

Pokročilé optimalizace pro webové vyhledávače

MARTIN TRNEČKA

Přírodovědecká fakulta UP, Olomouc

Tento článek navazuje na [3], kde jsme se věnovali problematice webového vyhledávání, základním principům a metodám SEO. Skončili jsme u optimalizací, které jsou úzce spojené se syntaxí jazyka HTML. V tomto článku na tyto optimalizace navážeme a budeme se věnovat pokročilým metodám SEO. Cílem obou dílů je podat srozumitelný výklad principů SEO. Učitelé informatiky a výpočetní techniky, kteří se ve své výuce zabývají problematikou webového vyhledávání, tak získají přehled o některých aktuálních tématech v této oblasti.

Pokročilé SEO

Pokročilé SEO zahrnuje především aspekty, které již nejsou zcela v režii tvůrce či majitele webu. Tvůrce webu ale může vhodnou úpravou webové stránky tyto aspekty alespoň částečně ovlivnit. Jako první se podíváme na optimalizaci ležící na pomezí mezi základní a pokročilou, která se týká URL adres.

URL adresy

URL (Uniform Resource Locator) adresa představuje jednoznačný identifikátor webových stránek a podstránek s touto stránkou spojených. Cílem optimalizace zahrnující URL jsou takzvaná „hezká URL“. Nejlepší jsou dobře zapamatovatelné a především logické URL adresy. Podíváme se nyní na několik hypotetických adres, které by mohly odkazovat na stránku zaměstnance, a zvažme jejich význam:

1. <http://www.example.com?location=staff&employee=berners-lee>
2. <http://www.example.com/staff/bernersleeinfo.html>
3. <http://www.example.com/staff/berners-lee-info.html>
4. <http://www.example.com/staff/tim-berners-lee>
5. <http://www.example.com/staff/berners-lee>
6. <http://www.example.com/berners-lee>

První URL adresa patří do kategorie označované jako *dynamická*. Obvykle je můžeme najít na automaticky generovaných stránkách jako jsou například diskuzní fóra. Tento typ URL adres je z hlediska SEO nejméně vhodný. V tomto konkrétním případě jsou parametry `location` a `employee` smysluplné. V praxi tomu ale často nebývá, nemluvě o tom, že pro uživatele je taková adresa těžko zapamatovatelná. Webové vyhledávače dokáží s dynamickými URL adresami pracovat, ale z celkového pohledu SEO jsou vhodné spíše na generované stránky, jako jsou výše zmíněná diskuzní fóra nebo různé výsledky vyhledávání. Při práci s dynamickými URL adresami je zapotřebí dát si pozor na prohazování pořadí parametrů, které může v některých případech působit zmatečně.

Druhá a třetí URL adresa se od sebe liší jen velmi nepatrně. Prostřední část URL adresy `/staff/` nese informaci o aktuální sekci stránek. Obě adresy jsou zakončeny příponou `.html`. Přípony určující typ souboru jsou považovány za přežitek a dnes se již moc nepoužívají. Z hlediska webového vyhledávání jsou názvy stránek `bernersleeinfo` a `berners-lee-info` v URL v podstatě srovnatelné. Z pohledu uživatele je preferovanější druhá varianta, která je lépe zapamatovatelná.

Čtvrtá adresa je tím nejlepším řešením. Neobsahuje žádné pro uživatele nepodstatné informace a je v ní jasně zanesena logická struktura webové stránky.

V páté adrese není na rozdíl od předchozí uvedeno křestní jméno. Z hlediska vyhledávání je lepší v URL uvést co nejvíce relevantních informací. Pátá varianta tedy nenese všechny potřebné informace.

Poslední adresa jako jediná neobsahuje informaci o sekci. Příslušný web není vůbec strukturován a všechny informace jsou prezentovány na jedné úrovni. To je v případě menšího webu s několika málo podstránkami snesitelné. V případě rozsáhlejšího webu je tato URL z hlediska SEO zcela nepřipustná.

Odkazy

Tím nejdůležitějším v kategorii pokročilé SEO jsou odkazy.¹ Pokud odkazujeme na důvěryhodné a kvalitní weby, zlepšuje se celkové hodnocení naší stránky. Je nutné si uvědomit, že takové odkazy nemohou být na stránku přidány uměle. Pokud není odkaz relevantní pro náš web, nemá žádný přínos. Odkazy mohou být přínosné, ale i naopak. Za nekvalitní odkazy, tedy například odkazy na spam, ilegální obsah či torrent sítě,² je celkové hodnocení stránky výrazně sníženo.

Důležitější než stránky, na které odkazuje náš web, jsou stránky, které naopak na náš web ukazují. Tyto odkazy se nazývají zpětné. Opět platí, že pokud na nás odkazují kvalitní stránky, je hodnocení našeho webu lepší. Stejně tak pokud na nás odkazují stránky nepřilíš valné reputace, může být hodnocení našeho webu sníženo. Málo kvalitních odkazů na webovou stránku je nejčastější důvod špatného umístění ve výsledcích vyhledávání.

Jak by tedy měl vypadat kvalitní odkaz? Především by měl být z důvěryhodného a tématicky příbuzného webu. Samotný odkaz by měl ležet uvnitř souvislého textu, měl by být dlouhodobě funkční a měl by vést na konkrétní podstránku našeho webu. Stránka, která odkazuje, by neměla mít mnoho dalších odkazů. Kvalita je v tomto případě důležitější než kvantita – jeden kvalitní odkaz je lepší než stovky nekvalitních.

Další SEO metody

Jako pokročilé SEO metody jsou dále označovány některé technické úpravy na webu. Příkladem je kvalitní navigační a organizační struktura, které není vždy jednoduché dosáhnout. Tím je myšleno, že náš web bude logicky strukturován do jednotlivých podstránek, které budou mezi sebou navzájem provázané pomocí snadno použitelné navigace. Dalším příkladem je sémantické pojmenovávání elementů. To znamená, že pokud chceme pracovat např. s elementem obsahujícím hlavní obsah, použijeme ve zdrojovém kódu pro jeho identifikaci³ místo pojmenování `class="hello1234"` pojmenování `class="main-content"`. Druhé pojmenování je logičtější a

¹Jazyk HTML je založen na jazyku SGML (Standard Generalized Mark-up Language). Jednou z mála odlišností jsou právě odkazy, které jazyk SGML vůbec neobsahoval. Bez odkazů by služba WWW měla zcela jinou podobu. Není tedy divu, že odkazy jsou tím nejdůležitějším elementem webové stránky, který určují její strukturu. Vyhledávací algoritmy tuto strukturu sledují a hodnotí.

²Torrent sítě slouží pro distribuci (zejména velkých) souborů. Obvykle jsou používány pro distribuci ilegálního obsahu jako například nelegální kopie filmů.

³Každá HTML značka může být pojmenována pomocí atributu `id` nebo `class`.

smysluplnější jak z pohledu návštěvníka, tak z pohledu tvůrce webové stránky. Do stejné kategorie patří i publikování obsahu pomocí standardů ATOM či RSS.

Další optimalizace se týkají duplicitního obsahu webových stránek. Pokud máme dvě stejné či jen velice podobné stránky, je novější stránka penalizována. Extrémním případem jsou obsahově chudé stránky, kde navigace, která je na všech podstránkách stejná, obsahuje více informací než jednotlivé podstránky. Stejně tak je dobré se vyvarovat přirozených duplicit v URL adrese. Tím je myšleno prohazování parametrů u dynamických adres o kterém jsme již mluvili a také uvádění URL s `www` a bez `www` prefixu. Prefix `www` v URL adrese má historické důvody a dnes již není nutný. Na stránku by mělo být odkazováno konzistentním způsobem a toho je možné docílit jediné tak, že stránka si sama vynutí správný typ adresy.

Novinky v SEO

Jak bylo napsáno v předešlém článku, algoritmy pro určování finální pozice webové stránky ve výsledcích vyhledávání se neustále mění. Z tohoto důvodu se některé techniky SEO mění s každým updatem vyhledávacího algoritmu. Někdy může být konkrétní metoda přínosná, jindy jsme za její uvedení na stránce naopak penalizováni. V následující části se podíváme na některé novinky v oblasti optimalizace stránek.

První novinkou v SEO je, že vyhledávání na mobilních zařízeních, jako je například chytrý telefon, není stejné jako vyhledávání na klasickém počítači. Vyhledávače preferují stránky, které jsou přizpůsobené pro zařízení, na kterém jsou vyhledávány. Je tedy vždy výhodné, aby byly stránky rozumně zobrazitelné a použitelné na co největším počtu různých zařízení.

Další záležitostí je celková velikost webu. Vyhledávače preferují stránky, které se načítají co nejrychleji. Především nezkušení programátoři se v posledních letech naučili používat celou řadu knihoven (například různé rozšíření vycházející z velice populární knihovny jQuery) usnadňující vývoj webu. Tyto knihovny jsou obvykle velké a pokud jich je na stránce mnoho, může se stát, že jednoduchá webová stránka má i několik megabajtů. Používání knihoven je jistě správná věc,⁴ ale je vždy dobré zvážit, zda je nutné vkládat velkou knihovnu kvůli jedinému grafickému efektu použitému na

⁴Knihovny dramatickým způsobem usnadňují vývoj webových stránek. Díky nim vytvoří působivé stránky i amatérský zájemce o tvorbu webu. Na druhou stranu bezmyšlenkovitě používání knihoven staví tvůrce webu do role uživatele. Tvůrce webu pak jen používá technologii, aniž by porozuměl hlubším principům, na kterých je založena. Tento fenomén je obecným problémem dnešní přetechnizované doby.

jediné stránce. S tím jde ruku v ruce i design webu, který je mnohdy přemrštěný a zahrnuje obrovské obrázky, které jen přidávají na velikosti stránky. Typickým příkladem jsou grafické slideshow na úvodní stránce, kvůli kterým se na mobilním zařízení stránka vůbec nezobrazí. Obecně platí, že čím menší je zařízení, na kterém chceme stránku zobrazit, tím menší by měla být i velikost⁵ této stránky.

Stáří webové domény je rovněž důležité. Dnešní vyhledávače zvýhodňují starší a zaběhnuté webové adresy. Zdali se nový webový projekt prosadí nebo zmizí v hlubinách internetu, je obvykle otázka času. Se stářím domény úzce souvisí pojem autoritativní doména. Za autoritativní domény jsou považovány ty domény, které jsou často odkazovány či citovány. K tomu se v dnešní době hojně používají různé sociální sítě. Aby se webová stránka stala autoritativní, musí si vybudovat dobrou pověst a to především kvalitním obsahem.

Dalším faktorem je, zda je daná stránka šifrována. Přesněji řečeno, jestli nabízí variantu přes `https` protokol. Bohužel i v dnešní době nejsou stále všechny webové stránky vyžadující přihlášení šifrovány. V roce 2014 společnost Google oznámila, že bude klást stále větší důraz na stránky poskytující šifrování. V současnosti není zcela jasné, jak velký tento důraz je, ale je zřejmé, že poskytnutí šifrování by mělo být naprostou samozřejmostí v případě, že vyžadujeme jakékoli informace od uživatele webové stránky.

SEO mýty

Jak již víme, algoritmy pro hodnocení webových stránek nejsou zcela známé a především se často mění. Jedním z důsledků je, že SEO je často zahaleno různými mýty, kterých se nejeden tvůrce webu zatvrzele drží. Navíc literatura věnující se problematice SEO je mnohdy nepřesná a zavádějící. Získat kvalitní znalosti z oblasti SEO lze jedinec praxí.⁶ Vyjmenovat všechny mýty, které se týkají SEO, by vydalo na celou knihu. Ve stručnosti se podívejme jen na některé z nich.

Jedním z nejstarších mýtů je, že tabulkový layout⁷ je pro SEO stejný jako layout vytvořený za použití CSS. Tabulkový layout je minulostí, ale zejména s příchodem moderního Material designu⁸ se spousta mladých

⁵Do velikosti stránky se započítává velikost všech souborů, které jsou nutné pro zobrazení stránky v prohlížeči.

⁶Studovat SEO znamená studovat ho neustále. Nelze se spolehnout na již jednou získané znalosti. To platí pro všechny webové technologie, které se neustále vyvíjejí.

⁷Tabulkovým layoutem je myšleno rozmístění hlavních prvků stránky (hlavičky, navigace, hlavního obsahu, patičky) do buněk tabulky.

⁸Grafický styl představený na konci roku 2014 společností Google, který se v sou-

programátorů nevědomky vrací právě k tomuto způsobu tvorby webu. Tabulkový layout byl zavržen zejména proto, že se samotná tabulka zobrazovala až poté, co se celá načetla do paměti. Dnešní moderní prohlížeče sice tabulku vykreslují postupně, tabulky by však měly být používány výhradně pro tabulková data. Jejich použití pro layout je sémanticky nesprávné a hodnocení stránky může uškodit.

Dnes je již historickým přežitkem, že `<description>` a `<keywords>` jsou nejdůležitější tagy na celé stránce. Tyto značky začaly být hojně zneužívány především za pomoci spamových klíčových slov (často hledaných slov, které s webem nesouvisí), takže k nim dnes přistupují webové vyhledávače obezřetně. Například Google vůbec nevyužívá obsah značky `<description>` pro vytvoření popisku stránky. Stejně tak obsah tagu `<keywords>` již neslouží pro určení klíčových slov. Dodejme, že i přesto mají oba tagy význam, a to především z pohledu sémantického webu.⁹

Dalším mýtem je, že podíl počtu klíčových slov k počtu všech slov na stránce musí být přesně x procent. Ještě před pár lety bylo ono číslo x magickým zaklínadlem, které mohlo webovou stránku vystřelit na první příčky ve výsledcích vyhledávání. Dnes je již situace jiná. Vyhledávače sice hustotu klíčových slov sledují, ale správná hodnota neexistuje. Pokud je obsah kvalitní, klíčová slova se v něm vyskytují přirozeně a to je více než dostatečné. Nad klíčovými slovy je prováděna spam analýza, která zajistí, že již není možné využívat často hledaných nerelevantních slov pro zlepšení pozice webové stránky. Navíc je počítána sémantická vzdálenost klíčových slov. Například klíčová slova „rohlík“ a „chleba“ mají menší sémantickou vzdálenost než slova „rohlík“ a „auto“. Slova s velkou vzdáleností nemohou být klíčová. S klíčovými slovy je také spojený mýtus, že se nesmí skloňovat ani časovat. Vyhledávače si s tím velice dobře poradí a to i v případě češtiny. Používání velice populárních tag listů (seznam klíčových slov obvykle v podobě, kdy jsou nejdůležitější slova psána většími písmeny) je v dnešní době penalizováno. Bez přínosu také zůstává použití klíčových slov v odkazech.

časnosti těší velké oblibě a to nejen na webových stránkách, ale i například v mobilních aplikacích pro platformu Android.

⁹Sémantický web je považován za budoucnost WWW. Základní myšlenka, poprvé přestavena v článku [1], spočívá ve strukturovaném uložení informací a jejich významu na webu. Takovéto informace jsou snadněji vyhledatelné a strojově zpracovatelné. Pokud například uložíme na web obrázek, nepřičítáme mu jen strohý popis, ale i informaci o tom, co se na obrázku nachází, kde a kdy byl pořízen a jak souvisí s obsahem webu. K tomuto účelu existuje celá řada technologií.

V souvislosti se SEO mýty je dále potřeba upozornit na nabídky mnohých SEO expertů, kteří garantují pozici ve vyhledávacích či dokonce registraci do webových vyhledávačů. Pozici ve vyhledávači totiž není možné nijak garantovat, stejně tak není potřeba registrovat stránku do webového vyhledávače. Fenomémem jsou analyzátoři SEO, které často nabízí naprosto zkreslené informace. Výsledky z těchto aplikací je vždy potřeba brát s nadhledem. Absence stoprocentního hodnocení nás vůbec nemusí znepokojovat – pro hodnocení jsou mnohdy použity zcela nesmyslné metriky.

Jak bylo napsáno v předchozím článku [3], SEO je především o obsahu. V tomto kontextu je v poslední době hojně diskutován content (obsahový) marketing [2]. Vymežit přesně tento pojem není jednoduché. Ve stručnosti je možné si ho představit jako nenucenou reklamu bez nabídky prodeje. Uživatelům poskytujeme zajímavý, zábavný a poučný obsah, který je upozorní na nás nebo náš produkt. Někteří jej považují za spásu, jiní za nesmysl.

Content marketing není ve skutečnosti nic nového. Poprvé byl představen již v roce 1895 a za více než sto let jeho existence bylo mnohokrát ověřeno, že funguje. Mezi velké propagátory content marketingu patří například Coca-Cola, Microsoft nebo Micheline. Nesmíme ale zapomínat, že se jedná o marketing. Jeho konkrétní podoba nemusí být vždy stejně úspěšná. Například reklama, která je úspěšná v zahraničí, nemusí oslovit české publikum. Content marketing je se SEO propojen tím nejdůležitějším, a sice obsahem. Mnozí SEO experti jej ale chápou nesprávně. Mnohdy je pak návštěvníkům webu vnucován obsah v podobě PR článků, videí a záplav příspěvků na sociálních sítích bez toho nejdůležitějšího — bez kvality. I zde platí, že přínos je přímo úměrný kvalitě.

Temná strana SEO

V předchozích částech jsme používali slovo penalizace. Penalizace znamená, že stránka získá záporné hodnocení nebo dokonce nebude vůbec zaindexována. Webové vyhledávače obvykle poskytují pouze informaci o tom, že použití zakázaných metod může být penalizováno. Situace je stejná jako u hodnotících algoritmů. Všechny postupy, které mohou být penalizovány nejsou zcela známé. Možnosti automatické penalizace jsou stále omezené. Velice populární se staly takzvané „bonzovací stránky“, na kterých je možné nahlásit nevhodné chování. Dodejme, že tyto stránky jsou hojně používány.

Hromadně se penalizované postupy označují jako neetické SEO metody. Mezi ty patří například umělé zvyšování četnosti klíčových slov za pomoci skrytého textu. Zvyšování kvantity odkazů na úkor kvality (např. spam v diskuzích nebo komentářích). Podvedení vyhledávacích robotů, kdy je vyhledávacímu programu podstrčen jiný obsah. V neposlední řadě se jedná o takzvaný „Ester Ládová fenomén“,¹⁰ kdy je na stránkách umístěna aktuálně hojně vyhledávaná informace za účelem zvýšení návštěvnosti.

Další neetickou techniku představují Google bomby, které ovlivňují vyhledávač Google. Tato technika je založena na následujícím principu. Pokud se odkazujeme na webovou stránku odkazem, který má určitý popis, je tento popis zohledňován při analýze klíčových slov. Pokud je takovýchto odkazů dostatečné množství, může se stát, že se tento popis stane hlavním klíčovým slovem či frází pro odkazovanou stránku. První známý případ použití Google bomby pochází z konce minulého tisíciletí, kdy po vyhledání fráze „more evil than Satan“ (horší než Satan) se ve výsledcích vyhledávání objevila na prvním místě domovská stránka společnosti Microsoft.

Ačkoliv jsou současné vyhledávače částečně imunní vůči výše popsaným neetickým metodám, neustále se objevují nové a nové podvodné přístupy, znesnadňující vyhledávačům jejich práci. Příkladem mohou být odkazující webové farmy, tedy obrovská armáda webů odkazujících se na konkrétní stránky, či metoda zvaná SEO poisoning, která je založena na infikování webového prohlížeče malwarem, který upravuje výsledky vyhledávání.

Dostáváme se tak k temné stránce webových optimalizací. SEO je často kritizováno, ale příliš se o tom nemluví. Hlavním důvodem kritiky je, že vyhledávací roboti nejsou dostatečně inteligentní. Stále nedokáží zcela automaticky rozpoznat podvodnou stránku. Další důvod pro kritiku představuje fakt, že veřejnost chápe SEO čistě jako pozici ve vyhledávacích, aniž by porozuměla hlubším mechanismům webového vyhledávání. Před rokem 2013, kdy Google poprvé představil updaty vyhledávacího algoritmu Panda a Penguin, byla mezi temnou a světlou stranou SEO vymezena jednoznačná mez. Příchodem těchto updatů se revolučně změnilo chápání SEO (viz [4]). Celková tolerance se posunula k temné straně a tento trend přetrvává až do současnosti. Některé na první pohled neetické metody jsou dnes webovými vyhledávači tolerovány. V případě dlouhodobě prověřených stránek je míra tolerance dokonce větší. Google se tak rozhodl dát zaběhnutým webovým stránkám prostředek pro boj s neetickou konkurencí.

¹⁰V anglicky mluvících zemích známí jako „Pamela Anderson fenomén“.

Závěr

Pokročilé SEO je spíše o technických maličkostech a především o chytrosti. Pokud bude majitel stránky přemýšlet o tom, jak jeho web vnímají uživatelé, a bude se jim snažit vyjít vstříc, výsledky se vždy dostaví a web bude prosperovat. Vždy je potřeba přemýšlet, zda má daná optimalizace přínos pro uživatele. Pokud tomu tak není, je zbytečná. Zatímco základní SEO by mělo být naprostou samozřejmostí, pokročilé SEO je mnohdy považováno za nadstandard. Správné principy SEO jsou součástí kvalitního webdesignu, které by měl dodržovat každý tvůrce webových stránek.

Poděkování. Tento článek byl vytvořen za podpory projektu IGA UP 2016, reg. č. IGA_PrF_2016_027. Děkuji Eduardu Bartlovi za jeho cenné rady a připomínky k textu.

Literatura

- [1] Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American, 2001, s. 29–37.
- [2] Clark, A.: Search engine optimization 2016: Learn SEO with smart internet marketing strategies. 2015.
- [3] Trnecka, M.: Základní optimalizace pro webové vyhledávače. MFI, roč. 25 (2016), č. 3, s. 227–234.
- [4] Williams, A.: SEO 2014 & Beyond: Search engine optimization will never be the same again! 2014.

Bobřík učí informatiku

6. díl – Kódování

DANIEL LESSNER – JIŘÍ VANÍČEK

Matematicko-fyzikální fakulta, UK Praha

Pedagogická fakulta, Jihočeská univerzita v Českých Budějovicích

V minulém dílu seriálu o situačních infromatických úlohách jsme se seznámili s pojmem informace a s tím, jak informace ubírá možnosti. Příliš jsme ale nezdůrazňovali, že každá informace, pokud ji chceme zpracovat, musí být nějak fyzicky zaznamenána.