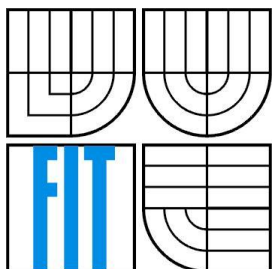


VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ

FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

SEO - OPTIMALIZACE PRO VYHLEDÁVAČE

SEO - SEARCH ENGINE OPTIMIZATION

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

PETR TRÁVNÍK

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. RUDOLF KAJAN

BRNO 2012

Abstrakt

Tato bakalářská práce obsahuje přehled faktorů ovlivňujících viditelnost webu ve vyhledávacích včetně optimalizace pro geografické oblasti, různé jazykové modifikace apod. Dále v práci najdete krátký přehled nástrojů, které usnadňují optimalizaci. Hlavním cílem je návrh a implementace webové aplikace, která slouží pro analýzu webových stránek včetně sledování pozic ve vyhledávacích.

Abstract

This Bachelor's thesis provides an overview of factors affecting site visibility in search engines including optimization for geographic areas, various language modifications, etc. Thesis gives a brief overview of tools that help in the process of web optimization. The main goal is to design and implement a web application capable of analyzing web pages and reporting positions in search engines.

Klíčová slova

SEO, Google, Seznam, internetové vyhledávače.

Keywords

SEO, Google, Seznam, search engines.

Citace

Trávník Petr: SEO - optimalizace pro vyhledávače, bakalářská práce, Brno, FIT VUT v Brně, 2012

SEO - optimalizace pro vyhledávače

Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně pod vedením Ing. Rudolfa Kajana. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

.....
Petr Trávník
15.5.2012

Poděkování

Chtěl bych poděkovat vedoucímu bakalářské práce Ing. Rudolfu Kajanovi za odbornou pomoc a rady. Dále bych chtěl poděkovat vlastníkům serverů, na kterých probíhalo testování optimalizací.

© Petr Trávník, 2012

Tato práce vznikla jako školní dílo na Vysokém učení technickém v Brně, Fakultě informačních technologií. Práce je chráněna autorským zákonem a její užití bez udělení oprávnění autorem je nezákonné, s výjimkou zákonem definovaných případů.

Obsah

Obsah.....	1
1 Úvod	2
2 Metodika vyhledávání.....	3
2.1 Internetové vyhledávače	3
2.2 Vyhledávací robot.....	4
2.3 Klíčová slova	5
2.4 On-page faktory	6
2.5 Off-page faktory	9
3 Nástroje pro analýzu	12
3.1 Vyhledávač Google	12
3.2 Google Analytics	13
3.3 Google Keyword Tool	13
3.4 Google Webmaster Tools	14
3.5 Validátory	14
4 Návrh aplikace pro SEO analýzu.....	15
4.1 Výběr klíčových slov.....	15
4.2 On-page analyzátor.....	15
4.3 Off-page analyzátor	16
4.4 Sledování pozic.....	16
5 Implementace.....	17
5.1 Technologie	17
5.2 Stavba aplikace	20
6 Testování.....	24
6.1 Volba klíčových slov	24
6.2 Efektivita optimalizací.....	25
6.3 Použití aplikace.....	28
7 Závěr	31
Literatura	32
A Uživatelská příručka	33
B Obsah DVD	38

1 Úvod

Internet obsahuje v dnešní době obrovské množství informací, které by byly bez vyhledávačů prakticky nepoužitelné. Vyhledávačů existuje velké množství a všechny si kladou za cíl podat uživateli co nejrelevantnější informace o vyhledávaném tématu. Téma uživatel definuje tzv. klíčovými slovy. Podle těchto klíčových slov pak vyhledávač určí relevantní odkazy a seřadí je podle kvality. Právě metody řazení a určení kvality stránky jsou tématem této bakalářské práce.

Díky znalostem o řazení výsledků pak můžeme provádět optimalizaci webové stránky pro vyhledávače - SEO (z angl. Search Engine Optimization). SEO nám může poskytnout za malé peníze mnoho nových návštěvníků případně nových zákazníků pro firmu.

Práce obsahuje obecné zásady platné pro standardní internetové vyhledávače, do hloubky se zaměřuje na vyhledávač Google a částečně také Seznam.cz. Google představuje zástupce celosvětových vyhledávačů přístupných ve všech jazycích a zároveň nejpoužívanější internetový vyhledávač na světě [3]. Seznam.cz je pak nejpoužívanější vyhledávač v Česku (viz Obrázek 2.1), jehož schopnost vyhledávání je omezená pouze na weby psané v českém jazyce.

Cílem této práce je všechny SEO techniky vyzkoušet v reálném prostředí s konkurencí a získat přehled o jejich efektivitě. Výstupem práce je pak aplikace, která usnadňuje SEO a zajišťuje metriku pro určení úspěšnosti optimalizací, tedy nástroj pro sledování pozic ve vyhledávačích.

Kapitola 2 - Metodika vyhledávání - se zabývá metodikou řazení ve vyhledávačích a snaží se informovat o všech faktorech, které ovlivňují pozice.

Kapitola 3 - Nástroje pro analýzu - rozebírá nástroje poskytované samotnými vyhledávači, které nám dávají informace potřebné k optimalizaci webových stránek. Díky těmto nástrojům můžeme určit vhodnou strategii pro SEO a optimalizaci tak zefektivnit.

Kapitola 4 - Návrh aplikace pro SEO analýzu - rozebírá vytvoření aplikace, která bude mít za úkol usnadnění, zefektivnění a měření úspěšnosti SEO.

Kapitola 5 - Implementace - obsahuje popis technologií použitých pro vytvoření a běh aplikace.

Kapitola 6 - Testování - se zabývá testováním SEO technik pro zvýšení pozic a testováním vytvořené aplikace.

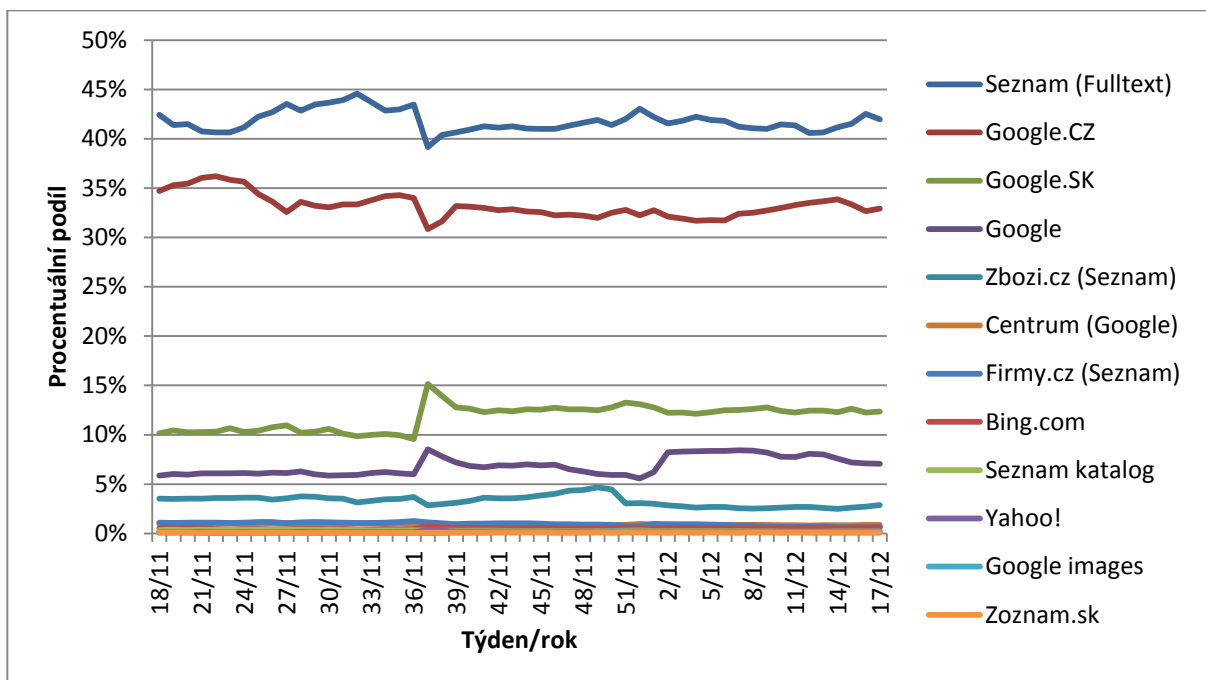
Kapitola 7 - Závěr - obsahuje zhodnocení práce.

2 Metodika vyhledávání

Řazení výsledků vyhledávání probíhá pomocí speciálních algoritmů. Tyto algoritmy jsou většinou tajné a mohou se měnit klidně i několikrát za rok. Algoritmy zohledňují mnoho faktorů a jsou navrženy tak, aby zařadily relevantní a kvalitní informace na nejvyšší pozice. Základním dělením faktorů řazení je rozdělení na skupiny *on-page* a *off-page*.

2.1 Internetové vyhledávače

Na internetu existuje mnoho různých vyhledávačů. Každý z vyhledávačů používá různé algoritmy pro řazení svých výsledků. Algoritmy jsou většinou pečlivě sříženy a postupně se vyvíjí například změnou vah jednotlivých faktorů. Jelikož jsou vývoj a údržba internetového vyhledávače neuvěřitelně náročné, menší vyhledávače si nechávají dodávat výsledky od externích search engineů. U nás například vyhledávač Centrum používá výsledky od firmy Google. I největší tuzemský vyhledávač Seznam používá vlastní search engine pouze pro vyhledávání na českém internetu, pro vyhledávání v zahraničním internetu používá Bing. Podíl vyhledávačů se ale mění v souvislosti s vyhledávaným tématem.



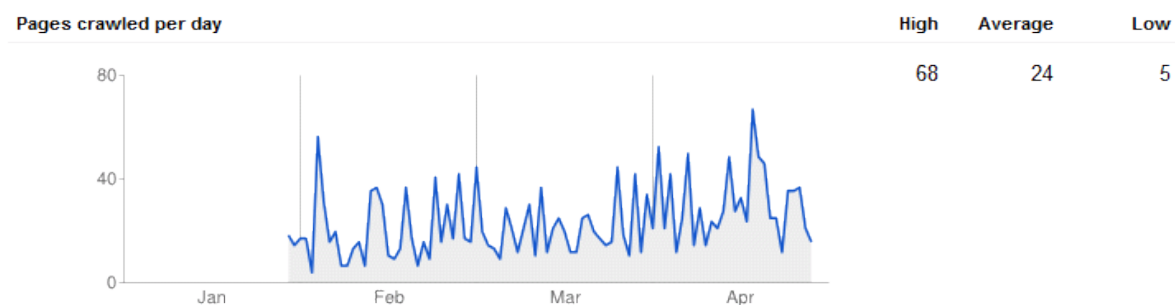
Obrázek 2.1- Podíl vyhledávačů v ČR (Zdroj: Toplist.cz [14])

Pro SEO je důležité vybrat správné vyhledávače, na kterých se budou výsledky sledovat. Česko je jeden z mála světových regionů, kde nemá naprosto dominantní pozici vyhledávač Google. Na obrázku Obrázek 2.1 vidíme procentuální podíl vyhledávačů za poslední rok. Statistiky pochází ze serveru Toplist.cz, který provádí audit návštěvnosti téměř milionu serverů v České Republice. Ze statistiky vyplývá, že v podílu vyhledávání vede s nadpoloviční většinou Google. Na druhém místě se pak nachází Seznam.cz s podílem přes čtyřicet procent objemu vyhledávání. Ostatní vyhledávače mají pouze zanedbatelný podíl.

2.2 Vyhledávací robot

Vyhledávače používají tzv. vyhledávací roboty. Ty periodicky prochází veškerý obsah internetu, ke kterému mohou přistoupit a o kterém vědí. Informovat vyhledávacího robota o existenci nového serveru by měl být první krok, který autor webu udělá. Indexovací robot může zaregistrovat nový webový server tak, že nalezne odkaz z již zaindexované stránky. Další možností je použít formuláře pro zapsání nového serveru do databáze odkazů, které by měl nabízet každý standardní internetový vyhledávač. Někteří vyhledávací roboti mají dokonce nastavený limit potřebných odkazů z externích webů a teprve když nalezne dostatek zmínek, zařadí nový server do své indexovací databáze. Následná perioda návštěv robota se průběžně mění. Webmaster Tools (kapitola 3.4) například nabízí přehled návštěv robota Google, viz Obrázek 2.2.

Crawl Stats Googlebot activity in the last 90 days



Obrázek 2.2 - Statistika návštěv stránek robotem

Z principu je vyhledávací robot aplikace s neuvěřitelně vysokými paměťovými a výkonnostními nároky. Robotům se tedy musí předhodit webová stránka s pokud možno jednoduchou strukturou odkazů, která usnadní průchod napříč obsahem celého serveru. Vyhledávací roboti se snaží indexovanou stránku pro své použití co nejvíce zjednodušit. Většina robotů tedy vynechává obsah, který je příliš složitý na zaindexování. Jedná se například o obsah generovaný pomocí JavaScriptu nebo Adobe Flash. Většina robotů má také problémy s indexováním webů vytvářených pomocí rámců, případně s použitím vložených rámců. Indexovací robot nevnímá jakoukoliv grafickou úpravu

stránky. Pokud obsahuje stránka obrazové materiály, pak vnímá robot pouze jejich alternativní popisky.

Vyhledávací robot také detekuje některé podvodné techniky, za které mohou dostat stránky postih ve formě nízké pozice ve výsledcích nebo naprostého vymazání stránek z databáze. Za jednu z nejrozšířenějších podvodných technik se dá považovat tzv. *hidden content*. Tato metoda spočívá v podstrčení obsahu s vysokým obsahem klíčových slov vyhledávacímu robotovi. Obsah je ale pro normálního uživatele stránek skryt například použitím stejné barvy písma a podkladu nebo překrytím textu obrázkem. Moderní roboti již takové případy částečně dokážou detekovat a podvodný obsah ze stránky vyřadit. Další technikou, za kterou následuje postih, je tzv. *cloaking*. Při cloakingu se podstrkává vyhledávacímu robotovi zcela jiná stránka, než která se zobrazí uživateli. Cloaking se provádí například pomocí skriptování na straně serveru. Technika *doorway page* se snaží o zajištění vysokých pozic ve vyhledávacích pomocí opakovaného vkládání klíčových slov na samostatnou stránku, z které pak teprve vede odkaz na prezentaci určenou návštěvníkovi. Podobná technika jako doorway page je tzv. *deceptive redirect*. Opět se jedná o vysoce optimalizovanou samostatnou stránku pro vyhledávače, ze které je návštěvník okamžitě přesměrován (např. pomocí JavaScriptu) na samotnou prezentaci.

Dalším problémem, který může způsobit snížení pozice stránek, je nedostupnost webového serveru. Pokud se indexovací robot setká s nedostupností stránek, tak server okamžitě ztrácí důvěryhodnost. Proto je dobré využívat prostředky pro monitoring přístupnosti a webovou prezentaci umístit na kvalitní server.

Je nutné, aby pro účely následného vyhledávání určil robot jazyk, ve kterém je obsah stránky napsán. Existuje meta tag v HTML hlavičce, který určuje jazyk stránky. V praxi však robot tento meta tag nepoužívá a provádí analýzu textu. V textech například vyhledává slova typická pro ten který jazyk a dle počtu výskytů určí jazykovou mutaci. Roboti v dnešní době již nemají problémy s detekcí kódování znaků. Kódování se definuje pomocí meta tagu v HTML hlavičce.

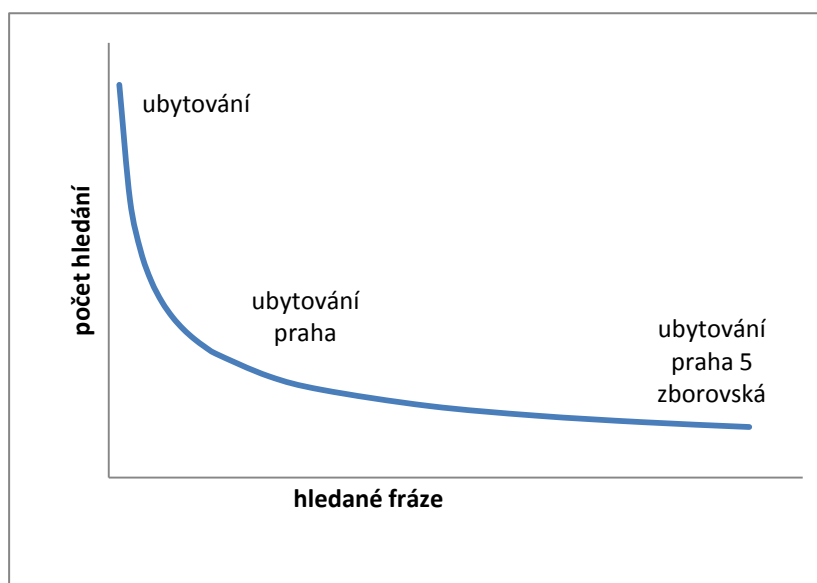
2.3 Klíčová slova

Klíčové slovo je pojem označující frázi, kterou uživatel zadává do vyhledávače. Tvoří se s ohledem na stručnost a maximálně zhušťují popis hledaného tématu nebo informace. Většinou se používají fráze v prvním pádě a neobsahují slovesa. Výsledkem hledání by pak měly být stránky, které se věnují tématu popsanému klíčovým slovem.

Optimalizace on-page i off-page faktorů se zaměřuje na stejná klíčová slova tak, aby bylo SEO dostatečně efektivní. Proces výběru klíčových slov je základem celé optimalizace a měl by být také prvním krokem. Existují dva problémy při volbě klíčových slov. Zaprvé je to volba příliš ambiciózních slov s velkou konkurencí. Probojovat se u takových slov na první stranu výsledků trvá velmi dlouho, ale uživatel vyhledávače obvykle volí odkazy hned z první strany, málokdy pokračuje dále. Druhým problémem je volba nevyhledávaných klíčových slov. Pokud se zaměříme na slova,

kteřá nikdo nehledá, pak nezískáme žádné návštěvníky, i když budeme na první pozici. Nástroji, které usnadňují výběr klíčových slov, se zabývá kapitola 3.

S výběrem klíčových slov souvisí pojem *Long tail*. Ten reflektuje trend vyhledávání, kdy existuje malý počet frází, které se denně vyhledávají mnohokrát a vedle toho obrovské množství frází, které se vyhledávají jen velmi málo. V součtu ale málo vyhledávané fráze zabírají větší objem vyhledávání než fráze s větší konkurencí. Viz Obrázek 2.3 – svislá osa zároveň také odpovídá ceně SEO a velikosti konkurence.



Obrázek 2.3 - Long tail – nejvyhledávanějších klíčových slov na internetu je málo, většinu objemu dotazů na vyhledávače dělají málo hledaná slova

2.4 On-page faktory

On-page faktory se vyznačují tím, že je ovlivňuje samotná webová stránka – jsou determinovány přímo autorem webu. Z hlediska SEO je optimalizace on-page faktorů základ, na kterém se dále staví. Tyto faktory se promítají do výsledků zejména, pokud jsou stránky menšího rozsahu. Při velké konkurenci a u rozsáhlých webů jsou většinou díky množství obsahu všechny on-page faktory na maximální úrovni a řazení dále ovlivňují především off-page faktory. On-page můžeme dále rozdělit na faktory určené obsahem a určené zdrojovým kódem. Výhodou on-page optimalizací je jejich jednoduchý a rychlý průběh s téměř okamžitou odezvou ze strany internetového vyhledávače. Projeví se hned, jakmile stránku navštíví indexovací robot.

2.4.1 Faktory určené zdrojovým kódem

Z hlediska zdrojového kódu získá stránka kladné body, pokud má správnou strukturu nadpisů a kód odpovídá standardům. Především první bod – struktura nadpisů a odstavců je důležitá, jelikož vyhledávače určují důležitost textu podle jeho označení. Web by tedy měl používat tagy H1–H6 a členit texty do odstavců. Menší vliv pak má také poměr zdrojového kódu k textu, nicméně toto je problém spíše pro uživatele prohlížečů v mobilních telefonech.

Dále je důležité, aby měl robot vyhledávače k dispozici odkazy na veškerá místa webu. Pro webové prezentace menšího rozsahu (do deseti stránek) většinou platí, že existují odkazy mezi všemi stránkami navzájem. U větších webů ovšem může nastat problém, kdy buď neexistuje odkaz na některou stránku, nebo se robot při procházení ztratí. Toto můžeme řešit vytvořením XML souboru formátu *Sitemap*, do kterého uvedeme odkazy na všechna místa, která se mají zaindexovat.

Další možností, jak můžeme ovlivnit sekce, které robot vyhledávače prohledá je vytvoření souboru *robots.txt*. Pomocí nastavení dle standardu, který lze nalézt v [1], můžeme zakázat jednotlivým nebo všem robotům indexaci určitých částí webu. Špatně vytvořený *robots.txt* je častou příčinou vyškrtnutí serveru z výsledků vyhledávání. Při aplikování pravidel platí zásada - co není zakázáno, je povoleno. Pro skrytí například placených článků před vyhledávači ale *robots.txt* není vhodný. V takovém případě se používá meta tag *robots* v hlavičce stránky.

```
# poznámky se uvozují znakem #
User-agent: * # obecné pravidlo pro všechny roboty
Disallow: /admin/
Disallow: /tmp/ # zákaz indexování obsahu ve složkách tmp a admin

User-agent: googlebot # následující pravidla platí pouze pro Google
Crawl-delay: 5 # doba mezi dotazy na server, odlehčuje serveru
Disallow: /odkazy/
```

Kód 2.1 – Příklad souboru *robots.txt*

Dalším místem, kam můžeme zakomponovat klíčová slova pro zvýšení relevance naší stránky je samotná URL adresa dokumentu. Pomocí různých technik (např. vytvoření souboru *.htaccess*) můžeme dynamický web, který si předává kontext v URL stránky upravit tak, aby se pro roboty vyhledávačů jevil staticky. Toto mělo smysl zejména v době, kdy Google.com neuměl analyzovat všechny parametry typu GET a klíčová slova v URL přidávaly stránkám lepší pozici. Nyní se ale stále najdou vyhledávače, které klíčová slova v URL hodnotí kladně (Seznam.cz).

```
http://www.studio-pure.cz/?lang=cs&page=sauna - před optimalizací  
http://www.studio-pure.cz/cs/sauna-praha-5 - po zapojení htaccess
```

Kód 2.2 – Příklad optimalizování URL

Ohledně zdrojového kódu pak musíme robotům vyjít vstříc i v případě prezentace obsahu. Problémy způsobují webové prezentace tvořené pomocí ráků – robot totiž považuje vnořené stránky za samostatné jednotky a vyhledávač pak odkazuje přímo na vnořené dokumenty. Dále není vhodné použití Javascriptu pro zobrazení důležitého obsahu. Robot Googlu je nyní jeden z mála, který dokáže zpracovat alespoň základní příkazy v Javascriptu, jinak je ale takový obsah pro roboty neviditelný. Dalším problémovým formátem pro prezentaci obsahu je Adobe Flash. V posledních verzích se sice objevily možnosti, jak udělat obsah Flashových aplikací indexovatelný, nicméně takový obsah nemá správnou strukturu.

Součástí optimalizace webové stránky by mělo být mimo jiné také minimalizování velikosti webové stránky. V dnešní době již není třeba se ohlížet na uživatele domácích PC. Vysoké oblíbenosti se ale těší mobilní připojení k internetu, které většinou nabízí omezenou rychlost. Zajímavým pro SEO tak například může být poměr zdrojového kódu k samotnému textu.

2.4.2 Faktory určené obsahem

Samotný text stránky je nejdůležitějším on-page faktorem, který určuje relevanci k hledaným klíčovým slovům. Podle členění stránky se pak určuje důležitost jednotlivých částí textu.

Nejdůležitější sekcí s největším vlivem na relevanci stránky je titulek. Tento se zadává do tagu `title`. Jeho délka by neměla přesáhnout 65 znaků, jelikož se zobrazuje ve výsledcích vyhledávání a při delších titulcích se zkracuje. Vzhledem k omezené délce je vhodné zde zadávat pouze klíčová slova s největší prioritou a konkurencí.

Dalšími sekcemi, které ovlivňují pozice, jsou meta tagy `description` a `keywords`. Druhý jmenovaný se již nyní ve výsledcích většinou nezohledňuje. Sloužil k přímému zadávání relevantních klíčových slov autorem. Meta tag byl ale příliš zneužíván a nyní ho například Google ve výsledcích naprosto zavrhuje [3]. Naproti tomu meta tag `description` je pro SEO velice důležitý. Obsahuje krátký popis obsahu stránky a Google jej používá při zobrazení výsledků vyhledávání. Description by mělo být stejně jako titulek stránky jedinečné pro každou stránku webu. Délka meta tagu by neměla přesáhnout 155 znaků. Důvod je stejný jako u titulku stránky – při delším textu je robotem zkrácen.

Samotný text je pak členěn do struktury určené zdrojovým kódem. Zde roboti rozlišují různé důležitosti textu dle umístění. Za nejdůležitější se považují hierarchické nadpisy (tagy H1 – H6). Texty v odstavcích se pak mohou zvýraznit pomocí tagů `strong`, které značí vyšší důležitost části textu. Zvýraznění a nadpisů je třeba využívat přirozeně, jinak hrozí penalizace. Počet klíčových slov

v textu pak také nesmí být uměle navyšován, tato snaha by byla kontraproduktivní. Někteří roboti dokážou rozeznávat také členění struktury na hlavičku, menu, obsah, patičku apod.

Obsah stránky by měl být jedinečný především v rámci webové prezentace. Stránky, které mají stejný obsah, se mohou navzájem vyrušit, ale nikdy nebudou ve výsledcích zobrazovány najednou. Pro SEO je nejlepší použití originálních textů, které v rámci možností obsahují klíčová slova – tímto se zabývá takzvaný *Copywriting*. Robotu Google nedělá problém skloňování a příbuzná slova. Může se vám tak například stát, že na první pozici vyhledávání se zobrazí stránka, která přesné klíčové slovo neobsahuje, ale přesto se váže na stejné téma. S obsahem stránky souvisí pojem *stuffing*. Ten znamená umělé doplňování klíčových slov do textu. Na webu se mnohdy najdou texty, které opakují klíčová slova i vícekrát v jedné větě. Stuffing je ale technika kontraproduktivní, jelikož vyhledávače takové snahy většinou ohodnotí negativními body a pozice stránky klesají. Klíčová slova se musí v textu objevovat v rozumné míře a tam, kde dávají smysl.

Vizuální obsah stránky jako například obrázky jsou pro nevidomé roboty problémem. Proto je důležité, aby všechny vizuální prvky s výpovědní hodnotou měly svůj alternativní popis. Tento je důležitý především při vyhledávání obrázků na Google. Vyhledávač obrázků může také přinést velké množství návštěvníků na web. S alternativními popisky souvisí pojem *alt spamming*. Jedná se o pojem příbuzný stuffingu. V tomto případě se ale klíčová slova uměle přidávají do alternativních popisků. Opět se jedná o techniku, kterou hodnotí vyhledávače negativně.

2.5 Off-page faktory

Zatímco u webů s velmi malou konkurencí stačí on-page optimalizace pro získání výhodných pozic ve vyhledávačích, u větších webů a větší konkurence je nutnost využít také off-page optimalizace. Hlavním off-page faktorem jsou zpětné odkazy na web. Vyhledávač předpokládá, že kvalitní weby odkazují opět na kvalitní weby. Off-page faktory tak neovlivňuje autor stránky přímo, ale musí se snažit získat odkazy z kvalitních webů. S tímto souvisí několik pojmů:

- *linkbuilding*
- *linkbaiting*
- *PageRank* společnosti Google
- *S-Rank* společnosti Seznam.cz

Linkbuilding se zabývá vytvářením zpětných odkazů na stránky. Způsobů existuje mnoho. Od registrace v katalozích, přes výměnu odkazů po nákup odkazů za peníze. Dalším způsobem je tzv. *linkbaiting* – vytváření takového obsahu, který přitáhne pozornost lidí a spolu s tím také odkazy z cizích webů.

```
<a href="/adresa-odkazu" title="popisek odkazu">anchor text</a>
```

Kód 2.3 – struktura odkazů v HTML

Kvalita off-page optimalizace je určena kvantitou, kvalitou (PageRank viz dále) a obsahem stránek, které na web odkazují. Obsahem se zde myslí anchor text odkazu a téma webu. Za nejhodnotnější považujeme odkazy, které vedou ze stejně tematicky zaměřených webů, které navíc v anchor textu obsahují naše klíčová slova. Síla anchor textu lze demonstrovat na fenoménu zvaném *Google Bomb*. Jedná se o SEO metodu na internetu většinou používanou k vyjádření negativních politických názorů. V praxi funguje na principu vytvoření velkého množství odkazů z webů zapojených osob s předem domluveným anchor textem. Klíčové slovo se tak vůbec nemusí objevovat na webových stránkách, na které se odkazuje, ale přesto bude na velice dobrých pozicích ve vyhledávači. Známy je například případ, kdy se na klíčové slovo *failure* (angl. selhání) pomocí Google Bomb protlačil na první pozici odkaz na oficiální stránky tehdejšího prezidenta USA Gerge W. Bushe.

PageRank je algoritmus, který určuje kvalitu jednotlivých stránek v indexovací databázi. PageRank ovlivňuje množství a kvalita webů, které na stránku odkazují. Pro ilustraci následuje vzorec z [2].

$PR(A) = (1-d)/m + d * (PR(T_1)/C(T_1) + \dots + PR(T_n)/C(T_n))$ <p>PR(A) - PageRank webu A</p> <p>T₁-T_n - stránky se zpětnými odkazy</p> <p>d - dampening faktor (určuje krok, tedy jak rychle vzorec konverguje)</p> <p>m - celkový počet zaindexovaných stránek</p> <p>C(T) - počet odkazů ze stránky T</p>

Kód 2.4: Výpočet PageRanku

V praxi vzorec znamená, že každá stránka, která někam odkazuje, předává část svého PR stránce odkazované. Čím více odkazů stránka obsahuje, tím menší hodnotu dále předává. Nejhodnotnější jsou tedy pro nás odkazy z webů, které dále již nikam neodkazují. Naopak stránky jako seznamy nebo katalogy mají nejmenší hodnotu. Zjistit ohodnocení URL nebo dokumentu můžeme například pomocí *Google Toolbar*. Ten zobrazuje hodnotu PR v rozmezí od 0 do 10. Tato hodnota je přepočtená, v originálu nabývá hodnot blízkým nule. PageRank se aktualizuje přibližně jednou za tři měsíce a jeho hodnota postupně konverguje v závislosti na použitých parametrech vzorce. Podobný systém hodnocení jako Google má také Seznam. Jeho verze se jmenuje *S-Rank*.

Předávání PageRanku se dá zamezit použitím atributu odkazu `ref` s hodnotou `nofollow`. Ten zamezí robotovi následovat odkaz a stane se tak prakticky neviditelným. Takováto technika je nutná pro použití v komentářových systémech a obecně v sekcích, jejichž obsah neovlivňujeme sami. Pokud se `nofollow` nepoužívá, zůstávají na stránkách otevřené dveře pro „kradení“ PageRanku. Toho využívají mimo jiné i automatictí roboti, kteří pak například do komentářů vkládají odkazy na jimi propagované stránky a tím získají cenný PageRank z naší stránky.

Dalšími faktory, které se dají považovat za off-page jsou geolokace a jazyk. Vyhledávač Google zohledňuje ve svých výsledcích aktuální polohu uživatele. Pokud například mají stránky jazykové mutace v češtině a angličtině a člověk bude z České republiky vyhledávat na google.co.uk název firmy, pak se ve výsledcích zobrazí automaticky česká mutace stránek. Z důvodu geolokace výsledků je důležité udržovat zápis stránek v katalogu Google Maps. Pokud uživatel vyhledává službu na specifickém místě, pak se ve výsledcích vždy nejdříve zobrazí relevantní zápisy z katalogu Google Maps. Počet klíčových slov, který lze do tohoto zápisu umístit je omezený a proto musíme volit pouze nejdůležitější fráze. Vyhledávač Seznam používá při řazení výsledků také geolokaci. Seznam zohledňuje v případě společností zápis v katalogu Firmy.cz a v případě ostatních objektů jsou směrodatné zápisy v Mapy.cz. Výsledky z různých regionů se ve výsledcích liší především na prvních pozicích.

Negativní vliv na off-page faktory může mít zapojení serveru do tzv. *link farm*. Jedná se o webové servery, jejichž jediným účelem je propojit odkazy navzájem pokud možno co největší počet webů. V důsledku toho se přerozdělí PageRank a odkazované weby získají lepší pozice. Takovéto servery jsou ale většinou robotem velmi brzy odhaleny a všechny stránky zapojené do sítě ztratí důvěryhodnost. Negativní hodnocení získají pouze weby, které na link farmy odkazují zpět (registrace do link farm je většinou zpětným odkazem podmíněna). Při SEO si tedy také musíme dávat pozor, na jaké weby odkazujeme.

3 Nástroje pro analýzu

Provádění optimalizačních kroků musí být postaveno na kvalitní analýze. Existuje několik nástrojů mj. i přímo od tvůrců vyhledávačů, které usnadňují provádění analýz. Následující výčet obsahuje především služby poskytované přímo vyhledávačem Google.

3.1 Vyhledávač Google

Prvním mocným nástrojem pro analýzu je samotný vyhledávač Google. Vyhledávání umožňuje použít několik přepínačů, které nám dají cenné informace o webu.

Za prvé je to přepínač `info`. Pomocí tohoto přepínače zjistíme, zda jsou stránky již zaindexovány robotem a dostaneme na výběr použití dalších přepínačů, které zmíním dále. Příklad použití ve vyhledávání:

```
info:www.ptravnik.cz
```

Kód 3.1 – přepínač `info`

Velice užitečným je přepínač `site`. Pomocí něho si můžeme nechat zobrazit všechny stránky zaindexované Google robotem. Přepínač nám dává možnost také omezit hledání na určitou subdoménu nebo podadresář, případně je z vyhledávání vyloučit. Příklad:

```
site:www.ptravnik.cz (všechny zaindexované stránky domény)
site:www.ptravnik.cz/blog (všechny stránky z adresáře blog)
site:www.ptravnik.cz -site:www.ptravnik.cz/blog (všechny stránky domény
kromě adresáře blog)
```

Kód 3.2 – přepínač `site`

Dalším přepínačem, který se při SEO analýze používá, je `link`. Tento přepínač zobrazí všechny stránky odkazující na daný web. Při vyhledávání se však bohužel zobrazují pouze weby s PageRankem vyšším než 3 nebo 4. Analýzu počtu zpětných odkazů se tak vyplatí dělat spíše pomocí vyhledávače `jyxo.cz` případně `yahoo.com`, kde existují také přepínače `link` a zobrazují většinou všechny existující odkazy. Příklad použití platný pro všechny tři vyhledávače naleznete v následujícím kódu:

```
link:www.ptravnik.cz
```

Kód 3.3 – přepínač `link`

Google dále poskytuje možnost zobrazit svou archivovanou verzi stránky, na které vyhledává. Používá se přepínač `cache`, který nám dává také možnost zobrazit textovou verzi

archivované stránky. Textová verze odpovídá tomu, jak web vidí robot Google. Pomocí `cache` také můžeme zjistit, kdy byla stránka naposledy zaindexována.

```
cache:www.ptravnik.cz
```

Kód 3.4 – přepínač `cache`

Posledním operátorem je `related`. Ten umožňuje vyhledávání podobných webů. V praxi pro SEO znamená vyhledání webů, z kterých bychom chtěli mít odkaz na naše stránky. Následuje příklad:

```
related:www.ptravnik.cz
```

Kód 3.5 – přepínač `related`

3.2 Google Analytics

Google Analytics [4] je webová aplikace od společnosti Google, která poskytuje uživatelům pokročilé statistiky návštěvnosti stránek. Kromě registrace musí uživatel pro plnou funkčnost aplikace vložit měřicí kód do zdrojového kódu sledovaných stránek. Tento měřicí kód generuje přímo Analytics a je to krátký úsek kódu v Javascriptu. Kód zavolá dále externí skript na serveru Google a posbírá všechny možné informace o aktuálním návštěvníkovi. Modifikací kódu a profilů v aplikaci se dají získat pokročilé statistiky.

Jednou ze statistik je například analýza klíčových slov, přes která se návštěvníci na stránku dostali. Statistiky jsou v Analytics úzce provázané a tak můžete zjistit například, kolik stránek si průměrně prohlédli uživatelé, kteří hledali specifické klíčové slovo. Analytics také nabízí statistiky používání odkazů, geolokaci návštěvníků apod.

3.3 Google Keyword Tool

Součástí souboru nástrojů pro internetovou reklamu Google AdWords je mj. nástroj Google Keyword Tool [5]. Nástroj usnadňuje výběr klíčových slov pro optimalizaci. Umí analyzovat obsah stránky a vybrat klíčová slova, která souvisejí s tématem webu. Případně po zadání vlastních klíčových slov nástroj nabídne další varianty frází. Pro všechna takto nabídnutá klíčová slova zobrazí statistiky – rozsah konkurence, celosvětový objem vyhledávání a místní objem vyhledávání. Pro SEO jsou tato velice cenná data, díky kterým se zmenšuje pravděpodobnost, že zvolená klíčová slova nebudou dostatečně efektivní. Keyword Tool poskytuje možnost provádět analýzu v jakýchkoliv jazykových mutacích a navíc se dá specifikovat cílové zařízení (počítače, mobilní zařízení apod.). Služba obsahuje také placené API, díky kterému mohou k datům přistupovat i externí servery.

3.4 Google Webmaster Tools

Tato služba [6] poskytuje možnost lépe provázat vlastní webové stránky s vyhledávačem Google. Pro využití služby musí uživatel mimo registraci prokázat přístup k webovým stránkám, které chce se službou svázat. To se provádí buď pomocí přidání kontrolní značky do hlavičky stránek nebo pomocí nahrání html souboru s kontrolním hash na server. Webmaster Tools umožňují nastavit soubor sitemap (viz 2.4.1). Nástroj poskytuje informace o návštěvách robota Google a přístupnosti stránek, umožňuje generovat `robots.txt` a další.

3.5 Validátory

Pro SEO je důležitá validace HTML a XHTML, která zvyšuje čitelnost webu pro roboty vyhledávačů. Pro maximální přístupnost by se měla také zajistit validita CSS. Pro tyto účely existuje mnoho nástrojů s více či méně detailní validací. HTML validátor [7] je nástroj asociace W3C pro kontrolu HTML a XHTML. Na stejném serveru lze také najít validátor kaskádových stylů. W3C určuje standardy pro používání HTML, XHTML, CSS a dalších webových technologií. Validátory konsorcia poskytují také API pro použití v externích službách.

4 Návrh aplikace pro SEO analýzu

Tato kapitola popisuje návrh aplikace, která by měla sloužit pro SEO analýzu webových stránek. Zde představeného modelu se bude držet kapitola 5 Implementace, nicméně mohou vzniknout drobné změny.

Aplikace bude on-line přístupná pomocí prohlížeče. Bude rozdělena na čtyři části – výběr klíčových slov pro následující analýzy, on-page analýza, off-page analýza a sledování pozic ve vyhledávačích. Cílem tohoto nástroje je dát člověku ucelený přehled o stavu a účinnosti SEO optimalizací určitého webového serveru. Výsledný nástroj by měl být určen pro urychlení a usnadnění práce člověka, který bude provádět SEO. Aplikace by měla být schopna udržet ve správě několik různých serverů (v řádu jednotek) a v rámci každého z nich několik klíčových slov (v řádu desítek).

4.1 Výběr klíčových slov

Výběr klíčových slov je první krok analýzy, bez kterého nelze plně zpřístupnit kroky další. V této sekci si uživatel zvolí klíčová slova, pro která se bude provádět analýza a sledování pozic. Uživatel musí mít možnost zadávat klíčová slova s několika parametry. Prvním parametrem je jazyková mutace slova – bude určovat, v jaké jazykové variantě vyhledávače Google se budou vytvářet statistiky o pozicích. Druhý parametr také souvisí se sledováním pozic a jedná se o periodu, v jaké se kontrola pozic bude provádět. Vybraná klíčová slova dále používají sekce On-page analyzátoru a Sledování pozic, tudíž bez provedení výběru slov nebudou plně přístupny. Off-page analýza s klíčovými slovy nepracuje a je vždy plně funkční.

4.2 On-page analyzátor

Po zadání klíčových slov bude mít uživatel možnost provést on-page SEO analýzu. Výsledky analýzy se budou ukládat na serveru a uživatel musí mít možnost požádat o provedení nové analýzy. Na internetu již existuje mnoho SEO analyzátorů, jejichž výsledek se prezentuje pomocí skóre. Tento model byl striktně zavrhnut proto, aby analyzátor nepodával zavádějící informace a bylo až přímo na uživateli, jak výsledky využije.

On-page analýza bude zahrnovat kontrolu zdrojového kódu. Zaprvé to bude analýza hlavičky, tedy výpis `title` a `description` a varování při příliš dlouhém textu. Pomocí API validátoru W3.org se provede kontrola validity kódu HTML/XHTML i CSS a zobrazí se případné chyby. Analyzátor pak zjistí poměr textu ke zdrojovému kódu na stránce. Dále se provede kontrola indexace robotem Google. Ta bude obnášet zjištění času poslední indexace a zobrazení celé stránky v textovém

formátu z cache vyhledávače Google. Dále je třeba informovat uživatele o počtu použitých frames, flashů a vložených objektů.

On-page analýza pokračuje analýzou obsahu. Z aktuální stránky se vypíše struktura nadpisů včetně jejich obsahu. Dále se zkontrolují všechny tagy `img` na přítomnost alternativního popisku. Závěrem analýzy obsahu je pak vyhledání všech zvolených klíčových slov v textu a zobrazení statistiky. Tzn., zda a kolikrát se vyskytují v titulku, popisku, nadpisech, obyčejných textech případně zvýrazněných textech atd.

4.3 Off-page analyzátor

Další sekci aplikace je off-page analýza. Ta by měla informovat uživatele o všech důležitých rancích stránky – tzn. PageRank, S-Rank a AlexaRank. Dále pak provést analýzu všech odkazů na stránce. Odkazy se rozdělí na interní a externí vedoucí na jiné stránky. Statistika odkazů by měla obsahovat kromě URL také titulek odkazu a anchor text pro lepší přehled. Nakonec to bude zjištění kompletního počtu odkazů, které na stránku vedou. Zde se zobrazí počet zpětných odkazů, tak jak o nich informuje Google.com. Pro informaci se dále použije počet zmínek o webu na službě Facebook.com a Twitter.com, které významně přispívají ke zvýšení důvěryhodnosti stránek u vyhledávačů.

4.4 Sledování pozic

Poslední funkcí aplikace je nástroj pro sledování pozic ve vyhledávači Google u klíčových slov, která si uživatel určí v první sekci. Kontrola pozic probíhá dle nastavení uživatele, kdy nejmenší periodou je jeden den. Uživatel musí také mít možnost zvolit si jazykovou mutaci Google, která se pro jednotlivá klíčová slova použije.

Výsledky sběru dat o pozicích si uživatel může zobrazit ve formě grafu. Uživatel musí mít možnost sám si zvolit časové období, pro které se graf vykreslí. Pro každé klíčové slovo by se měl vykreslovat samostatný graf kvůli přehlednosti. Pozice se mohou totiž pohybovat v řádu jednotek, desítek i stovek a pomocí grafu bychom nebyli schopni pojmout takové rozdíly.

5 Implementace

Tato kapitola se zabývá převodem návrhu aplikace do funkčního programu. Cílem implementace je vytvořit funkční aplikaci běžící na serveru a přístupnou pomocí webového prohlížeče. Aplikace musí být schopná automaticky spouštět průběžné kontroly stavu optimalizací, z čehož vyplývají specifické požadavky na technologie použité na straně serveru. Na straně klienta by pak měla aplikace využívat dynamického zasílání požadavků na server pro plynulou obsluhu a používání programu.

5.1 Technologie

K finální aplikaci se přistupuje pomocí webového prohlížeče. Funkčnost zajišťuje PHP (Hypertext Preprocessor) na straně serveru a JavaScript na straně klienta. Pro ukládání a předávání dat se používá kombinace XML (Extensible Markup Language) a MySQL.

5.1.1 JavaScript

JavaScript [9] je interpretovaný programovací jazyk, který se používá především pro tvorbu interaktivního obsahu webových stránek. Termín *Java* je v názvu zavádějící, jelikož jediným pojítkem těchto dvou jazyků je podobná syntaxe. JavaScript je zpravidla prováděn průběžně na straně klienta, což dává možnost měnit obsah stránky bez nutnosti načítání celého webu. Dá se použít také na straně serveru nebo při tvorbě pluginů pro různé oblíbené nástroje (např. Mozilla Firefox nebo Adobe Acrobat). Pro použití v naší aplikaci je JavaScript rozšířen o framework jQuery pro usnadnění práce s DOM (Document Object Model) a Google Chart Tools pro vykreslování grafů.

AJAX

AJAX [9] je pojem označující spojení několika technologií. Jedná se o XML, JavaScript, HTTP a HTML. Zkratka samotná znamená Asynchronous JavaScript and XML. Základním kamenem technologie je funkce `XMLHttpRequest` implementovaná v JavaScriptu, s jejíž pomocí se vytváří asynchronní požadavky na webový server. Odpovědi na takový požadavek jsou pak data ve formátu XML, JSON, HTML, případně čistý text nebo jiné vlastní formáty. Požadavky se dokončují nezávisle na sobě, a proto se musí při více dotazech implementovat synchronizační mechanismy.

jQuery

jQuery [8] je aplikační framework rozšiřující JavaScript o funkce usnadňující vytváření dynamických stránek. Jeho výhodou je také multiplatformnost – funkčnost je ve všech prohlížečích stejná. jQuery ušetří velké množství kódu a přitom je rychlé a jednoduché s ním manipulovat. Na webu navíc existuje množství pluginů postavených na tomto frameworku. jQuery implementuje několik funkcí

obsahujících XMLHttpRequest pro použití technologie AJAX. jQuery umožňuje jednotlivými funkcemi asynchronně zasílat požadavky typu POST i GET. Funkce automaticky parsuje odpovědi typu JSON, XML případně HTML. Příklad AJAX požadavku zapsaného v jQuery naleznete v následujícím kódu.

```
$.ajax({
  url: "test.php",
  context: document.body
}).done(function() {
  alert("Požadavek byl odeslán.");
});
```

Kód 5.1 – AJAXový požadavek zapsaný v jQuery

Google Chart Tools

Google Chart Tools [13] je knihovna funkcí v JavaScriptu a API poskytované zdarma firmou Google. Obsahuje funkce pro vykreslování grafů nebo demografických map, vytváření QR kódů apod. Tato práce využívá Chart Tools pouze pro vykreslování dynamických grafů. Vykreslování probíhá čistě v režii JavaScriptu a pomocí něho se také předávají data.

5.1.2 PHP

PHP [10] je skriptovací programovací jazyk, který se používá především pro webové aplikace, a sice pro použití na straně serveru. Momentálně se jedná o nejrozšířenější programovací jazyk pro webové aplikace. Zkratka znamená PHP: Hypertext Preprocessor. Jazyk je interpretovaný, nicméně existuje i jeho kompilovaná forma. Podporuje velké množství knihoven, mezi nimiž jsou funkce pro spolupráci s většinou databázových systémů. V tomto projektu byla využita databáze MySQL pro její snadnou dostupnost. Pro získání dat z externích serverů používá aplikace knihovnu cURL. Platformou pro provoz projektu je standardně Linux ve spojení s webovým serverem Apache a databází MySQL. Pro PHP existuje stejně jako pro JavaScript mnoho frameworků usnadňujících práci, nicméně tento projekt využívá pouze čisté PHP.

cURL

PHP podporuje knihovnu *libcurl*, která slouží pro komunikaci s externími servery. Díky cURL PHP rozšíření tak můžeme komunikovat na úrovni *http*, *https*, *ftp* a podobných protokolů. Čisté PHP sice také obsahuje funkce pro komunikaci s externími servery, nicméně jejich syntaxe je velmi nepřehledná a jsou při provádění více dotazů pomalejší než cURL. Vzhledem k množství dotazů na externí servery používá tedy tato aplikace cURL.

CRON

Cron slouží jako takzvaný plánovač úloh v operačních systémech unixového typu. Cron démon slouží především pro administrátory ke spouštění údržbových skriptů. Navíc ale může pomáhat také vývojářům internetových aplikací. Konfigurace démona probíhá pomocí utility *crontab*. V tomto nástroji se nadefinují časové intervaly spouštěných skriptů a jejich cesty. Pro správnou funkci tohoto projektu je nutné Cron používat pro automatické spouštění skriptu pro kontrolu pozic ve vyhledávačích.

MySQL

Databázový systém MySQL je oblíbený v kombinaci s programovacím jazykem PHP. Využívá základní syntaxe jazyka SQL se specifickými rozšířeními. MySQL nenabízí výkon jako například databáze Oracle, nicméně je poskytována pod bezplatnou licenci. Tento databázový systém byl pro aplikaci zvolen vzhledem k malému množství dat, které jsou potřeba ukládat do databáze. Pro data, která se ukládají pouze pro čtení, byl zvolen formát XML.

MySQL nabízí několik druhů úložných enginů, které ovlivňují způsob a strukturu, jakými jsou data fyzicky uložena. Volba se ale ve finále zužuje pouze na *InnoDB* a *MyISAM*. *MyISAM* je nejstarší systém pro uložení dat. Nabízí mnoho rozšiřujících funkcí, ale nepodporuje práci s cizími klíči. *InnoDB* je již v posledních verzích MySQL výchozím úložištěm a na rozdíl od *MyISAM* přímo podporuje cizí klíče. V tomto projektu jsou veškerá data ukládána pomocí *MyISAM* především kvůli jednoduchosti.

5.1.3 XML

Extensible Markup Language je standardizovaný značkovací jazyk. V aplikaci je využit za prvé pro ukládání dat určených pouze pro čtení. Za druhé pak pro komunikaci mezi PHP na straně serveru a JavaScriptem na straně klienta. XML bylo zvoleno především proto, že PHP i JavaScript obsahují předem vytvořené funkce pro parsování a práci s tímto formátem.

XML-RPC

XML-RPC (*Remote Procedure Call*) je protokol postavený na XML a slouží pro vzdálené volání procedur. PHP nabízí funkce, které automaticky zformátují XML-RPC požadavky dle zadaných parametrů. V projektu se tato technologie využívá pro komunikaci se vzdáleným serverem (konkrétně požadavky na Seznam.cz). Pro ilustraci XML-RPC požadavku slouží kód 5.2.

```
<?xml version="1.0"?>
<methodCall>
  <methodName>hodnotaPrvku</methodName>
  <params>
    <param>
      <value><int>1</int></value>
    </param>
  </params>
</methodCall>
```

Kód 5.2 – příklad XML RPC požadavku

5.2 Stavba aplikace

Při implementaci aplikace byly použity techniky dekompozice a objektově orientovaného programování. Aplikace je rozdělena do několika částí a spojena do jednoho celku pomocí kostry v PHP. Základní částí, která určuje funkčnost ostatních sekcí, je výběr klíčových slov. Zbylé tři sekce zajišťují samotnou funkčnost aplikace a jedná se o on-page analýzu, off-page analýzu a sledování pozic. Pro psaní veškerých zdrojových kódů byl použit nástroj *PSPad*.

5.2.1 Proxy

Třída sloužící pro komunikaci se vzdálenými servery. Tato třída se používá ve všech sekcích aplikace, které zajišťují funkčnost. Pro komunikaci nabízí dvě různé metody. První metodou je `get()`, která posílá obyčejný GET požadavek na URL určenou atributem. Požadavek se vytváří pomocí rozšíření `cURL`. Ovládání `cURL` probíhá pomocí nadefinování konstant a následnému spuštění pomocí `curl_exec()`. Výsledek se pak uloží do proměnné a rovnou zpracuje pro další použití v rámci DOM.

Druhou metodou třídy `Proxy` je `xmlrpc_post()`. Ta slouží k odesílání XML-RPC požadavků na vzdálený server. Na rozdíl od metody `get()` vyžaduje také druhý argument, a sice zformátovaný XML-RPC požadavek. Tento pak funkce doplní o xml hlavičku a odešle na adresu určenou prvním argumentem. Výsledek se opět nahraje do DOM.

5.2.2 Robot

Třída sloužící pro periodickou kontrolu pozic webů ve vyhledávači. Instanci této třídy je třeba pomocí nástroje `Cron` spouštět každý den. Její funkčnost je založena na spolupráci s několika dalšími třídami. Veškerá funkčnost je rozdělena do dvou metod. Metoda `GetPosition()` vysílá a

zpracovává požadavek na server Googlu. Metoda `Walkthrough()` zajišťuje smyčku, ve které se `GetPosition()` volá.

Walkthrough

Všechna data, se kterými pracuje třída `Robot`, se ukládají do MySQL databáze. Pro spolupráci s databází slouží třídy `Servers`, `Keywords` a `Positions`. Při spuštění robota metodou `Walkthrough()` se nejprve vytvoří instance třídy `Server`. Pomocí metody `Load()` se naplní pole všech serverů. Naplněné pole se prochází pomocí smyčky. V této smyčce se vytvoří instance třídy `Keyword` a naplní se opět metodou `Load()` klíčovými slovy k příslušnému serveru. Pro klíčová slova je pak volána druhá metoda robota – `GetPosition()`, jejímž atributem je identifikátor klíčového slova. Poté se vytvoří instance třídy `Position` a naplní se daty získanými metodou `GetPosition()` spolu s aktuálním datem. Data se ukládají pomocí metody `Save()`.

Get

Třída `Proxy` v případě robota zajišťuje komunikaci se serverem Googlu. Požadovanou URL na serveru googlu určuje jazyková mutace klíčového slova a klíčové slovo samotné. Tyto atributy se zakódují do URL a dále robot využívá `Proxy`. Konkrétně metodu `get()` pro komunikaci s externím serverem požadavkem typu GET. Odpověď serveru Google se pak zpracovává pomocí DOM. V DOM robot vyhledá nejdříve dle identifikátoru blok s výsledky vyhledávání a zde následně dle tagů vyhledá jednotlivé odkazy. Tyto odkazy se pomocí regulérního výrazu porovnávají s adresou serveru. Dle konfigurace robota probíhá vyhledávání až do x-té stránky s výpisem po y výsledcích. Výsledkem metody `GetPosition()` je pozice hledaného webu včetně čísla strany, na které se nachází.

5.2.3 Grafy

Vykreslování grafů probíhá pomocí Google Chart Tools. Data, která se budou vykreslovat, získává klient pomocí AJAXového dotazu na skript `getgraphdata.php`. JavaScript odesílá v rámci požadavku dva atributy, které určují, za jaké období chceme grafy vykreslit. Skript odpovídá na dotazy ve formátu JSON pro úsporu velikosti dat oproti XML.

5.2.4 Analyzátoři

Analyzátor je rozdělen do dvou sekcí – On-page a Off-page. Tyto jsou dále rozděleny do tří podsekcí. Všechny podsekcce jsou založeny na spolupráci se třídou `Proxy`. Analýzy se provádějí na straně serveru. Komunikaci se serverem a zobrazení výsledků zajišťuje JavaScript na straně klienta. Výměna dat probíhá pomocí XML. Výsledky analýz se navíc ve formě XML ukládají na lokální úložiště, aby byla přístupna i archivní data. Výsledky minulých analýz jsou přístupny uživateli dle

časů vykonání analýzy. Díky ukládání archivních dat ve stejném formátu, v jakém probíhá komunikace, stačí na straně klienta jedna funkce pro zpracování dat.

Uživatel má možnost zažádat o provedení nových analýz. V tomto případě JavaScript vyšle AJAXové požadavky na adresu příslušné analýzy. Následně se spustí analýza na straně serveru a klient bude čekat na odpovědi na všechny tři požadavky. Pokud uživatel žádá výsledek starého testu, pak server pouze přepošle obsah příslušného souboru s výsledky analýzy. Defaultně se při načtení stránky s analýzou načte výsledek posledního testu.

On-page analyzátor

On-page analýza sestává ze tří oddělených tříd. Třída `Own_Page` zajišťuje všechny testy, které pracují se zdrojovým kódem. Používá třídu `Proxy` pro načtení zdrojového kódu do DOMu, z kterého následně získává data. Získání obsahu hlavičky probíhá pomocí vyhledání meta tagů. Výpis struktury nadpisů se provádí pomocí regulérních výrazů. Dále se vyhledávají všechny obrázky a jejich atributy a detekuje se přítomnost rámců ve zdrojovém kódu. Jako poslední se provádí analýza textu. Pomocí třídy `Keywords` přistupuje analyzátor do databáze a následně pomocí regulérních výrazů vyhledává klíčová slova ve všech textech. Pro analýzu textů se rozlišuje umístění klíčových slov v nadpisech, hlavičce a obyčejném textu.

Třída `Validator` kontroluje validitu zdrojového kódu. Pomocí třídy `Proxy` zasílá GET požadavky na servery `validator.w3.org` pro kontrolu HTML, respektive `jigsaw.w3.org` pro validaci CSS. Odpověď ve formě XML se parsuje opět pomocí DOM.

Třída `Google_Index` slouží k získání informací o aktivitách Googlebota na analyzovaném serveru. Pomocí `Proxy` přistupuje do webcache archivu Google. Kompletní obraz stránky v archivu se pak uloží do odpovědi a extrahuje se datum poslední návštěvy indexovacího robota.

5.2.4.1 Off-page analyzátor

Off-page analýza se skládá stejně jako on-page analýza ze tří částí. První část zajišťuje třída `Ranks`. Ta získává hodnoty PageRanku, S-Ranku a AlexaRanku pro analyzovaný server. PageRank a AlexaRank se získává pomocí GET požadavku na příslušná API. S-Rank se získává pomocí XML-RPC požadavku na server Seznamu.

Třída `Outlinks` analyzuje opět samotný server. Pomocí regulérních výrazů a DOM extrahuje ze zdrojového kódu všechny odkazy. Odkazy se při analýze dělí dle URL na externí a interní.

Třída `Backlinks` spolupracuje se servery Google. Pomocí `Proxy` zasílá tři GET dotazy na server Google. Odpověď se parsuje pomocí DOM a regulérních výrazů a výsledkem je počet zpětných odkazů a počet zmínek na sociálních sítích Facebook.com a Twitter.com.

```
<?xml version="1.0"?>
<root>
  <html_val>
    <validity>true</validity>
    <errors>0</errors>
    <warnings>0</warnings>
  </html_val>
  <css_val>
    <validity>>false</validity>
    <errors>13</errors>
    <warnings>0</warnings>
  </css_val>
</root>
```

Kód 5.3 – ukázka výsledku testu validace ukládaného na serveru

5.2.5 XML

Pro ukládání výsledků analýz i pro komunikaci mezi analyzátořem a klientem se používá značkovací jazyk XML. Výsledkem všech analýz je asociativní pole, které se dále zpracovává pomocí funkce `Results_to_XML()`. Funkce převádí pomocí rekurzivní funkce asociativní pole do XML formátu (ukázka v kódu 5.3 výše). Výsledné XML následně vytiskne na standardní výstup jako odpověď AJAX požadavku a dále ho uloží do souboru pojmenovaný dle aktuálního data a času testu.

6 Testování

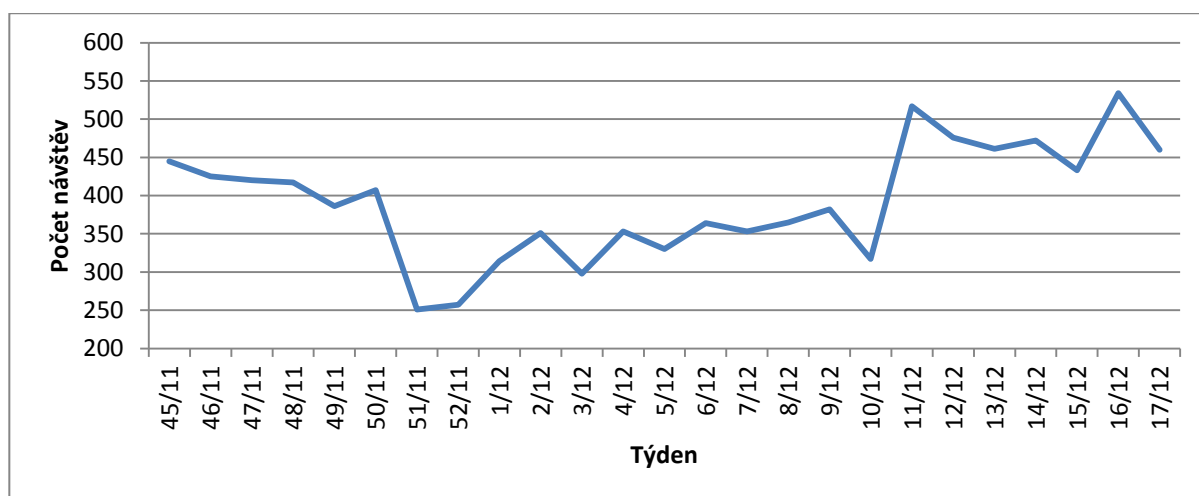
Pro účely testování se pro tuto práci podařila získat práva na editaci tří různě specifických webů. Weby se odlišují především svým stářím, ale jejich témata a region působnosti jsou velice podobná. Proto bylo možné objektivně porovnat efektivitu veškerých provedených optimalizací. Všechny servery pro testování jsou komerčního charakteru. V rámci spolupráce s vlastníky webu probíhala pouze diskuze k výběru klíčových slov. Ve všem ostatním se optimalizace řídily pouze pomocí dostupných zdrojů a následně také dle vytvořené aplikace.

- www.svornost.cz – web fungující již dlouhou dobu (přes deset let), určený především pro testování účinnosti on-page optimalizací
- www.studio-pure.cz – nový server (stáří jeden rok), vhodný pro porovnání váhy on-page a off-page optimalizací
- www.rezidencevysehrad.cz – středně starý server (5 let), vhodný pro testování optimalizací v různých jazykových mutacích

6.1 Volba klíčových slov

Volba klíčových slov je první a základní krok, který se při SEO provádí. Při špatném prvním kroku mohou všechny další optimalizace ztratit svůj účel, i když budou provedeny správně. Při testování se mimo jiné potvrdilo, jak důležitá volba klíčových slov je. I když se podařilo získat první pozice, tak se návštěvnost webu nemusela automaticky zvýšit. Toto platilo nejen u slov, které měly malou konkurenci a hledanost, nýbrž také u slov, která byla vyhledávána hojně.

Proces výběru klíčových slov vyžadoval podrobnou analýzu hledaností a konkurence v rámci jednotlivých slov a slovních spojení. Pro každý testovací server bylo nakonec vybráno kolem desítky klíčových slov. Jako zdroj informací při výběru klíčových slov sloužil nástroj Google Keyword Tool (kapitola 3.3). Při výběru byla snaha zajistit klíčová slova, která přinesou větší návštěvnost webu. Bohužel se toto nepodařilo dodržet u všech tří serverů. Při optimalizaci webu Rezidence Vyšehrad byla analýza klíčových slov evidentně provedena špatně. Se začátkem SEO zde klesla návštěvnost zhruba o 40% (týden 50/2011 viz Obrázek 6.1) pravděpodobně ztrátou pozic na klíčových slovech, která se na stránkách vyskytovala původně. Následně po dosažení prvních míst se návštěvnost vyšplhala na zhruba 85% původních hodnot. Proto se musely zvolit nové fráze a provádět optimalizace znovu od začátku. Výsledná návštěvnost se na konci optimalizací vyšplhala na zhruba 125% původních hodnot i za přispění aplikace vytvořené v rámci tohoto projektu. Na serveru neprobíhaly žádné reklamní kampaně. Veškerá návštěvnost pochází z vyhledávačů, pouze část jsou vracející se uživatelé (jejich počet se ale v průběhu optimalizací neměnil).



Obrázek 6.1 - Měření návštěvnosti www.rezidencevysehrad.cz (Zdroj: Google Analytics [4])

6.2 Efektivita optimalizací

V průběhu vypracování této práce byly průběžně testovány SEO techniky nastíněné v kapitole 2. Efektivita jednotlivých optimalizací byla velmi odlišná.

6.2.1 On-page

Velká výhoda on-page optimalizací oproti off-page byla v okamžitém odrazu ve výsledcích vyhledávání. Stačilo sledovat návštěvy robota Google a po zaindexování se mohla ihned kontrolovat zpětná odezva v pozicích webu. Jako nejdůležitější se ukázaly tři faktory:

- titulek stránky
- meta tag description
- nadpis 1. úrovně

Základním stavebním kamenem SEO tedy bylo rozmístit klíčová slova dle jejich důležitosti a především konkurence mezi tuto skupinu HTML tagů. Titulek stránky má naprosto fatální dopad na pozice – největší ze všech testovaných faktorů. Pro ilustraci příklad v Tabulka 6.1 pro klíčová slova optimalizovaná na podsekcích daných serverů. Ve srovnání s ostatními způsobila vždy změna titulku největší skok v pozicích ať už výš nebo níž.

	Počet výskytů klíčového slova	pozice
Hotel Praha 4 - www.rezidencevysehrad.cz		
Před optimalizací	1(1x nadpis text)	186
Po změně	3(1x titulek, 1x nadpis h1, 1x text)	21
firemní večírky Praha - www.svornost.cz		
Před optimalizací	2(1x nadpis h1, 1x text)	59
Po změně	3(1x titulek, 1x nadpis h1, 1x text)	5
jóga Praha 5 - www.studio-pure.cz		
Před optimalizací	1(1x text)	64
Po změně	3(1x titulek, 1x description, 1x text)	16

Tabulka 6.1 - optimalizace titulku stránek v podsekcích uvedených serverů

Meta tag `description` a nadpis 1. úrovně měly zhruba stejné účinky na postup pozic. Jejich výhodou oproti titulku stránky je větší počet znaků. Tag `description` je zkracován na prvních 155 znaků (titulek pouze na 65), nadpis 1. úrovně pak nemá omezenou délku, avšak je nutné ho používat s mírou a proto se nikdy nevolil nadpis delší než 100 znaků. V praxi se osvědčilo používat `description` pro klíčová slova ze spodní vrstvy long-tail (Obrázek 2.3).

Druhou skupinou on-page faktorů byly ty, u kterých se očekával menší vliv na pozice. Konkrétně se pak jedná o tyto:

- URL adresa
- nadpisy nižších úrovní
- samotný text
- alternativní popisky

Změny URL adresy byly testovány na serveru `www.studio-pure.cz`, kde byl vytvořen systém pro změnu adresy, včetně použití HTTP kódu *301 Moved Permanently* pro rychlejší reakci vyhledávače. Bylo provedeno několik editací URL adres, nicméně nebyla zaznamenána jakákoliv změna pozic na základě těchto kroků. Změna se promítla pouze v zobrazovaném výsledku vyhledávání, kde je adresa obsahující klíčová slova zvýrazněna. Dalo by se tedy uvažovat o vlivu na uživatele, který je nabádán kliknout na takto zvýrazněné zápisy.

Samotný text stránek včetně nadpisů nižších úrovní má kupodivu velmi malý vliv na výsledné pozice. Při změně textu a nadpisů byly pozorovány změny maximálně o jednu až dvě pozice (Tabulka 6.2).

	Počet výskytů klíčového slova	pozice
apartments Prag zentrum - www.rezidencevysehrad.de		
Před optimalizací	1(1x titulek)	39
Po změně	2(1x titulek, 1x nadpis h2, 1x text)	38
hotel Praha 9 - www.svornost.cz		
Před optimalizací	2(1x description, 1x nadpis h1)	15
Po změně	4(1x titulek, 1x nadpis h1, 2x text)	12
taneční kurzy Praha - www.studio-pure.cz		
Před optimalizací	2(1x titulek, 1x description)	23
Po změně	5(1x titulek, 1x description, 1x nadpis h2, 2x text)	24

Tabulka 6.2 - optimalizace textu stránek

Z hlediska on-page optimalizací byl dále sledován počet zaindexovaných stránek. S tímto faktorem se nevyskytly žádné problémy, především díky menšímu rozsahu testovaných serverů. Díky častému provádění optimalizací se perioda návštěv robota ustálila na zhruba 6 dnech.

Nejmenší vliv na změnu pozic byl pozorován u meta tagu `keywords` a alternativních popisků obrázků. U `keywords` se potvrdilo tvrzení z [3], že není vyhledávačem Google bráno v potaz. Alternativní popisky a přejmenování obrázků bylo testováno na www.studio-pure.cz. Toto pomohlo k zviditelnění obrázků na obrázkovém vyhledávači Google, nicméně generovaná návštěvnost byla pro server minimální až žádná.

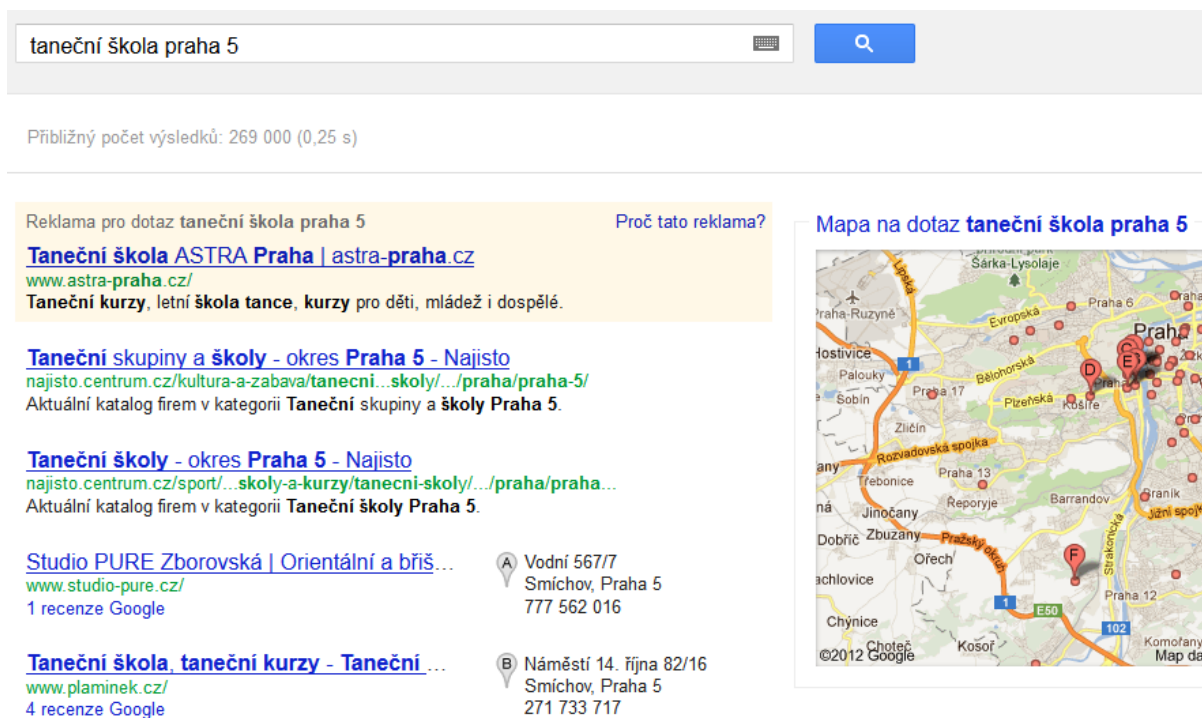
6.2.2 Off-page

U off-page faktorů nástává problém s metrikou pro měření úspěšnosti. Dosáhnout zvýšení PageRanku se podařilo pouze u jediného webu, a sice www.rezidencevysehrad.cz. Web nebyl registrován do katalogů ani nebyly vytvářeny zpětné odkazy. Jediným zde provedeným krokem v rámci off-page bylo prolinkování všech jazykových verzí webu mezi sebou. Jazykové mutace jsou totiž hostovány každá na své vlastní doméně (čtyři domény s koncovkami *cz*, *de*, *ru* a *com*). PageRank se tímto krokem zvýšil z hodnoty 3 na 4, což je ve srovnání s weby stejného rozsahu lehce nadprůměrná hodnota. V rámci multijazyčných stránek se tedy dá doporučit umístění jednotlivých mutací na samostatné domény.

Server www.studio-pure.cz byl z hlediska off-page optimalizován pomocí vzájemného prolinkování se sesterským webem (PageRank 3) se stejným tématem. Vzhledem k tomu, že tyto optimalizace se neprojeví okamžitě, nebylo možné určit, zda postupy pozic v té době ovlivnily také tyto off-page kroky. Na posledním optimalizovaném serveru se provedla registrace do tematických katalogů, nicméně zde žádné změny pozorovány nebyly.

Na druhou stranu dobře měřitelnou z hlediska off-page se stala optimalizace zápisů v katalogu Google Maps a vliv geolokace na výsledky vyhledávání. Příklad výsledků vyhledávání, kdy se používají výsledky z Google Maps naleznete na Obrázek 6.2. V příkladu vidíte výsledky vyhledávání klíčového slova „*taneční škola praha 5*“. Takováto forma výsledků vyhledávání se

používá při zadání klíčových slov obsahujících službu/zařízení spolu s názvem lokace (může být přesnější např. obsahovat přímo název ulice) nebo bez lokace (zde se použije geolokace a určí město, ve kterém se nacházíte). Na prvních místech se pak zobrazí soubor výsledků z hledání na Google Maps spolu s mapou. Algoritmus řazení výsledků v takových případech se z dostupných zdrojů nepodařilo zjistit, avšak po vytvoření zápisu v katalogu se klíčová slova zobrazující se do té doby na prvních dvou stranách přesunuly na první stranu do souboru výsledků z mapy.



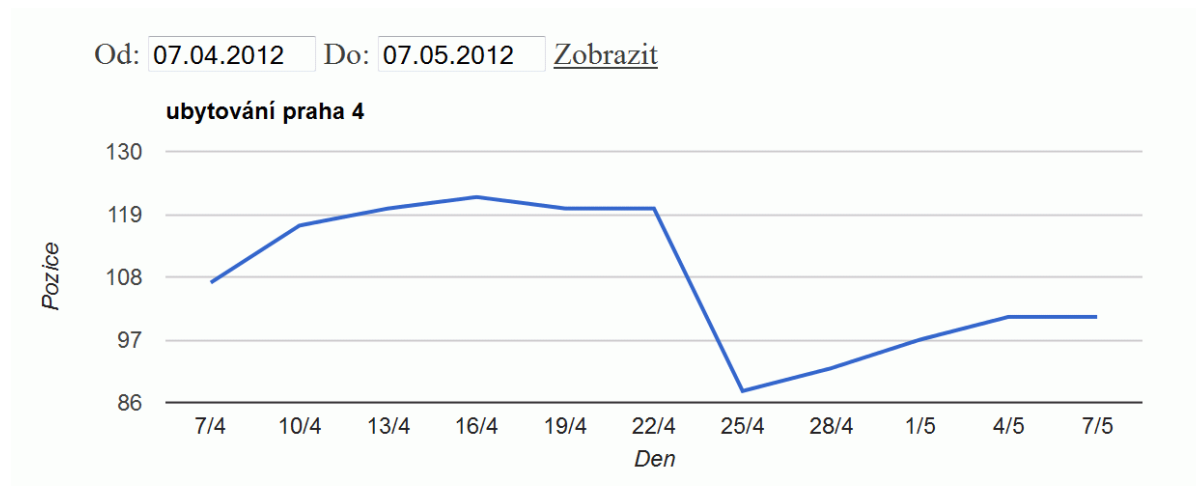
Obrázek 6.2 - výsledky vyhledávání klíčových slov obsahujících místo

V rámci optimalizace pro geografické oblasti je tedy třeba především provést registraci do Google Maps a v zápisu poté zajistit správné atributy. V průběhu testování byly zápisy měněny zhruba čtyřikrát dle aktuálních pozic klíčových slov. Tento atribut byl mj. kontrolován také na Seznam.cz, kde bylo zapotřebí zápisu v přidruženém katalogu Firmy.cz, případně pak přímo na Mapy.cz.

6.3 Použití aplikace

Na testování finální aplikace byl určen jeden měsíc. Během testování se vývojový cyklus několikrát vrátil do fáze implementace kvůli rozšíření funkčnosti, případně také opravě chyb. Aplikace byla testována za použití stejných serverů a klíčových slov jako v kapitole 6.2. Hlavní metrikou použitou pro hodnocení finální aplikace byl čas ušetřený oproti použití standardních metod SEO analýzy.

On-page analyzátor sleduje všechny faktory potřebné pro úspěch SEO optimalizací. Sledované faktory vychází ze závěrů vlastního testování a zkušeností získaných v celém průběhu vypracování této práce. Výhodou používání analyzátoru je především okamžitý přehled o stavu. Odpadá tak nutnost procházení a zkoumání zdrojového kódu stránek, kde je lehce možné mnoho nedostatků přehlédnout. On-page analyzátor pomáhá především ke zhodnocení výskytu klíčových slov. Podává report zahrnující všechna důležitá místa na stránce, která jsou pro optimalizované fráze důležitá. Informace z validátoru šetří čas (manuálně zabere validace zhruba pět minut) a pomáhá rychle zareagovat na nekorektnost zdrojového kódu. Nicméně validita se ukázala v průběhu optimalizací jako okrajový faktor. Většinou se tak na tento faktor přihlíželo, až pokud byl počet chyb opravdu velký. Jako velice praktická se ukázala sekce *Indexace* v rámci on-page analýzy. Údaj o poslední indexaci byl užitečný při sledování dopadů on-page změn. Textová podoba stránky v cache paměti Google je pak výborná pro revizi nejdůležitějšího obsahu stránky, tedy textu, a opět pomáhá s udržováním optimálního množství klíčových slov.



Obrázek 6.3 - výstup nástroje pro sledování pozic

Off-page analyzátor zaostává ve srovnání s přínosem on-page analyzátoru. Sledování SRanku a PageRanku se ukázalo jako nepříliš důležité, jelikož hodnoty se aktualizují pouze několikrát za rok. Nicméně jsou tyto hodnoty stále relevantní a jejich hodnota by měla být SEO optimalizátorovi známa. Užitečným prvkem pro sledování účinnosti SEO byl AlexaRank. Jelikož je založen na návštěvnosti webových stránek, tak ukazuje skutečné výsledky SEO optimalizací. Navíc je neustále aktualizován, a proto bylo sledování AlexaRanku jedním z nejčastějších důvodů spuštění off-page analyzátoru. Analýza zpětných odkazů byla bohužel založena na údajích z vyhledávače Google. Ten při výpisu zohledňuje pouze zpětné odkazy s vysokou relevancí, nicméně pro SEO je důležité získat informace i o těch méně kvalitních. Naproti tomu užitečná se ukázala analýza zmínek na sociálních sítích pomocí Google. Odkazy především ze sítě Facebook generují pro web poměrně velkou relevancí a navíc také návštěvnost. Proto je nutné hodnoty počtu zmínek ze sociálních sítí sledovat a

snažit se je neustále zvyšovat. Přehled odkazů na závěr off-page analýzy dává dobrý přehled o struktuře webu a o tom, komu dále poskytneme náš PageRank pomocí externích odkazů.

Nástroj pro sledování pozic (příklad výstupu na Obrázek 6.3) pomáhá efektivně měřit úspěšnost SEO optimalizací. Oproti ruční kontrole pozic, která byla prováděna po dobu vypracování práce, ušetří aplikace velké množství času. Při dvaceti až třiceti klíčových slovech trvala jedna kontrola pozic zhruba půl hodiny a déle. Z časového důvodu tak byla prováděna kontrola pouze jednou za měsíc. Používání aplikace ale znamenalo v posledním měsíci velkou úsporu času a větší přehled, jelikož pozice byly kontrolovány denně.

Čas, který aplikace SEO optimalizátorovi celkově ušetří je relativní a souvisí s povahou optimalizovaného serveru. U statických prezentací, kde se obsah serveru mění pouze sporadicky je časová úspora menší, jelikož se optimalizace provádí méně často. Aplikace nicméně vytvoří optimalizátorovi okamžitý přehled bez nutnosti procházet zdrojový kód a navíc hned zjistí, jak si klíčová slova stojí v konkurenci. V časové tísně při ruční kontrole se navíc většinou prochází pouze ty nejdůležitější faktory a snadno se tak například přehlédne, že se objevily nové externí odkazy, které ubírají stránkám PageRank.

7 Závěr

Cílem práce bylo v první fázi vyzkoušet různé SEO techniky a na základě nich pak ve druhé fázi vytvořit nástroj pro usnadnění provádění SEO. Pro celou práci bylo zvoleno zaměření na vyhledávač Google. V první fázi se na testovaných serverech za pomoci SEO technik podařilo dosáhnout na první strany výsledků ve vyhledávači. Optimalizace a testování bylo prováděno na komerčních serverech mezi skutečnou konkurencí, a proto se aplikované SEO metody dají považovat za úspěšné. Na základě získaných zkušeností s optimalizacemi pak byla vytvořena aplikace sloužící pro usnadnění a zefektivnění SEO.

K aplikaci se přistupuje pomocí webového prohlížeče a obsahuje SEO analyzátor a nástroj pro sledování pozic ve vyhledávači. Byla úspěšně testována k běhu na lokálním serveru i bezplatném webhostingu. Může tedy být přístupná jak pro lokální prostředí, tak sdílena přes internet. Při testování aplikace spravovala SEO analýzu pro několik menších komerčních webů se zhruba třiceti klíčovými slovy, což odpovídá zamýšlenému zatížení. Aplikace především zajistila úsporu času oproti manuálnímu provádění kontrol a analýzy. Největším přínosem aplikace je přehled všech potřebných faktorů pro SEO na jednom místě. Pro získání takovéto funkčnosti by se jinak muselo používat několik různých nástrojů.

Velký potenciál pro další vylepšení aplikace nabízí sekce s volbou klíčových slov. Data, která by usnadnila výběr klíčových slov, jsou však špatně dostupná. Z dostupných zdrojů byla nalezena pouze jedna použitelná báze dat přístupná vzdáleně, a sice Google Keyword Tool. Toto API je bohužel placené a v rámci projektu tedy nebylo možné ho zakomponovat. Další možností rozšíření by byla integrace nástroje pro sledování návštěvnosti webu. Ve spojení s nástrojem sledování pozic by uživatel měl možnost vytvořit si lepší přehled o účinnosti SEO.

SEO je nikdy nekončící proces, každý server by se měl o svůj obsah starat průběžně a s ohledem na praktiky uvedené v této práci. Jedině tak lze zajistit úspěšnost serveru v obrovské konkurenci, která v dnešní době na internetu panuje.

Literatura

- [1] Webová stránka: RobotsTXT.org [online]. Prosinec 2011 [cit. 2011-12-30]. Dostupné z WWW: <<http://www.robotstxt.org/>>
- [2] JANOVSKEÝ, Dušan. *Jak psát Web* [online]. Prosinec 2011 [cit. 2011-12-30]. Google PageRank, vysvětlení a odpovědi. Dostupné z WWW: <<http://www.jakpsatweb.cz/seo/pagerank.html>>
- [3] KUBÍČEK, Michal. *Velký průvodce SEO*. Vyd. 1. Brno: Computer Press, 2008, 318 s. ISBN 978-80-251-2195-5.
- [4] Webová stránka: Google Analytics [online]. Google Inc. [cit. 2011-12-27]. Dostupné z WWW: <<http://www.google.com/analytics/>>
- [5] Webová stránka: Keyword Tool - Google Adwords [online]. Google Inc. [cit. 2011-12-27]. Dostupné z WWW: <<http://adwords.google.com/select/KeywordToolExternal>>
- [6] Webová stránka: Webmaster Tools - Google [online]. Google Inc. [cit. 2011-12-27]. Dostupné z WWW: <<http://www.google.com/webmasters/tools/>>
- [7] Webová stránka: The W3C Markup Validation Service [online]. World Wide Web Consortium. [cit. 2011-12-27]. Dostupné z WWW: <<http://validator.w3.org/>>
- [8] MARGORÍN, Marián. *JQuery bez předchozích znalostí*. Vyd. 1. Brno: Computer Press, 2011, 253 s. ISBN 978-80-251-3379-8.
- [9] ODELL, Den. *JavaScript: průvodce programováním ajaxových aplikací*. Vyd. 1. Brno: Computer Press, 2010, 368 s. ISBN 978-80-251-2733-9.
- [10] GILMORE, W. *Velká kniha PHP 5 a MySQL: kompendium znalostí pro začátečníky i profesionály*. Nové, 3. vyd. Brno: Zoner Press, 2011, 736 s. ISBN 978-80-7413-163-9.
- [11] DOMES, Martin. *SEO: jednoduše*. Vyd. 1. Brno: Computer Press, 2011, 141 s. ISBN 978-80-251-3456-6.
- [12] Webová stránka: Audit návštěvnosti TOPlist [online]. TOPlist s.r.o. [cit. 2012-04-27]. Dostupné z WWW: <<http://www.toplist.cz>>
- [13] Webová stránka: Google Chart Tools – Google Developers [online]. Google Inc. [cit. 2012-04-27]. Dostupné z WWW: <<https://developers.google.com/chart/>>

A Uživatelská příručka

Návod pro instalaci

K aplikaci se přistupuje pomocí webového prohlížeče a její jádro běží ve skriptovacím jazyce PHP. Ten potřebuje pro svůj chod nainstalovaný server.

- První možností je pronajmout si webhosting u některé z četných firem působících na internetu a aplikaci tak mít přístupnou kdekoliv přes web
- Druhou možností je pak nainstalovat si webový server na lokální počítač (např. XAMPP pro Windows <http://www.apachefriends.org/en/xampp.html>)

Pro správný běh aplikace je nutné, aby webový server podporoval následující:

- PHP verze 5 a vyšší s povolenými knihovnamy cURL, mysql a xmlrpc
- MySQL databáze verze 4 a vyšší
- CRON

Pokud splňujete všechny požadavky, můžete začít s instalací a nastavením aplikace.

1. Nejprve rozbalte obsah instalačního archivu.
2. V prohlížeči zadejte cestu k souboru install.php.
3. Pokud server splňuje všechny požadavky, objeví se formulář pro nastavení připojení k databázi (je potřeba mít k dispozici prázdnou databázi s kódováním utf-8). Nejčastějším problémem při instalaci jsou nedostatečná práva pro editaci souborů.
4. Pokud instalace proběhla úspěšně, pak zbývá pouze nastavit periodické spuštění skriptu robot.php každý den pomocí CRON (zde se postupy liší v závislosti na použitém webhostingu či lokálním serveru).

Návod pro používání

Aplikace je přístupná pomocí webového prohlížeče. Analyzátor je rozdělen do čtyř sekcí:

1. Výběr klíčových slov
2. On-page analýza
3. Off-page analýza
4. Sledování pozic

Při prvním spuštění aplikace je nejprve potřeba přidat servery, které se budou sledovat. To se dá provést přímo na úvodní straně pomocí funkce „Přidat server“. Většina prvků v aplikaci reaguje okamžitě bez nutnosti znovu načítat stránku. Po přidání serverů pak zvolíte jeden, s kterým chcete aktuálně pracovat.

Výběr klíčových slov

Po volbě serveru Vás aplikace zavede do sekce klíčových slov. Zde je nutné vyplnit klíčová slova, jejichž pozice se posléze budou sledovat a která budou použita pro on-page analýzu. Při vkládání klíčových slov je potřeba zvolit jejich atributy pro sledování pozic.

- Perioda – slouží pro určení periody, s jakou se budou pozice kontrolovat. Hodnota je určena ve dnech a nejmenší možná je jeden den. Většinou není potřeba sledovat pozice s takovou frekvencí, a proto je dobré např. u méně důležitých slov zvolit periodu menší a ulevit tak zátěži serveru.
- Mutace – jazyková mutace vyhledávače Google, na které se má klíčové slovo sledovat. Aplikace nabízí volby cz (google.cz), uk (google.co.uk), de (google.de) a com (google.com).

WWW.REZIDENCEVYSEHRAD.CZ SEO ANALYZÁTOR

Výběr serveru | Klíčová slova | On-page analýza | Off-page analýza | Sledování pozic

Výběr klíčových slov

klíčové slovo	perioda [d]	mutace	
ubytování praha 4	1	cz	Upravit Smazat
krátkodobý pronájem apartmánu v praze	1	cz	Upravit Smazat
hotel praha 4	1	cz	Upravit Smazat
hotely praha 4	1	cz	Upravit Smazat
Praha luxusní apartmány	1	cz	Upravit Smazat
luxus apartment Prag	1	de	Upravit Smazat
mietwohnung Prag	1	de	Upravit Smazat
ferienwohnung Prag	1	de	Upravit Smazat
hotel Prag centrum	1	de	Upravit Smazat
serviced apartments Prague	1	uk	Upravit Smazat
luxury Prague apartments	1	uk	Upravit Smazat
apartments Prague	1	uk	Upravit Smazat
hotel Prague 4	1	uk	Upravit Smazat
<input type="text"/>	<input type="text" value="1"/>	cz ▾	Přidat

Aplikace vznikla v rámci BP SEO - Optimalizace pro vyhledávače na FIT VUT, Brno
Autor: Petr Trávník 2011-2012

Obrázek 1 – sekce výběru klíčových slov

On-page analýza

Pokud jste již zvolili klíčová slova, můžete naplno využít on-page analyzátor. Provádí analýzu serveru, se kterým se aktuálně pracuje. Veškeré ovládací prvky zde opět fungují okamžitě bez nového načítání stránky. Analýza je rozdělena na tři části, jejichž výsledky se ukládají jednotlivě na serveru ve formátu XML (ve složce tests/server). Při otevření sekce analyzátoru se vždy načte výsledek poslední analýzy a uživatel má dále přístupny výsledky všech předchozích testů v menu v horní části stránky.

- Obsah – zobrazí metainformace, poměr zdrojového kódu k textu, vypíše strukturu nadpisů, spočítá výskyt jednotlivých klíčových slov v prvcích na stránce, zkontroluje, zda stránka

nepoužívá vložený obsah (rámy, objekty, apod.) a vypíše veškerou strukturu obrázků na stránce s kontrolou, zda všechny obsahují alternativní popisky. Při nalezení problémů jako dlouhý titulek stránky nebo například chybějící alternativní popisky budou takovéto prvky označeny a zobrazí se upozornění. SEO je velice subjektivní pro každý internetový server, a proto analyzátor může strojově vyhodnotit pouze část prvků stránky. Pro celkové vyhodnocení je pak nejlépe informovat se v dalších zdrojích, případně přečíst bakalářskou práci, v rámci které byla tato aplikace vytvořena.

- Validita stránky – kontrola validity zdrojového kódu dle standardů W3C. Pro maximální přístupnost stránek pro indexovací roboty by měl být (X)HTML i CSS kód validní.
- Indexace – zobrazí informaci, kdy byly naposledy stránky navštíveny robotem vyhledávače Google a zobrazí stránky tak, jak je vidí robot. Tato analýza je velice důležitá a vyžaduje další pozornost člověka, který bude provádět následné SEO optimalizace.

PROVÉST KOMPLETNÍ ANALÝZU

Obsah	Historie testů	Indexace
	Validita	
04.05.2012 09:57:09	04.05.2012 09:57:13	04.05.2012 09:57:08
02.05.2012 16:25:42	02.05.2012 16:26:01	02.05.2012 16:25:37
02.05.2012 14:57:59	02.05.2012 14:58:07	02.05.2012 16:01:47
02.05.2012 09:50:53	02.05.2012 09:50:59	02.05.2012 14:57:58
19.04.2012 13:16:39	19.04.2012 13:16:43	02.05.2012 09:50:53
19.04.2012 13:15:29	19.04.2012 13:15:33	19.04.2012 13:16:39
19.04.2012 13:14:06	19.04.2012 13:14:10	13.04.2012 13:13:01
19.04.2012 13:11:14	19.04.2012 13:11:18	13.04.2012 13:11:39
19.04.2012 13:05:16	19.04.2012 13:05:20	13.04.2012 13:09:29
19.04.2012 09:59:49	19.04.2012 09:59:52	13.04.2012 13:09:19

Obsah stránky

Metainformace
 Titulek stránky: Hotel Praha 4 - Rezidence Vyšehrad
 Popisek stránky: Hotel Praha 4, Praha luxusní apartmány, krátkodobý pronájem apartmánu v Praze, Ubytování Praha 4
 Klíčová slova: ubytování, praha 4, hotel
 Poměr text/zdrojový kód: 0.17

Struktura nadpisů
 h2 - Online rezervace
 h1 - Luxusní apartmány se službami hotelu Praha 4
 h2 - Speciální nabídka
 h3 - Slevový voucher na thajskou masáž
 h3 - Má Vaše společnost potřebu ubytování na více než 50 nocí ročně?
 h2 - Nadcházející akce
 h3 - Volkswagen Prague Marathon PIM 2012
 h3 - IFA - Global Conference on Ageing
 h3 - ENS 2012 – 22nd Meeting of the European Neurological Society
 h3 - 25th IUA Congress of International Union of Agiology
 h3 - EUROGIN 2012
 h3 - 4th EuCheMS Conference 2012

Obrázek 2 – sekce on-page analýzy

Off-page analýza

Off-page analýza není závislá na volbě klíčových slov. Funkčnost je stejná jako u on-page analyzátoru, změna je ale v sekcích, které off-page analýza obsahuje.

- Hodnocení webu – v této části se vypisují hodnocení (tzv. ranky) z různých serverů. Konkrétně se jedná o PageRank vyhledávače Google, SRank vyhledávače Seznam a AlexaRank, který je založený na odhadované návštěvnosti serveru.
- Zpětné odkazy – vypisuje se počet zpětných odkazů, které na stránky vedou z internetu. Data pocházejí z vyhledávače Google, který zde zohledňuje pouze servery s vysokým PageRankem (ty jsou zároveň pro SEO nejdůležitější). Dále obsahuje také analýzu oblíbenosti na sociálních sítích Facebook a Twitter, které tvoří důležitou část zpětných odkazů pro SEO.
- Odkazy – tato část vypisuje veškeré odkazy, které se na stránce vyskytují. Odkazy jsou rozděleny na interní (v rámci webu) a externí (vedoucí do internetu). U odkazů se dále vypisuje atribut `title` a `anchor text` (tj. text odkazu, v případě obrázkového odkazu obsahuje hodnotu „Obrázek“).

PROVĚST KOMPLETNÍ ANALÝZU

Historie testů

Hodnocení	Odkazy	Zpětné odkazy
02.05.2012 16:03:39	02.05.2012 09:50:04	02.05.2012 16:03:39
02.05.2012 09:50:04	13.04.2012 13:15:50	02.05.2012 09:50:05
13.04.2012 13:15:50	13.04.2012 11:20:22	13.04.2012 13:15:50
13.04.2012 11:20:21	13.04.2012 11:17:59	13.04.2012 11:20:22
13.04.2012 11:17:59	13.04.2012 11:13:04	13.04.2012 11:18:00
13.04.2012 11:13:03	13.04.2012 11:12:19	13.04.2012 11:13:04
13.04.2012 11:12:22	13.04.2012 11:11:55	13.04.2012 11:11:56
13.04.2012 11:11:55	13.04.2012 11:11:15	12.04.2012 16:49:24
13.04.2012 11:11:14	13.04.2012 11:10:39	12.04.2012 16:47:05
13.04.2012 11:10:39	12.04.2012 16:49:27	12.04.2012 16:46:15

Hodnocení webu

Google PageRank: 4
Seznam SRank: 30
AlexaRank: 6218154

Zpětné odkazy

Zpětných odkazů: 21
Zminky na www.facebook.com: 67
Zminky na ww.twitter.com: 0

Odkazy

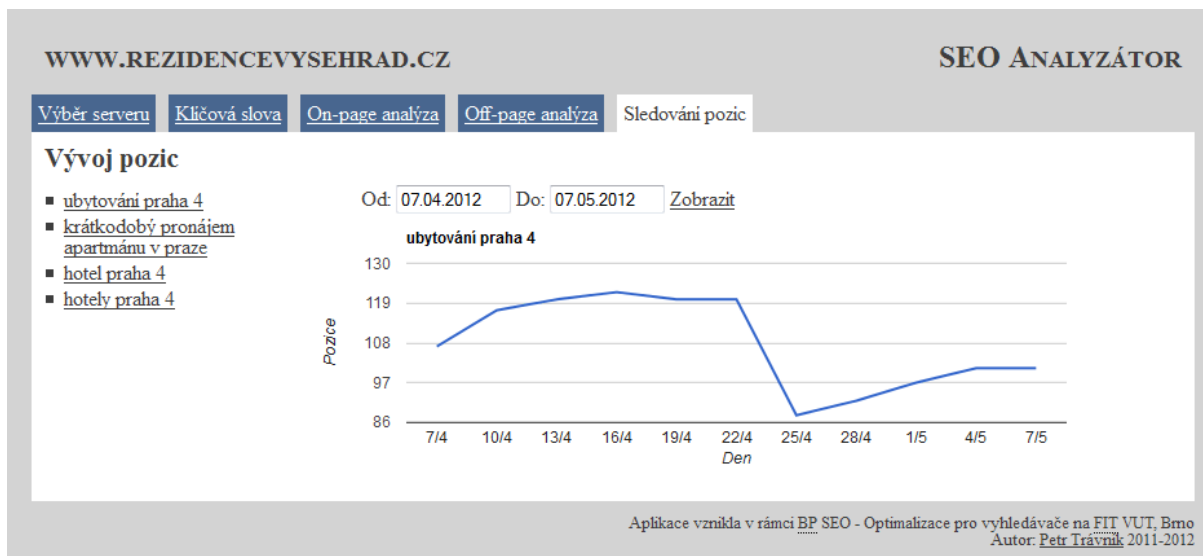
Interní odkazy

Adresa	Title	Anchor text
http://www.rezidencevysehrad.cz/		Obrázek
http://www.rezidencevysehrad.cz	Rezidence Vyšehrad	Obrázek
http://www.rezidencevysehrad.cz/uvod.html	úvod	úvod
http://www.rezidencevysehrad.cz/sluzby.html	Služby	Služby
http://www.rezidencevysehrad.cz/apartmany.html	Apartmenty	Apartmenty

Obrázek 3 – off-page analyzátor

Sledování pozic

Poslední sekci aplikace je sledování pozic. Je závislá na datech zadaných v sekci klíčová slova. Slouží jako prostředek pro sledování účinnosti SEO optimalizací. Získaná data o pozicích se pak vykreslují v této sekci do grafů. Pro omezení výpisů lze zvolit časový úsek, který se má na grafu zobrazit. Při velkém rozsahu dat jsou pozice před vykreslením grafu zhuštěny do vícedenních intervalů a zobrazená pozice je pak průměrem v tomto intervalu.



Obrázek 4 – sledování pozic

B Obsah DVD

Na přiloženém DVD se nachází následující adresářová struktura:

- */aplikace/* - zdrojové soubory aplikace
- */text-prace/* - zdrojový text této bakalářské práce
- */plakat/* - zdrojové soubory s plakátem projektu
- */web/* - zdrojové kódy webové prezentace projektu
- */pozice/* - tabulky s průběžným měřením pozic