWSKI, Tomasz. Web Design. Birmingham: Friends of ED, 2003. 423 s. ISBN 1-904344-16-x.
3. BHARAT, Krishna – MIHAILA, George. *Hilltop : A Search Engine based on Expert Documents.* [online]. [cit. 2007-06-01]. <ftp://ftp.cs.toronto.edu/pub/reports/csri/405/hilltop.html>.
4. BRIN, Sergey – PAGE, Lawrence. *The anatomy of a Large-Scale Hypertextual Web Search Engines*. [online]. [cit. 2007-06-01]. <http://infolab.stanford.edu/~backrub/google.html>.
5. BROŽOVÁ, Helena – HOUŠKA, Milan – ŠUBRT, Tomáš. *Modely pro vícekriteriální rozhodování.*. Praha : ČZU v Praze, 2003. 178 s. ISBN 80-213-1019-7.
6. BRUCE CLAY, Inc. *Search Engine Relationship Chart*. [online]. [cit. 2007-08-30]. <http://www.bruceclay.com/searchenginechart.pdf>.
7. BURNS, Enid. *Top 10 Search Providers, July 2007*. [online]. [cit. 2007-08-30]. <http://searchenginewatch.com/showPage.html?page=3626903>.
8. BURNS, Enid. *Top 10 Search Providers, April 2007*. [online]. [cit. 2007-08-30]. <http://searchenginewatch.com/showPage.html?page=3626208>.
9. BUSINESSWORLD. *Encyklopedie podnikové Informatiky*. *[*online]. [cit. 2007-04-28]. <http://www.businessworld.cz/bw.nsf/page/encyklopedie>.
10. CALISHAIN, Tara – DORNFEST, Rael. *100 způsobů jak vyzrát na GOOGLE*. Gliwice : Helion, 2004. 327 s. ISBN 83-7361-565-2.
11. The ClickZ. *Search-Style Ads Lift Brand Awareness : Study Says*. [online]. [cit. 2007-06-02]. <http://www.clickz.com/showPage.html?page=3381481>.
12. CROWEL, Grant. *Search Engine Marketing and Branding*. [online]. [cit. 2007-06-02]. <http://searchenginewatch.com/showPage.html?page=2200961>.
13. CZADEK, Miroslav. *Projektové řízení jako zdroj konkurenceschopnosti v ICT.* In: Systems Intergration 2006. Sborník. Praha : Prague University of Economics, 2005. s. 362 – 369. ISBN 80-245-1050-2.
14. ČESKÝ STATISTICKÝ ÚŘAD. *Informační společnost v číslech : 2007 Česká republika a svět*. [online]. [cit. 2007-05-30]. <http://www.czso.cz/csu/redakce.nsf/i/informacni_spolecnost_v_cislech_2007/$File/infospol2007kv.pdf>.
15. DoubleClick. *DoubleClick Performics 50 : Q1 2007 Search Trend Report*. [online]. [cit. 2007-08-01].

<http://www.doubleclick.com/us/knowledge_central/documents/TREND_REPORTS/DoubleclickPerformics50_Q1_2007.pdf>.

16. GÁLA, Libor – POUR, Jan – TOMAN, Prokop. Podniková informatika. Praha: Grada Publishing, 2006. 484 s. ISBN 80-247-1278-4.
17. GUPTA, Atul. *Analysis and Implications of Hilltop Algorithm*. [online]. [cit. 2007-06-01]. <http://www.seorank.com/analysis-of-hilltop-algorithm.php>.
18. HENNYEYOVÁ, Klára. *Informačné technológie v riadení.* In: Mezinárodné vedecké dni 2001. Sborník. Nitra : Slovenská poľnohospodárská univerzita v Nitre, 2001, s. 631 – 635. ISBN 80-7173-868-2.
19. INTERNET INFO. *NA VRCHOLU.cz : Seznam má na vyhledávání téměř dvoutřetinový podíl*. [online]. [cit. 2007-05-30]. <http://www.iinfo.cz/tiskove-centrum/tiskove-zpravy/navrcholu-vyhledavace-seznam-dvoutretinovy-podil/>.
20. JAK PSÁT WEB. *O tvorbě internetových stránek.* [online]. [cit. 2007-05-05]. <http://www.jakpsatweb.cz>.
21. LAURIE, Ben – LAURIE, Peter. *Apache : Správa webového serveru.* Praha : Computer Press, 1997. 257 s. ISBN 80-7226-043-X.
22. MASLAKOWSKI, Mark. Naučte se MySQL za 21 dní. Praha : Computer Press, 2001. 478 s. ISBN 80-7226-448-6.
23. MATĚNA, Petr. *Efektivnější využití IT infrastruktury.* In: Systems Intergration 2006. Sborník. Praha : Prague University of Economics, 2006. s. 399 – 404. ISBN 80-245-1050-2.
24. MSN MONEY. *Google Inc : Stock Chart, Index Chart*. [online]. [cit. 2007-09-01]. <http://moneycentral.msn.com/investor/charts/chartdl.aspx?symbol=GOOG&CP=0&PT=8>.
25. NĚMEC, Robert. *Letní škola SEO*. [online]. [cit. 2007-06-02]. <http://www.lupa.cz/>.
26. OHLÍDAL, Martin. *Proč nám dělá sdílení informací problémy.* In: Systems Integration 2007. Sborník. Praha : Prague University of Economics, 2007, s. 388 – 390. ISBN 978-80-245-1196-2.
27. PAYNE, Chris. *Naučte se ASP.NET za 21 dní.* Praha : Computer Press, 2002. 763 s. ISBN 80-7226-605-5.
28. PÁLENÍK, Richard. *Flexibilita vs. problém změny – dilema dnešního manažera IT.* In: Systems Intergration 2006. Sborník. Praha : Prague University of Economics, 2006. s. 56 – 57. ISBN 80-245-1050-2.
29. PETERKA, Jiří. *eArchiv*. [online]. [cit. 2007-05-03]. <http://www.earchiv.cz>.
30. POLÁK, Jiří – MERUNKA, Vojtěch – CARDA, Antonín. *Umění systémového návrhu : Objektově orientovaná tvorba informačních systémů pomocí původní metody BORM.* Praha : Grada Publishing a. s., 2003. 195 s. ISBN 80-247-0424-2.

31. PONKRÁC, Miloslav. *XHTML – proč vznikl tento standard?*. [online]. [cit. 2007-05-03]. <http://interval.cz/clanky/xhtml-proc-vznikl-tento-standard/>.

32. POUR, Jan. *Informatika a konkurenceschopnost – téma standardní, ale stále aktuální.* In: Systems Intergration 2006. Sborník. Praha : Prague University of Economics, 2005. s. 377 – 384. ISBN 80-245-1050-2.

33. POWELL, Thomas. *Web Design : Kompletní průvodce.* Brno : Computer Press, 2004. 818 s. ISBN 80-7226-949-6.

34. PROKOP, Marek. *Co je Search Engine Marketing*. [online]. [cit. 2007-06-02]. <http://www.lupa.cz/clanky/co-je-search-engine-marketing/>.

35. SCAMBRAY, Joel – SHEMA, Mike. *Hacking bez tajemství : Webové aplikace.* Brno : Computer Press, 2003. 328 s. ISBN 80-7226-769-8.

36. SEOmoz. *Search Engine Ranking Faktors V2*. [online]. [cit. 2007-08-01]. <http://www.seomoz.org/article/search-ranking-factors/>.

37. SLÁNSKÝ, David. *Abeceda konkurenceschopnosti aneb jak na ni s Business Intelligence.* In: Systems Intergration 2006. Sborník. Praha : Prague University of Economics, 2006. s. 63 – 71. ISBN 80-245-1050-2.

38. SMIČKA, Radim. *Optimalizace pro vyhledávače – SEO : Jak zvýšit návštěvnost webu*. Dubany : Jaroslava Smičková, 2004. 160 s. ISBN 80-239-2961-5.

39. ŠTĚPÁN, Karel. *Může vůbec IS/ICT v dnešním integrovaném světě zvyšovat konkurenceschopnost podniků na trhu s elektřinou a plynem?* In: Systems Intergration 2006. Sborník. Praha : Prague University of Economics, 2006. s. 58 – 62. ISBN 80-245-1050-2.

40. TOPLIST. *Podíl vyhledávačů v refereru : historie*.. [online]. [cit. 2007-09-01]. <http://history.toplist.cz/stat/?a=history&type=4>.

41. VANÍČEK, Jiří – PAPÍK, Martin – PERGL, Robert – VANÍČEK, Tomáš. *Teoretické základy informatiky*. Praha : Kernberg Publishing, 2007. 436 s. ISBN 978-80-903962-4-1.

42. VARGA, Andras. *Materials for Implementing Interactive Media*. University of Tampere, [cit. 2007-05-10]. <http://www.uta.fi/hyper/opetus/05kevat/p5/>.

43. VOŘÍŠEK, Jiří – NOVOTNÝ, Ota. *Cíle a metodika projektu analýzy nabídky a poptávky po ICT specialistech v ČR.* In: Systems Integration 2007. Sborník. Praha : Prague University of Economics, 2007, s. 19 – 35. ISBN 978-80-245-1196-2.

13 Seznamy

13.1 Seznam publikační činnosti autora (výběr)

1. ŠIMEK, Pavel. *Agrarian WWW cooperation*. In: Informační systémy v zemědělství a lesnictví 2007. Sborník. Praha: IPC PEF ČZU v Praze, 2007, s. 180 – 190. ISBN 978-80-213-1643-0.
2. ŠIMEK, Pavel. *Informační gramotnost : Základy dalšího rozvoje.* Klatovy: Úhlava, o. p. s., 2007. 212 s. ISBN 978-80-903851-1-5. Kapitola 5, Dostupnost informací a možnost komunikace na internetu, s. 77 – 89. Kapitola 6, Využití elektronického obchodování, s. 90 – 109. Kapitola 7, Prezentace na internetu, s. 110 – 141.
3. ŠIMEK, Pavel. *Marketing - inovace zemědělských a potravinářských výrobků, ochranné známky, schvalovací procesy dle podmínek EU, zásady propagace.* Praha: IPC PEF ČZU v Praze, 2006, s. 311 – 349. ISBN 80-213-1495-X.
4. ŠIMEK, Pavel. *Multiuživatelský systém pro publikování a prezentování zájmových sdružení regionů.* In: Informační systémy v zemědělství a lesnictví 2006. Sborník. Praha: IPC PEF ČZU v Praze, 2006, s. 190 – 195. ISBN 80-213-1494-X.
5. ŠIMEK, Pavel. *Realizace multiuživatelského systému pro publikování v prostředí WWW.* In: Agrární perspektivy XV. Sborník. Praha: PEF ČZU v Praze, 2006, s. 919 – 923. ISBN 80-213-1531-8.
6. ŠIMEK, Pavel. *Agrarian WWW portal Agris.* In: Informační systémy v zemědělství a lesnictví 2005. Sborník. Praha: IPC PEF ČZU v Praze, 2005, s. 4 – 11. ISBN 80-213-1429-X.
7. ŠIMEK, Pavel. *Architektura systému Agris*. In: Objekty 2005. Sborník. Ostrava: VŠB – Technická univerzita Ostrava, 2005, s. 291. ISBN 80-248-0595-2.
8. ŠIMEK, Pavel – ČERMÁK, Vratislav. *Elektronická prezentace zájmových sdružení v regionech.* In: Agrární perspektivy XIV. Praha: PEF ČZU v Praze, 2005, s. 885 – 890. ISBN 80-213-1372-2.
9. ŠIMEK, Pavel. *Hodnocení produktů internetového bankovnictví.* In: Sborník příspěvků z doktorského semináře. Sborník. Praha: PEF ČZU v Praze, 2005, s. 574 – 581. ISBN 80-213-1314-5.
10. ŠIMEK, Pavel. *Moderní přístup koncepčního a programového řešení agrárního WWW portálu Agris.* In: Objekty 2005. Sborník. Ostrava: VŠB – Technická univerzita Ostrava, 2005, s. 273 - 279. ISBN 80-248-0595-2.
11. ŠIMEK, Pavel. *Hodnocení nabídek registrátorů doménových jmen v doméně .cz.* In: Sborník příspěvků z doktorského semináře. Sborník. Praha: PEF ČZU v Praze, 2004, s. 86. ISBN 80-213-1150-9.
12. ČERMÁK, Vratislav – ŠIMEK, Pavel. *Internet as a Medium for Digital Photography Sharing.* In: Internet and Informations Systems II. Sborník. Praha: PEF ČZU v Praze, 2004, s. 14 – 20. ISBN 80-213-1282-3.

13. ČERMÁK, Vratislav – ŠIMEK, Pavel. *Uložení, distribuce a sdílení dat internetovými portály.* In: Agrární perspektivy XIII. Sborník. Praha: PEF ČZU v Praze, 2004, s. 645 – 650. ISBN 80-213-1190-8.
14. ČERMÁK, Vratislav – ŠIMEK, Pavel. *ASP Model Utilization.* In: Proceedings of Workshop Internet and Information Systems. Sborník. Praha: PEF ČZU v Praze, 2003, s. 6 – 9. ISBN 80-213-1118-5.
15. ŠIMEK, Pavel. *Bezdrátové sítě standardu IEEE 802.11.* In: Sborník příspěvků z doktorského semináře. Sborník. Praha: PEF ČZU v Praze, 2003, s. 25. ISBN 80-213-1016-2.
16. ŠIMEK, Pavel – ČERMÁK, Vratislav. *Implementace moderních metod při tvorbě internetového portálu.* In: Agrární perspektivy XII. Sborník. Praha: PEF ČZU v Praze, 2003, s. 919 – 924. ISBN 80-213-1056-1.

13.2 Seznam grafů

Graf 1: Nejpoužívanější světové vyhledávače (dle počtu vyhledávání - červenec 2007) [7]. 32
Graf 2: Podíl jednotlivých vyhledávačů v ČR, leden 2007 [19]. 34
Graf 3: Podíl vyhledávačů v refereru [40]. 34
Graf 4: Vývoj akcií společnosti Google (1. září 2004 – 31. srpen 2007) [24]. 35
Graf 5: Vývoj ceny za klik (CPC) a ceny za klíčové slovo (CPK) [15]. 80
Graf 6: Vývoj ceny za klik (CPC) za vyhledávače Google a Yahoo! [15]. 81
Graf 7: Nárůst drahých klíčových slov [15]. 81
Graf 8: Meziroční změna transakcí a prodejů (pouze kampaně s přímými příjmy) [15]. 82
Graf 9: Návratnost investice (pouze kampaně s přímými příjmy) [15]. 83
Graf 10: Útraty inzerentů na Yahoo! [15]. 84
Graf 11: Průměrný rank (pozice) a podíl kliků [15]. 84
Graf 12: Meziroční vývoj transakcí (pouze kampaně s přímými příjmy) [15]. 85
Graf 13: Grafické znázornění vah kritérií dle Saatyho metody. 109
Graf 14: Grafické znázornění výchozích variant pro analýzu klíčových slov (Veteran Car Club Dvůr Králové nad Labem. 143
Graf 15: Vývoj návštěvnosti a žádostí o testování historického vozidla. 150
Graf 16: Vývoj návštěvnosti webu ctvrtecka.cz. 157

13.3 Seznam obrázků

Obrázek 1: Schéma URI [29]. 23
Obrázek 2: překlad národního tvaru domény na IP adresu [29]. 24
Obrázek 3: Princip překladu do ACE tvaru [29]. 24
Obrázek 4: Vztahy mezi předními světovými vyhledávači [6]. 31
Obrázek 5: Matice FCB [25]. 74
Obrázek 6: Tři faktory ovlivňující Search Engine Marketing [25]. 75

13.4 Seznam příkladů

Příklad 1 – základní struktura XHTML dokumentu. 20
Příklad 2: Základní vyhledávací formulář Googlu. 45
Příklad 3: XHTML kód tagu body hlavní stránky webu ctvrtecka.cz. 154

13.5 Seznam schémat

Schéma 1: Metodický postup optimalizace dokumentu pro WWW vyhledávače. 104
Schéma 2: Metodický postup pro analýzu klíčových slov. 111
Schéma 3: Metodický postup pro analýzu konkurence. 118
Schéma 4: Metodický postup pro analýzu aktuálního stavu website. 120
Schéma 5: Metodický postup pro copywriting. 124
Schéma 6: Metodický postup pro optimalizaci dokumentu a serveru, část 1. 131
Schéma 7: Metodický postup pro optimalizaci dokumentu a serveru, část 2. 132
Schéma 8: Metodický postup pro tvorbu zpětných odkazů (zajištění link popularity). 136

Schéma 9: Metodický postup pro analýzu efektivity. 140

13.6 Seznam tabulek

Tabulka 1: Počty vyhledávání za duben a červenec 2007 [7] [8]. 32
Tabulka 2: Nejčastější kritéria pro výpočet váh klíčových slov na WWW stránce [20]. 86
Tabulka 3: Ukázková tabulka pro výpočet pořadí klíčových slov a frází..... 106
Tabulka 4: Ukázka tabulky pro analýzu konkurence, část 1..... 115
Tabulka 5: Ukázka tabulky pro analýzu konkurence, část 2..... 116
Tabulka 6: Ukázka rozboru části zpětných odkazů k dokumentu umístěné na první pozici ve výsledcích vyhledávání. 116
Tabulka 7: Výchozí tabulka pro analýzu klíčových slov Veteran Car Clubu..... 142
Tabulka 8: Výsledná pořadí klíčových slov (Veteran Car Club Dvůr Králové nad Labem). 143
Tabulka 9: CVD za klíčová slova testování historických vozidel. 144
Tabulka 10: CVWi za klíčová slova testování historických vozidel. 144
Tabulka 11: CVD za klíčová slova testace historických vozidel..... 144
Tabulka 12: CVWi za klíčová slova testace historických vozidel..... 145
Tabulka 13: CVD za klíčová slova veteran car club..... 145
Tabulka 14: CVWi za klíčová slova veteran car club..... 145
Tabulka 15: CVD za klíčová slova historické vozidlo. 145
Tabulka 16: CVWi za klíčová slova historické vozidlo. 146
Tabulka 17: CVD za klíčová slova historie motorismu..... 146
Tabulka 18: CVWi za klíčová slova historie motorismu..... 146
Tabulka 19: CVD za klíčová slova veteran pojištění..... 146
Tabulka 20: CVWi za klíčová slova veteran pojištění..... 147
Tabulka 21: CVD za klíčová slova klub historických vozidel..... 147
Tabulka 22: CVWi za klíčová slova klub historických vozidel. 147
Tabulka 23: CVW za všechna klíčová slova. 148
Tabulka 24: Umístění odkazu na website Veteran Car Clubu..... 151
Tabulka 25: Umístění odkazu na website Veteran Car Clubu II. 151
Tabulka 26: Umístění odkazu na website Veteran Car Clubu III..... 151
Tabulka 27: CVW za všechna klíčová slova po optimalizaci (mimo slovního spojení *klub historických vozidel*)..... 152
Tabulka 28: Výsledná pořadí důležitosti klíčových slov..... 153
Tabulka 29: Seznam odkazů na ctvrtecka.cz. 155
Tabulka 30: Kolik návštěvníků přivedla klíčová slova na ctvrtecka.cz..... 156

14 Přílohy

14.1 Použité zkratky

zkratka	vysvětlení
ACE	ASCII Compatible Encoding
ASCII	American Standard Code for Information Interchange – americký standardní kód pro výměnu informací – kódová tabulka definující znaky anglické abecedy a jiné znaky používané v informatice
b	bit
B	Byte
BI	Business Intelligence - zpracovávající informace z interních provozních systémů, případně externích informací a poskytující výstupní informace pro řízení společnosti na různých úrovních
BSC	Balanced Scorecard - koncept sledování výkonnosti (a systémy pro toto sledování) organizace pomocí tzv. vyvážených ukazatelů, zahrnující finanční i nefinanční ukazatele
CASE	Computer Aided Software Engineering – nástroje pro podporu analýzy, plánování a rozvoje informačních systémů
CPC	Cost Per Click - cena za klik
CPK	Cost Per Keyword - cena za klíčové slovo
CRM	Customer Relationship Management – systémy podporující řízení vztahu se zákazníky
CSS	Cascading Style Sheets – kaskádové styly – kolekce metod pro grafickou úpravu webových stránek
CTR	Click Through Rate - míra prokliku – poměr počtu prokliknutých inzerátů vůči počtu zobrazených inzerátů – vyjadřováno v procentech
CVD	Competitive Value of Document - ukazatel konkurenční hodnoty dokumentu
CVW	Competitive Value of Website - ukazatel konkurenční hodnoty website
ČSÚ	Český statistický úřad
DNS	Domain Name Systém
DSS	Decision Support System – označení pro software pro podporu rozhodování, poskytující souhrnné informace z provozních systémů umožňující kvalifikovanější rozhodnutí řídících nebo odborných složek organizace
DTD	Document Type Definition – definice typu dokumentu
ERP	Enteprise Resource Planning – plánování podnikových zdrojů
FTP	File Transfer Protocol
GTPR	Google Toolbar PageRank
HR	Human Resources – lidské zdroje
HTML	HyperText Markup Language
HTTP	HyperText Trasfer Protocol
IANA	Internet Assigned Numbers Authority
IBL	InBound Link – příchozí odkaz - odkaz vedoucí z cizího website na website provozovatele - v češtině bývá tento termín překládán jako zpětný
ICT	Information and Communication Technology - informační a komunikační technologie
IDN	Internationalized Domain Names – standard, podle kterého mohou doménová jména obsahovat i znaky národních abeced
IRI	Internationalized Resource Indicators
IT	Information Technology – informační technologie
JIT	Just In Time – filozofie zefektivnění dodavatelsko-odběratelského řetězce pomocí odstranění skladových zásob dynamickým propojením dodavatelů komponent a materiálu a výrobců finálních výrobků průběžné zásobování v čase
KM	Knowledge Management – je většinou chápán jako systém pro správu expertních znalostí, případně se může jednat o systémy uchovávající organizační znalosti – směrnice, postupy, integrované workflow apod.
KW	Key Word - klíčové slovo
MFF	Matematicko-fyzikální fakulta, Univerzita Karlova

UK	
MRP	Manufacturing Resource Planning – plánování výrobních zdrojů
MRP	Material Requirements Planning – plánování materiálních požadavků
PDA	Personal Digital Assistant – v současné době víceúčelový kapesní počítač
PPC	Pay Per Click - platba za klik
PR	PageRank
RBSE	Repositury Based Software Engineering
RFC	Request For Comments – žádost o komentáře - používá se pro označení řady standardů a dalších dokumentů popisujících internetové protokoly, systémy apod.
ROI	Return On Investment – metrika posuzování přínosu informačního systému do organizace porovnáním investičních a provozních nákladů rozložených v čase s měřitelnými přínosy (zisk, úspora nákladů)
SCM	Supply Chain Management – řízení řetězce zásob (dodavatelsko odběratelského řetězce)
SE	Second Edition – druhé vydání
SEM	Search Engine Marketing
SEO	Search Engine Optimization – optimalizace stránek pro vyhledávací roboty
SERP	Stránka zobrazující výsledky vyhledávání konkrétního termínu v podobě řady odkazů na různé WWW stránky
SGML	Standard Generalized Markup Language - nejobecnější značkovací jazyk, ovšem svým způsobem komplikovaný na implementaci, obsahuje velkou spoustu rysů, které jsou zřídkakdy využity
SQL	Structured Query Language - standardizovaný dotazovací jazyk používaný pro práci s daty v relačních databázích
TCO	Total Cost of Ownership – celkové náklady na implementaci a provoz IT systémů
TLD	Top Level Domain - doména nejvyšší úrovně
URI	Uniform Resource Indicator – jedinečná identifikace zdroje – obecně použitelná množina na všechny druhy adres, URL je zkratka pro Uniform Resource Locator a je podmnožinou URI
URL	Unique Resource Locator – jednoznačné určení zroje – způsob, jak jednoznačně zapsat umístění souboru na Internetu nebo na intranetu
VPN	Virtual Private Network – virtuální privátní síť vytvářená ve sdílené WAN infrastruktuře (nejčastěji v internetu) pro bezpečné propojení dvou a více lokalit
W3C	World Wide Web Consorcium – mezinárodní konsorcium, jehož členové společně s veřejností vyvíjejí webové standardy pro WWW
WAIS	Wide Area Information Servers – systém který transformuje dotaz uživatele do formální podoby, kterou vyžaduje databáze, do níž se uživatel hlásí prostřednictvím služby Telnet
WHOIS	v současné době služba WHOIS slouží pro zjištění informací o doménách a detailů s nimi souvisejících (registrátor, technický správce, majitel, administrativní kontakt, apod.)
WWW	World Wide Web
XHTML	eXtensible HyperText Markup Language
XML	eXtensible Markup Language – značkovací jazyk pro uchovávání a předávání dat ve strukturované a semistrukturované formě

14.2 Pojmy World Wide Web

WWW, World Wide Web

jméno dnes nejpopulárnější služby Internetu, určené pro zpřístupnění informací. Původně byla vyvinuta pro potřeby sdílení textových informací, později byla obohacena o možnost práce s grafikou a dalšími uživatelsky atraktivními prvky (animace, zvuky atd.). Služba World Wide Web má architekturu klient/server, což znamená že pro její fungování jsou zapotřebí WWW servery, a dále klientské programy (prohlížeče, browsery), prostřednictvím kterých uživatelé pracují s touto službou.

Hypertext

text, členěný na relativně menší celky (označované jako stránky), které jsou mezi sebou provázány prostřednictvím tzv. odkazů (links). Smyslem je umožnit čtenáři přijímat informace po menších a snáze srozumitelných a přehledných částech (jednotlivých stránkách), a podle vlastní potřeby si volit postup přijímání informací – uživatel má možnost přecházet podle vlastního uvážení z jedné stránky na druhou prostřednictvím příslušných odkazů, a tím si volit vlastní sekvenci stránek, které mu budou zobrazovány.

Page, stránka

základní jednotka informací, předkládaná uživateli v rámci služby WWW. Kromě textu mohou být součástí takovéto stránky i obrázky a další vizuální prvky (např. animace), a dokonce i výkonné prvky (programy, např. applety jazyka Java). Odkazy na jiné stránky (tzv. hypertextové odkazy) mohou být reprezentovány jak libovolně velkými či malými částmi textu (jednotlivými znaky, celými slovy), tak i celými obrázky (např. ikonami), ale také jednotlivými částmi obrázků (v případě tzv. citlivých map). Tyto odkazy přitom nemusí vést jen na jiné stránky, ale také na jiná místa téže stránky.

Browser, prohlížeč

klientský program služby WWW, pomocí kterého uživatel pracuje s touto službou. Browser zobrazuje uživateli jednotlivé stránky a dává mu možnost přecházet mezi stránkami prostřednictvím odkazů, vedoucích mezi nimi. Uživatel vyvolá tento přechod aktivováním příslušného odkazu, typicky poklepáním myší na to, co odkaz reprezentuje (což může být část textu, celý obrázek nebo jen konkrétní část obrázku).

HTML, HyperText Markup Language

jazyk, ve kterém je zapisován obsah jednotlivých WWW stránek. Je založen na principu označování částí textu pomocí značek (anglicky: mark-up), které specifikují význam textu - například to, že určitá část textu představuje nadpis, či má být zobrazena tučně. Jiné značky zase specifikují, že na určité místo má být vložen obrázek, že určitá část textu představuje hypertextový odkaz. Podle těchto značek se pak při zobrazování stránek v prohlížeči (browseru) určuje jejich grafická reprezentace.

WWW server

server, který uchovává u sebe jednotlivé WWW stránky jako textové soubory, napsané v jazyku HTML. Na explicitní žádost zasílá jednotlivé WWW stránky WWW klientům (browserům, prohlížečům), které podle HTML kódu vytvoří grafickou

reprezentaci stránky (vytvoří její grafickou podobu) a tu zobrazí uživateli. Kromě samotného uchovávání WWW stránek a jejich poskytování na žádost klienta může WWW server plnit i další doprovodné funkce – může například zajišťovat překódovávání obsahu stránek z jednoho kódování češtiny do jiného (podle konkrétních požadavků klienta). Nebo může zajišťovat dynamické generování WWW stránek, které nejsou připraveny dopředu (tj. nejsou statické), ale vznikají dynamicky až v okamžiku, kdy o ně klient požádá – například jde-li o dotaz do databáze, server nejprve zprostředkuje položení dotazu databázi, a následně zajistí dynamické vygenerování stránky s výsledky dotazu.

HTTP, HyperText Transfer Protocol

přenosový protokol, pomocí kterého komunikují servery a klientské programy (browsery) v rámci služby WWW. Podle pravidel jazyka HTTP tedy klient formuluje svůj požadavek na novou WWW stránku, kterou mu má server zaslat, a opět podle jazyka HTTP pak server formuluje svou odpověď. Klientský program (browser) přitom generuje jednotlivé požadavky na popud uživatele (jako důsledek toho, že uživatel klikne na určitý odkaz, tím vyjádří přání jej aktivovat, a klient si od serveru vyžádá WWW stránku, na kterou tento odkaz směřuje).

URL, Uniform Resource Locator

jednotný způsob označování objektů v Internetu (tj. jednotný způsob označování WWW stránek na WWW serverech, souborů v FTP archivech, uživatelských poštovních adres atd.). V rámci služby WWW jsou tímto způsobem řešeny odkazy na jednotlivé WWW stránky (jako tzv. URL odkazy). V URL odkazu na WWW stránku je obsažena informace jak o konkrétním WWW serveru, na kterém se stránka nachází, tak i informace o souboru ve kterém je obsažena (o jeho jménu a příponě), a stejně tak i informace o adresáři, resp. o přístupové cestě k tomuto souboru. Příklad URL odkazu na WWW stránku: http://www.pef.czu.cz/index.php.

WWW site, website

WWW stránky typicky tvoří celky, odpovídající logicky uceleným expozicím. Příkladem takovýchto "expozic" mohou být WWW stránky konkrétní instituce či firmy (prezentující její profil, obchodní nabídku, reference atd.), WWW stránky věnované konkrétnímu produktu, městu, kulturní památce, problematice (např. spammingu). Každá takováto „expozice“ (site) je vždy umístěna na nějakém konkrétním WWW serveru, ale na jednom serveru může být takovýchto „expozic“ více. WWW server je tedy fyzickým celkem, a WWW site („expozice“) celkem logickým.

Home page, domovská stránka

v užším slova smyslu jde o jednu konkrétní stránku v rámci celé „expozice“ (WWW site), která je zamýšlena jako vstupní, a jsou na ní tudíž umístěny základní orientační prvky. Také konkrétní způsob provázání stránek pomocí hypertextových odkazů je uzpůsoben předpokladu, že uživatel poprvé vstoupí do příslušné expozice skrz tuto zamýšlenou vstupní (domovskou) stránku. V širším slova smyslu se jako „domovské stránky“ označují celé skupiny WWW stránek, které prezentují určitý subjekt a obsahují základní informace o něm (např. domovské stránky firmy nebo domovské stránky určité fyzické osoby).

Cookie

charakteristickým rysem komunikace mezi WWW serverem a jeho klientem (daný protokolem HTTP) je jejich bezestavový charakter – každý jednotlivý požadavek klienta je posuzován samostatně a nezávisle na případných jiných požadavcích od téhož klienta či jiných klientů. To na jedné straně značně usnadňuje práci WWW serveru (který si nemusí pamatovat nic o předchozí komunikaci s různými klienty, a jednotlivé požadavky od různých klientů může vyřizovat v libovolném pořadí). Na druhé straně to ale komplikuje fungování některých mechanismů – z pohledu klienta je totiž někdy nutné, aby nový požadavek logicky navazoval na jeden či několik předchozích požadavků, a aby určitá (stavová) informace z předchozích požadavků byla zachována i pro požadavky nové. Příkladem může být volba kódování češtiny, kdy si uživatel jednou vybere požadované kódování, a jeho volba by měla být automaticky respektována i při dalších požadavcích. Jiným příkladem může být přihlášení uživatele, který zadá své jméno a heslo, a teprve pak se dostává ke stránkám, ke kterým by se jinak nedostal. Jedním z možných řešení je použití tzv. cookies (doslova: koláčky): jde o malé kusy textu, které generuje WWW server na základě jednoho požadavku klienta, a zasílá je klientovi v rámci odpovědi na jeho požadavek (tj. spolu s obsahem stránky, kterou si klient vyžádal). Obsahem takovéhoto „cookie“ pak jsou právě ty informace, které je třeba uchovat pro další návazné požadavky, ale které si server není schopen pamatovat. Místo toho si obsah „cookie“ pamatuje klient ve své paměti (na svém disku), a při dalším požadavku na tentýž server mu obsah cookie zase vrátí (čímž mu připomene předchozí požadavek a umožní mu na něj navázat). Celý mechanismus práce s „cookies“ je pro uživatele neviditelný, a příjem a zpětné zasílání se plně stará WWW browser (ale uživatel má možnost si nechat zobrazovat text všech přicházejících cookies, a eventuelně je odmítnout).

14.3 Internetové organizace

ISOC (Internet Society)

v době své akademické minulosti byl Internet do značné míry pod „nadvládou“ USA jakožto státu, který financoval rozhodující část chodu Internetu, a potřebným způsobem také „zaštiťoval“ různé technické orgány, které se staraly o chod a rozvoj Internetu, o jeho standardizaci atd. Jakmile ale počátkem devadesátých let Internet začal přecházet do rukou komerční sféry, role „státního deštníku“ nad Internetem se začala vytrácet (samotné státní orgány USA se samy začaly postupně stahovat z nejrůznějších aktivit kolem Internetu, s cílem přenechat je komerční sféře). Tím vznikla potřeba vytvořit „deštník“ nový, a tato potřeba byla v roce 1992 naplněna vznikem organizace ISOC (doslova: společnost pro Internet). Je nutné ji chápat nikoli jako orgán, který sám sebe pasuje do role jakéhosi vlastníka či „šéfa“ Internetu, a tuto globální síť sítí si přivlastňuje. Není tomu tak, Internet Society si nečiní nárok na vlastnictví Internetu či na výlučnou pravomoc rozhodovat o dění v něm. Je pouze sdružením lidí, kterým Internet leží na srdci a chtějí pro něj něco udělat – ISOC nemá formální mandát, ale její postavení vychází spíše z její morální autority a z autority lidí, kteří v ISOC působí.

IAB (Internet Architecture Board)

po celou dobu existence Internet, už od jeho samotného počátku vždy existoval nějaký orgán, který se staral o jeho technický rozvoj. Tento orgán se v různých časových údobích jmenoval různě – v době největšího rozkvětu akademického Internetu se jmenoval IAB, od: Internet Activities Board, a byl fakticky podřízen vládním orgánům USA, které také nepřímo financovaly jeho činnost (skrz grantovou agenturu NSF). Po přechodu Internetu do rukou komerční sféry a po vzniku ISOC přešel orgán IAB pod „zastřešující deštník“ Internet Society, která také převzala na svá bedra úkol zajišťovat finanční prostředky, nezbytné pro fungování IAB – ta si při svém přechodu sice zachovala své zkratkovité jméno, ale význam zkratky se poněkud změnil: nyní znamená: Internet Architecture Board. Souvisí to ostatně i s celkovým posláním IAB jako spíše technického orgánu, zabývajícího se celkovou architekturou Internetu, zatímco obecnějšími aktivitami kolem Internetu (například sháněním finančních prostředků, vystupováním jako partner vládních orgánů při jednání ohledně Internetu) se zabývá samotná Internet Society. Orgán IAB, jakožto součást ISOC, se fakticky zabývá vývojovými trendy v Internetu, a zastřešuje celý proces standardizace a tvorby standardů. Je také formálním vydavatelem dokumentů RFC. Neformálně jej lze přirovnat k jakési „radě starších“, která řeší zásadní a koncepční technické otázky (ale již se nezabývá konkrétní technickou prací, to přenechává svým podřízeným orgánům, které sama zastřešuje).

IETF (Internet Engineering Task Force)

faktická „technická práce“, neboli faktická příprava nejrůznějších standardů, se odehrává na půdě orgánu jménem IETF, který je podřízen orgánu IAB. Samotný IETF je přitom velmi volným sdružením odborníků, bez formálního členství – jeden oblíbený slogan dokonce říká, že členství v IETF je rozpoložením mysli. Vyjadřuje to skutečnost, že do aktivní práce v rámci IETF se může zapojit skutečně kdokoli po celém světě, kdo o to má zájem (a činí tak svou aktivní účastí v elektronických konferencích, v rámci kterých se diskutují jednotlivá řešení, případně svou fyzickou účastí na některém ze tří ročních plenárních zasedání IETF, která se konají v různých částech světa). V dobách akademického Internetu byl orgán IETF skutečně místem,

kde nejrůznější řešení skutečně vznikala – hlavně proto, že členy byli především lidé z akademické sféry, kteří obvykle pracovali na základě různých grantů. Když ale později Internet přešel do rukou komerční sféry, akademická sféra měla méně prostředků na vysílání svých zástupců do IETF a na financování jejich rozsáhlejší práce na půdě tohoto orgánu. Navíc vývoj nejrůznějších řešení se významně prodražil, a spíše se přesunul na půdu komerčních subjektů, které přeci jen mají více možností jak investovat do vývoje potřebné finanční prostředky. Dnes se tedy nová technická řešení pro potřeby Internetu rodí spíše mimo IETF, u různých firem, které pak svá řešení předkládají IETF jako návrhy na nové standardy. Na půdě IETF pak probíhá jejich schvalování – hlavně ale hledání konsensu o tom, zda dané řešení je či není vhodné, případně výběr z několika alternativ, které se nabízí.

IESG (Internet Engineering Steering Group)

orgán IETF přeci jen má určitou vnitřní strukturu – je členěn na tzv. oblasti (ureas), které se věnují určitým širším tématům, a v rámci nich pak vznikají jednotlivé pracovní skupiny, které se zabývají dílčími aspekty celé problematiky. Jednotliví odborníci pak mohou pracovat v jedné, nebo několika pracovních skupinách současně. Jelikož ale jejich zapojení do práce IETF je velmi volné a neexistuje zde ani formální členství, je celková organizace IETF příliš volná na to, aby tento orgán mohl přijímat určité formální kroky – mezi které nutně patří i schvalování návrhů nových standardů. Proto „nad“ IETF existuje jeho „dohlížecí orgán“ (jakási dozorčí či správní rada), s formálním členstvím a mnohem přesněji definovanou organizační strukturou. Jmenuje se IESG (Internet Engineering Steering Group) a je tím, kdo formálně vykonává nejrůznější kroky typu schvalování návrhů, resp. jejich přechodu mezi jednotlivými fázemi (Proposed Standard, Draft Standard a Internet Standard).

IRTF (Internet Research Task Force)

vedle orgánu IETF existuje ještě jeden další, obdobně zaměřený a také obdobně organizovaný orgán jménem IRTF, který se také zabývá novými technickými řešeními. Dělba práce mezi IETF a IRTF je přitom dána „aktuálností“ zkoumaného řešení: pokud má bezprostřední význam a využití, spadá do kompetence IETF. Pokud se jedná spíše o záležitosti, které se sice zkoumají již dnes, ale jejich nasazení bude aktuální až někdy v budoucnu, pak spadají do kompetence IRTF. Toto rozdělení lze interpretovat také tak, že IRTF se zabývá oblastí výzkumu (proto také slovo Research v jeho názvu), a jakmile příslušná řešení najdou praktické uplatnění, přechází do kompetence IETF.

IRSG (Internet Research Steering Group)

vnitřní organizace orgánu IRTF je stejně volná a neformální, jako v případě IETF. Proto i zde vznikla analogická „dozorčí rada“ s mnohem striktnější organizací, která zajišťuje potřebné formální kroky.

IANA (Internet Assigned Numbers Authority)

pro praktické fungování celosvětového Internetu jsou nutné i nejrůznější parametry, adresy a jiné číselné identifikátory, které musí mít stejný význam v rámci celého Internetu, nebo musí být v celém Internetu unikátní (použity jen jednou). Příkladem mohou být nejen IP adresy (které musí být celosvětově unikátní), ale například také tzv. čísla portů (kterými jsou identifikovány různé služby, např. služba WWW je standardně provozována na portu č. 80). V Internetu proto musí existovat centrální autorita (centrální subjekt), který bude používání a přidělování

takovýchto číselných hodnot zastřešovat a koordinovat – například vznikne-li nová služba, přidělí jí tato centrální autorita nové, dosud nepoužité číslo portu. Touto organizací je v současné době organizace jménem IANA, a ve skutečnosti jde o malé středisko při jedné z univerzit v USA, které stále ještě funguje z grantů vlády USA. Tato organizace pak má formálně „pod sebou“ například i celý systém doménových jmen DNS, a například právě ona formálně přiděluje správu nově vzniklých domén nejvyšší úrovně (TLD domén).

14.4 Standardizační instituce

ITU (International Telecommunications Union, mezinárodní telekomunikační unie)

organizace spadající pod OSN, je nejstarší organizací produkující standardy z oblasti spojů telekomunikací (z roku 1865). Jejími členy je více než 160 zemí, a v současné době má dvě části: ITU-Telecommunications (ITU-T) a ITU-Radiocommunications (ITU-R), přičemž každá z nich je členěna na studijní skupiny (Study Groups), které se dále dělí na pracovní skupiny (Working Parties), které se ještě dále člení podle konkrétních řešených otázek (Questions). Práce v rámci těchto oddílů zvaných „otázky“ je řízena zpravodajem (francouzsky Rapporteur). Část ITU-T se původně jmenovala International Telegraph and Telephone Consultative Committee (zkratkou CCITT, podle svého názvu ve francouzštině), a je známa svými standardy (formálně: doporučeními) i ve světě počítačů:

- doporučení řady V, týkající se modemů (např. V.32, V.42, nejnověji V.90)
- doporučení řady X, týkající se datových sítí a koncepce OSI (např. X.400, X.500)
- doporučení řady I a Q, která se týkají ISDN,
- doporučení řady T, která pokrývají oblast textových komunikací (teletex, fax, videotex atd.)
- doporučení řady H, která se zabývají práce s digitalizovaným zvukem a obrazem (např. standard H.323, využívaný i pro internetovou telefonii).

ITU je považována za nejformálnější ze všech formálních organizací. Její domovskou stránku lze nalézt na adrese http://www.itu.ch.

ISO (International Organization for Standardization)

mezinárodní nevládní organizace, založená v roce 1947, sdružuje národní standardizační instituce (např. ANSI, DIN, a další). Usiluje o harmonizaci národních standardů po celém světě – její zkrácený název (ISO) není zkratkou, ale převzat z řecké předpony „iso“, která znamená „stejné“ (čím jsou myšleny stejné standardy po celém světě). Výsledky svých aktivit publikuje ISO ve formě svých standardů, které pak jsou přejímány národními standardizačními organizacemi. Pokrývají širokou škálu technických aspektů, od tvaru a rozměrů šroubků a matiček, přes normy technických výkresů, elektrické konektory, přenosové protokoly počítačových sítí, až třeba po klasifikaci publikací (systém ISBN). Ve světě počítačů a počítačových komunikací je zřejmě nejznámější koncepce tzv. referenčního modelu ISO-OSI (ISO Open Systems Interconnection). Organizace ISO je známa také tím, že své standardy neposkytuje k volnému šíření (jako je tomu např. u standardů Internetu), a pouze je prodává, ze nemalé částky. Domovskou stránku ISO lze nalézt na http://www.iso.ch.

IEEE (Institute of Electrical and Electronics Engineering)

největší profesní (professional) organizace na světě. Byla založena roku 1884 skupinkou praktiků z nově vznikajícího oboru elektrotechnického inženýrství. Dnes má přes 320 000 členů z celého světa, kteří se podílí na aktivitách IEEE ve více než 147 zemích světa. Konkrétními aktivitami je například pořádání odborných konferencí, vydávání odborných časopisů, a také příprava a vydávání standardů. Pro počítačové sítě má největší význam standardizační orgán, založený v rámci IEEE

v únoru roku 1980 (a proto označovaný jako IEEE 802), který je specificky zaměřen na problematiku standardů lokálních sítí. Ten si pak vytvoří pracovní skupiny pro jednotlivé oblasti, které do této problematiky spadají. Snad nejznámější pracovní skupinou je IEEE 802.3, která se věnuje standardizaci Ethernetu. Zajímavé je, že standardy pocházející od této skupiny samy nepoužívají jméno „Ethernet“, které má registrováno firma Xerox. Místo toho hovoří o „sítích na bázi přístupové metody CSMA/CD“. Domovská stránka je na adrese http://www.ieee.org.

W3C (World Wide Web Consorcium)

konsorcium, vzniklé v roce 1994 v důsledku neobyčejného rozvoje služby World Wide Web a potřeby její standardizace. Konsorcium W3C nevděčí za svůj vznik středisku CERN, odkud pochází samotná služba WWW – jako první „hostitelská organizace“ se konsorcia ujala prestižní univerzita MIT (Massachussets Institute of Technology, druhou se pak stala francouzská INRIA, a třetí japonská univerzita KEITO). Konsorcium W3C se zabývá celým širokým spektrem otázek kolem populárního webu: počínaje jazykem HTML a protokolem HTTP, přes problematiku práce s grafikou, lokalizace a internacionalizace webu, možnosti zpřístupnění webu tělesně postiženým, až například po otázky duševního vlastnictví či digitální podpisy. Působnost konsorcia W3C na poli vydávání standardů pro Internet sice koliduje s působností standardizačních orgánů jako IETF (viz minule), ale vzhledem k existujícímu personálnímu propojení a blízké mentalitě obou orgánů zde nevznikají větší problémy – podle gentlemanské dohody IETF přenechává oblast WWW právě konsorciu W3C. To vydává své standardy ve formě vlastních dokumentů (formálně doporučení), ale obvykle jsou tyto dokumenty publikovány i jako standardy RFC. Domovská stránka konsorcia W3C se nachází na adrese http://www.w3c.org.

ANSI (American National Standards Institute)

dobrovolná standardizační organizace z privátního sektoru USA, založená roku 1918 pěti technickými společnostmi a třemi vládními agenturami (dnes má přes 1300 členů, mezi nimi nechybí žádná velká počítačová firma) . Byla zakládající členem ISO, ve kterém stále hraje významnou roli. Produkuje standardy pro oblast počítačů, mezi její nejznámější standardy patří např. kód ASCII (American Standard Code for Information Interchange), standard programovacího jazyka ANSI C, tzv. ANSI terminál, a od ANSI pochází např. většina přenosových standardů technologie FDDI. Domovská stránka je na adrese http://www.ansi.org.

EIA (Electronic Industries Association)

asociace komerčních subjektů z USA, reprezentující komunitu výrobců pokročilých technologií. Vznikla v roce 1924, původně jako asociace výrobců radiotechniky (Radio Manufacturers Association). Sponzoruje mnoho odborných aktivit, pořádá vlastní semináře a konference, a vydává také vlastní standardy. Mezi nejznámější zřejmě patří standard RS-232-C, definující chování sériového rozhraní. Domovská stránka EIA se nachází na adrese http://www.eia.org/.

VESA (Video Electronics Standards Association)

konsorcium výrobců videoadaptérů a monitorů, usilující o standardizaci zobrazovacích systém a video protokolů. Od tohoto subjektu pochází např. standard SVGA, či sběrnice VLB (VESA Local Bus), předchůdce dnešní sběrnice PCI. Domovská stránka: http://www.vesa.org.

IEC (International Electrotechnical Commision)

tato organizace, založená roku 1904, se zabývá přípravou a vydáváním mezinárodních standardů v oblasti elektrotechnologií, elektronicky a souvisejících oblastí. Členem je více než 50 států z celého světa, domovská stránka: http://www.iec.ch.

NIST (National Institute for Standards and Technology)

vládní agentura USA, zabývající se vývojem síťových protokolů pro oblast průmyslu. Dříve známá jako NBS (National Bureau of Standards). Nejznámější standard: MAP. Domovská stránka: http://www.nist.gov.

ČSNI (Český normalizační institut)

zabývá se tvorbou norem, jejich vydáváním a distribucí, poskytováním informací o národních, mezinárodních, evropských a zahraničních normách a o veškerých aktivitách souvisejících s normalizací. Vydává normy ČSN, je členem ISO a IEC za ČR. Domovskou stránku ČSNI lze nalézt na adrese http://netra.felk.cvut.cz/CSNI/.

14.5 České fulltextové vyhledávače a katalogy

Seznam

Seznam vznikl v květnu 1996 a jeho autorem je Ivo Lukačovič. Když se objevily první zahraniční vyhledávače, fascinovaly ho natolik, že v roce 1995 v programovacím jazyce Perl napsal svůj vlastní, který byl nápadně inspirován portálem Yahoo. Původní Seznam byl však pouze katalog webových stránek s možností vyhledávání v jeho položkách. První fulltext pro Seznam napsal Lukačovič ve spolupráci se svým bývalým spolužákem z ČVUT Štěpánem Škrobem a společností Solamylware. Světlo světa tak tento projekt spatřil dne 1. května 1997. Na jeho konci byl vyhledávač, který bral v úvahu i českou diakritiku a různé způsoby kódování. V současné době je Seznam nejen nejstarší, ale také nejnavštěvovanější český internetový portál. Od svého vzniku prošel mnoha změnami. Od srpna roku 2006 se vyrovnává s prvním zahraničním konkurentem. Na český trh v tomto roce vstoupila česká verze světově známého vyhledávače Google.

Google.cz

Od 10. srpna 2006 je doménové jméno google.cz registrováno na americkou společnost Google, Inc. a téměř okamžitě byly zaznamenány velké změny ve významu jednotlivých národních verzí vyhledávače Google. Lednová studie Navrcholu.cz poukazuje na razantní přesun zájmu k české verzi, a k tomu úměrný pokles role Google.com. V úvahu je však třeba brát fakt, že při zadání adresy Google.com mohou být někteří čeští uživatelé automaticky přesměrování na lokální verzi.

Centrum

Portál Centrum byl spuštěn 7. září 1999 a za jeho vznikem stojí dva absolventi Dopravní fakulty Univerzity Pardubice – Ondřej Tomek a Oldřich Bajer. Ti ještě během studií v roce 1997 založili společnost NetCentrum, která již několik let provozovatelem Centrum.cz a Centrum.sk. V současnosti je tento portál silnou dvojkou na českém trhu, která se neustále dynamicky vyvíjí. V roce 2006 společnost utržila 367,6 miliónů korun. V současnosti zaměstnává přes 200 lidí. V srpnu 2007 spustilo Centrum.cz specializovaný magazín digitálně.cz nebo unikátní vyhledávání MP3 skladeb na Internetu. Dne 5.9.2007 pak koupilo největší a nejnavštěvovanější český on-line vyhledávač software Stahuj.cz od společnosti Miton. To Centru umožnilo nabídnout svým uživatelům další specializovanou vyhledávací službu a svým obchodním partnerům větší reklamní plochy. Centrum využívá vlastní vyhledávací technologii Morfeo založenou na systému Sherlock Holmes, do které každoročně investuje kolem deseti miliónů korun. Dnes je jediným českým portálem, který v celé své historii používal jen vlastní fulltextový vyhledávač, jehož podíl na českém trhu byl na počátku roku 2007 zhruba pětiprocentní.

Atlas

Internetový portál Atlas.cz vznikl v roce 1997 a jako první portál v ČR nabídl fulltextové vyhledávání. Dnes je viditelnou trojkou na poli nejvýznamnějších českých portálů. Od roku 1999 jej provozuje společnost Atlas.cz, a.s., jejímž majoritním akcionářem je II. Epic Holding. V březnu roku 2006 nastoupil do Atlasu na pozici generálního ředitele David Duroň, který odstartoval rozsáhlý projekt pod názvem „Restart“, jehož cílem je stát se do roku 2008 dvojkou české portálové scény. V rámci tohoto programu bylo uvedeno nové logo Atlasu – modrozelený panáček, nová

personalizovaná Homepage a také odstartována reklamní kampaň „Pepík a Péťa z Pneuservisu“. V roce 2006 se Atlas.cz stal zástupcem ICQ na českém. Zároveň z této aplikace umožnil zasílání SMS zdarma prostřednictvím mobilního operátora O2, o rok později pak přidal SMS zdarma ve spolupráci s Vodafone. Klíčovou událostí roku 2007 je pro tento portál spuštění fulltextového vyhledávání ve spolupráci se společností Seznam. Od 1. srpna tohoto roku tak Atlas poskytuje výsledky nejpoužívanějšího českého vyhledávače.

Jyxo

Jyxo je v České republice čtvrtý nejpoužívanější internetový vyhledávač, který se na Internetu objevil koncem srpna roku 2002. Doména Jyxo.cz byla zaregistrována již 16. dubna 2002 samotným tvůrcem vyhledávací technologie, Michalem Illichem. Svojí technologií se Jyxo trochu podobá Google, zdaleka však není tak úspěšný a používaný. Již od svého prvopočátku oslovilo část internetové populace svými revolučními funkcemi a možnostmi na překvapivě vysoké technologické úrovni. Mezi jeho přední vlastnosti patří zejména kontrola pravopisu, české skloňování a časování, řazení nalezených výsledků do skupin či každodenní aktualizace. Jyxo vyhledává nejen stránky psané v jazyce HTML, ale také dokumenty ve formátech PDF a .doc, pomocí modulů pak může prohledávat v podstatě jakékoli datové sklady, včetně rozsáhlých SQL databází. Jyxo ovšem není jen vyhledávač, ale především firma, která vyvíjí sofistikovaný a často velmi inovativní software. Zároveň provozuje několik zajímavých projektů, mezi než patří databáze znalostí Mozek.cz, tematicky členěný agregátor článků a v neposlední řadě vynikající blokovací služba blog.cz. Přestože je Jyxo technologicky hodně na výši, již od svého vzniku nemá jednoduchou pozici na trhu. V podstatě všichni velcí hráči na českém Internetu mají fulltextové vyhledávání vyřešeno a Jyxo v tomto směru není mezi širokou veřejností dostatečně známé.

Caramba

Caramba je katalogový internetový vyhledávač, který pracuje na principu otevřeného systému. To znamená, že jeho obsah je budován relativně velkým počtem webmásterů, z nichž každý má na starost jednu či více stran katalogu odkazů, kterou vytvořil a kterou průběžně aktualizuje. Základní myšlenkou projektu je skutečnost, že jednotlivá témata zde zpracovávají lidé, kteří se na daný obor specializují a rozumí mu. Ti jsou pak schopni odkazy posuzovat, vybírat a zatřiďovat lépe, než daleko menší týmy zaměstnanců, jež spravují katalogy vyhledávačů budovaných na komerčním principu.

14.6 Zahraniční fulltextové vyhledávače a katalogy

Google

Tvůrci Google.com jsou dva studenti postgraduálního studia ze Stanfordovi univerzity, Larry Page a Sergey Brin. Společný zájem si našli v řešení otázky, jak z velkého množství dat, která se na Internetu shromažďují, získat relevantní informace. V lednu 1996 tak začali spolupracovat na vyhledávacím robotu BackRub. Na jaře 1998 si v Larryho ložnici postavili vlastní počítačový server a začali hledat investora, který by jejich projekt zafinancoval. Poté, co se jim nepodařilo oslovit hlavní vyhledávací portály, se rozhodli jít vlastní cestou. Na jejím počátku bylo potřeba sehnat základní kapitál pro start firmy. Jako první navštívili Andyho Bechtolsheima, jednoho ze zakladatelů Sun Microsystem. Tomu stačil jediný pohled na demo Google, aby v něm viděl budoucí potenciál. Obratem pak vypsal společnosti Google, Inc. šek na 100 000 dolarů. Zásadní problém byl v tom, že firma s názvem Google, Inc. v té době ještě neexistovala. Toto zdánlivé dilema se rychle vyřešilo. Larry a Sergey obešli většinu známých a příbuzných, následně našli několik dalších sponzorů a založili společnost Google, Inc. se základním jměním téměř milion dolarů. Od září 1998 se scházeli ve své první kanceláři v garáži, kde jejich projekt podporoval i nový kolega Craig Silverstein. Beta verze Google v té době odpovídala asi na 10 000 dotazů. O webový vyhledávač vracející relevantní dotazy se začal zajímat tisk a články vychvalující Google se objevili v Today a Le Monde. V prosinci vyhlásil PC Magazín Google nejlepší webovou stránkou a vyhledávačem za rok 1998.

Prestiž Google rok od roku rostla. Se vstupem investorů a příchodem nových kapacit začala být malá i nová kancelář v University Avenue v Palo Alto. Google měl v té době již osm zaměstnanců a denně odpovídal na 500 000 dotazů. Problémy s nedostatkem prostoru a techniky zmizely, až když se Google přestěhoval do Googleplexu, současného ředitelství v Mountain Wiew v Kalifornii.

Google je dnes nejpoužívanějším, nejpopulárnějším a nejúspěšnějším vyhledávačem na světě. Jeho vyhledávací robot, který byl pojmenován Googlebot, využívá k hodnocení stránek algoritmus PageRank a denně zpracovává přes 200 milionů dotazů. Vyhledávač se honosí řadou ocenění a lokalizací do desítek světových jazyků. V roce 2006 se na základě pravidelného průzkumu Brandz Top 100 agentury Millward Brown ve spolupráci s ekonomickým deníkem Financial Times stal nejcennější značkou světa. Hodnota Google se v loňském roce vyšplhala na 66,434 milionů dolarů, což znamená meziroční růst o 77 %. Popularita samotného výrazu narostla do takových rozměrů, že sloveso „google“ bylo oficiálně dne 15. června 2006 přidáno do Oxford English Dictionary jako synonymum pro hledání na webu. Samotný název Google však pochází z anglického výrazu Googol. Jedná se o matematický termín pro jedničku následovanou sto nulami. Výsledkem je nesmírně veliké číslo. To symbolizuje obrovské množství informací, které je nejvýkonnější světový vyhledávač schopen zpracovávat.

Yahoo!

Yahoo vzniklo v roce 1994. Za jeho tvorbou stojí dva studenti doktorského programu elektrotechniky na Stanfordské univerzitě, David Filo a Jerry Yang. Celá jejich myšlenka se zrodila z nepřehledné situace, která panovala na Internetu. Napadla je koncepce kompletně nového webu, kterou v podstatě představoval on-line seznam nejoblíbenějších a nejdůležitějších internetových stránek, které tématicky roztřídili

do jednotlivých kategorií. Svůj katalog pak pojmenovali Jerry´s Guide to the World Wide Web. Aniž by to předem tušili, katalog začaly navštěvovat stovky lidí a to i z míst mimo Stanfordskou univerzitu. Jerry s Davidem si postupem času uvědomili, že jejich akademický projekt lze snadno proměnit v projekt komerční. V červnu 1994 přejmenovali stránku na Yahoo!, jako zkratku ze slov „Yet Another Hierarchical Officious Oracle“. Den co den do katalogu přibývaly stovky odkazů a služba Yahoo na konci roku 1995 zodpovídala denně již na 4 milióny dotazů.

Dnes je Yahoo druhý světově nejpoužívanější portál, který zaměstnává 12 000 lidí po celém světě a do roku 2007 ho využilo více jak 500 miliónů unikátních návštěvníků. Jeho hlavním konkurentem je Google, díky němuž bylo Yahoo nuceno vyvinout svůj vlastní fulltextový vyhledávač.

MSN/Live

Služba MSN Search se poprvé objevila ke konci roku 1998. V té době využívala výsledky vyhledávání licencované technologie Inktogmi. Na počátku roku 1999 byla představena veřejnosti nová verze, která zobrazovala výpis Inktogmi v kombinaci s výsledky vyhledávače Looksmart.

V listopadu 2004 Microsoft po roce a půl usilovného vývoje zpřístupnil veřejnosti ukázkovou verzi vlastní vyhledávací technologii, kterou nasadil jako konkurenční zbraň proti odvětvovým leaderům, společnostem Google a Yahoo. V únoru 2005 byla spuštěna ve verzi ostré. V té době mělo MSN zaindexováno více jak 5 biliónu dokumentů, 400 milionů obrázků a 3 milióny znalostí. Brzy tato služby změnila uživatelské rozhraní a výsledky jejího vyhledávání začaly využívat i jiné portály. 8. března 2006 byla uvolněna beta verze nazvaná Live Search, která se stala výsledkem dlouhodobé práce Microsoftu na inovaci MSN Search. Ostrá verze této aplikace se dostala na dosah internetové veřejnosti dne 11. září 2006. V současné době je služba Microsoft Live Search třetím nejpoužívanějším vyhledávačem na světě s tržním podílem 8,4%.

Ask.com

Ask.com, původně známý jako Ask Jeeves, byl založen v roce 1996 jako divize společnosti IAC Search & Media v Barkeley v Kalifornii. S jeho vznikem jsou spojena dvě jména, Garrett Gruener a David Warthern. Původní myšlenka, na které byl Ask Jeeves postaven, byla velice zajímavá. Stejnojmenný ilustrovaný sluha nehledal relevantní stránky pouze na základě klíčových slov, ale přímo jako odpověď na přirozené otázky. To ho postavilo do pozice prvního komerčního vyhledávače, který dokázal relevantně odpovídat na ucelené dotazy, samozřejmě pouze v anglickém jazyce. Svojí intuitivností se tak stal velice oblíbeným vyhledávačem, který v dnešní době patří mezi nejsilnější americkou čtyřku.

Srdce této vyhledávací technologie tvoří od roku 2001 algoritmus Teoma známý také jako ExpertRank. V současnosti je tento impozantní vyhledávací systém založen na principu otázka-odpověď, kde je uživatel veden k zpřesňování dotazu za pomoci systému kladených otázek. Využívá tak jednak technologii pro zpracování přirozeného jazyka, tak i bázi znalostí, která v současné době obsahuje zhruba 7 000 000 otázek. Vyhledávač nabízí i svůj vlastní toolbar, který obsahuje spoustu užitečných funkcí a spolupracuje s prohlížeči Mozilla Firefox a MS Internet Explorer.

DMOZ

DMOZ neboli Open Directory Project (ODP) je největší a nejkvalitněji ručně setříděný katalog odkazů na světovém webu. Zakladateli tohoto serveru jsou Rich Skrenta a Chris Tolles, kteří se nechali inspirovat úspěchem jiných otevřených projektů. ODP byl spuštěn 5. června 1998 a jeho vlastníkem je firma Netscape Communications Corporation, která převzala původní rozcestník NewHoo.com. Cílem tohoto neziskového projektu je vybudovat největší katalog na Internetu s uživatelsky generovaným obsahem. Svým nekomerčním charakterem se tak odlišuje od většiny katalogů a staví se do pozice významného zdroje informací o internetových stránkách, který nesklouzl ke zvýhodňování určitých odkazů za provizi.

Projekt vytváří dobrovolní editoři z celého světa, kteří mají nejrozmanitější zájmy. Každý z nich má na starosti jednu či více kategorií, do nichž zařazuje navržené odkazy spadající do jeho oblasti. Záleží jen na něm, jestli stránky zavrhne, zařadí nebo přesune do jiné kategorie. Editorem se může stát každý, kdo má chuť nezištně pracovat, ochotu spolupracovat s jinými lidmi a vůli dodržovat základní stanovy. Pro vlastníky stránek z řad veřejnosti je v příslušných kategoriích k dispozici formulář pro podání návrhu editorům katalogu. Návrh neznamená nic jiného, než tip na stránku. Cílem ani smyslem návrhů není totiž prosadit zařazení příslušných stránek na DMOZu, ale snaha pomoci katalogu tím, že jeho editorům je zaslán do správné kategorie tip na kvalitní a užitečné stránky, které jsou označeny požadovaným titulkem a popiskem. Základní roli v tomto procesu však hrají velmi přísná pravidla pro navrhování stránek, z kterých je důležité vycházet. Nesprávné či nekvalitně podané návrhy jsou z katalogu rychle vyřazovány.