

## Základní optimalizace pro webové vyhledávače

MARTIN TRNEČKA

Přírodovědecká fakulta UP, Olomouc

Od dob, kdy světlo světa poprvé spatřil internet, tehdy ještě v podobě experimentální vojenské sítě ARPANET, urazil technologický vývoj pořádný kus cesty. Mnohé technologie, spadající ještě před nedávnem pouze do žánru science fiction, se staly součástí běžného života. Dnešní podoba internetu nám dává široké možnosti, k dispozici máme všudypřítomný GSM signál a veřejné wi-fi sítě. Díky tomu se na internet dostaneme ve vlaku, letadle a dokonce i ve vesmíru. Pro přístup k internetu již není zapotřebí počítačů, nyní můžeme využít chytré telefony, televize, hodinky či brýle. Nacházíme se nyní na počátku epochy internetu věcí, kdy do internetové nebo na podobném principu fungující sítě jsou připojeny námi běžně používaná zařízení jako je například lednička či stolní lampa. Internet je zkrátka všude a množství informací v něm uložených neustále narůstá.

Naopak znalostí potřebných pro vytváření jednoduché webové stránky, nedílné součásti<sup>1</sup> internetu, je zapotřebí čím dál méně.

Zatímco v raných dobách internetu dokázal vytvořit webovou stránku jen velmi schopný technický nadšenec, dnes se učí tvorbě webových stránek děti na základní škole. I při současných možnostech však nelze přesně určit, kolik webových stránek internet obsahuje; jejich počet i různorodost neustále narůstá. Samostatnou kapitolu tvoří internetové obchody, které se

---

<sup>1</sup>Webové stránky jsou součástí World Wide Web (WWW), jedné z mnoha služeb poskytovaných na internetové síti.

v posledních letech těší obrovské popularitě. Důvody jsou zřejmé – provoz internetového obchodu je výrazně levnější než provoz kamenné prodejny, navíc je možné je navštívit kdykoli a z pohodlí svého domova. Pokud byste se před pár lety rozhodli na internetu prodávat pečivo, lidé by vás považovali za blázný. Pokud se pro stejnou věc rozhodnete dnes, neprorazíte kvůli konkurenci.

Na internetu je toho tolik, že je mnohdy obtížné nalézt přesně to, co potřebujeme. Uživatelé mají k dispozici za účelem vyhledávání webové (internetové) vyhledávače, jako jsou například portál `google.cz` společnosti Google, `bing.cz` společnosti Microsoft nebo portál `seznam.cz` provozovaný stejnojmennou společností.

Z celé řady studií [1] analyzujících chování uživatelů vyplývá, že se běžní uživatelé potýkají při vyhledávání se třemi zásadními problémy. První dva spolu úzce souvisí – uživatel mnohdy sám nemá přesnou představu o tom, co hledá a následně neumí zformulovat požadavek do vyhledávacího pole tak, aby mu vyhledávač porozuměl a vrátil relevantní výsledky. Třetím problémem je neochota uživatelů procházet více než prvních pár vrácených výsledků vyhledávání.

Jak ale zajistit, aby v případě, že již vlastníme nebo vytváříme webovou stránku, byla tato stránka na internetu vidět? Jednou z možností jsou právě optimalizace pro webové vyhledávače, kterými se budeme v tomto článku zabývat. Nejprve se ale podívejme, jak vlastně funguje webové vyhledávání.

## Webové vyhledávání a webové stránky

Předpokládejme, že máme již hotovou webovou stránku, která je umístěna na internetu. Naši stránku navštíví robot, přesněji řečeno automatický program, který stránku analyzuje a přiřadí jí číselné ohodnocení. Návštěva robota se nazývá indexace a obvykle vyústí v zařazení webu do webového vyhledávače. Přiřazené hodnocení se nazývá rank webové stránky a je ukazatelem důležitosti webu. Vyhledávač společnosti Google používá PageRank [2], který nabývá hodnot  $0, 1, 2, \dots, 10$ , vyhledávač `seznam.cz` pak S-rank, který nabývá hodnot  $0, 1, 2, \dots, 100$  a je založen na podobném principu jako PageRank.

Je důležité si uvědomit, že samotný rank neurčuje výslednou pozici ve webovém vyhledávači. Konečné umístění webové stránky je výsledkem algoritmu, který zvažuje celou řadu kritérií a rank stránky je pouze jedním z nich. Většina kritérií pro hodnocení se však nezveřejňuje. Na druhou

stranu tvůrci algoritmů sloužících k určení konečné pozice webové stránky ve vyhledávání své výtvořky často komentují a popisují jejich základní chování. Z těchto informací je možné odvodit, co je zapotřebí udělat, aby byla naše webová stránka snadno k nalezení. Tento postup se nazývá optimalizace pro webové vyhledávače, často označovaný jako SEO z anglického Search Engine Optimization.

Pro úplnost ještě dodejme, že algoritmy určující pozici ve vyhledávači se neustále vyvíjí. Menší úpravy jsou prováděny v několikadenních intervalech, větší změny jsou prováděny s odstupem několika málo měsíců. Modifikace algoritmů jsou řešeny pomocí updatů (aktualizací), které jsou někdy mylně zaměňovány se samotnými algoritmy. Tyto updaty jsou dále rozšiřovány a upravovány. V zásadě se jedná o filtry, které se aplikují na výsledky vyhledávacích algoritmů. Mezi nejaktuálnější updaty vyhledávače od Googlu [3] patří Panda, Penguin (Tučňák) a Pigeon (Holub). Poslední update vyhledávače seznam.cz se nazývá Jalapeño, který opět upravuje výsledky vrácené vyhledávacím algoritmem.

Webová stránka je ve skutečnosti produktem celé řady technologií. Dnes je každá webová stránka vytvořena pomocí jazyka HTML (HyperText Markup Language), který slouží pro popsání základní struktury a sémantiky webové stránky. Dále je použita technologie CSS (Cascading Style Sheets), která dává webu grafickou podobu a skriptovací jazyk JavaScript, který přináší na web dynamičnost. Tyto tři technologie se souhrnně označují jako client-side<sup>2</sup> technologie. Pro účely tohoto článku si vystačíme s obyčejným HTML, ostatní technologie zasahují do SEO jen velmi nepatrně. Pojdme si tedy jazyk HTML ve stručnosti představit.

Počátky jazyka sahají do roku 1990. V této době Tim Berners-Lee<sup>3</sup> vytvořil jazyk HTML pro účely výzkumného centra CERN ve Švýcarsku. Dnes je jazyk HTML základním stavebním kamenem každé webové stránky. Za aktuální verzi je považován jazyk HTML 5, který byl po zdoluhavém schvalovacím procesu 28. října 2014 přijat za oficiální standard. O jeho údržbu a vývoj se stará World Wide Web Consortium (W3C).

---

<sup>2</sup>Označení client-side je převzato z klient-server architektury v počítačových sítích. Webové stránky zahrnují i server-side technologie. Mezi ně patří například programovací jazyky PHP, ASP.NET, Java nebo Ruby on Rails a databázové technologie jako například MySQL, MariaDB či PostgreSQL, které z webů dělají plnohodnotné internetové aplikace.

<sup>3</sup>Sir Timothy „Tim“ John Berners-Lee je považován za tvůrce World Wide Web. Mimo jiné vytvořil protokol http, na kterém je založena služba WWW a úplně první webový prohlížeč.

HTML je značkovací jazyk, neobsahuje tedy žádné (až na několik okrajových výjimek) konstrukty pro řízení výpočtu, funkce ani cykly. Obsahuje značky, také nazývané tagy nebo elementy, které přiřazují obsahu webové stránky její sémantiku. Příkladem je značka `<nav>`, která slouží pro označení navigace. Většina značek má otevírací a uzavírací část (párové značky). Uzavírací část se zapisuje stejně jako otevírací část, jen je v ní navíc uveden symbol lomítka. Úplná syntaxe párové značky `<nav>`, je tedy `<nav></nav>`. To co je uvedeno mezi otevírací a uzavírací částí se nazývá obsah značky. Ten může být tvořen libovolným textem či dalšími HTML značkami. Nepárové značky nemají ukončovací část a tudíž nemají ani samotný obsah. Příkladem takové značky je značka `<img>`, která slouží pro vložení obrázku. Párové a nepárové značky mohou mít atributy, které se vždy zapisují do otevírací části. Příkladem je atribut `src`, který udává jaký obrázek se má zobrazit při použití značky `<img>`. Celý zápis i s atributem `src` je ``. Jaké atributy lze u konkrétní značky použít, je dáno specifikací jazyka [4]. Celá webová stránka je tvořena hierarchií HTML značek.

Co znamená přiřazení sémantiky? Představte si obyčejné noviny. Než je uvidíme v podobě, v jaké si je přečteme u ranní kávy, jsou noviny vyhotoveny ve formě čistého, neformátovaného textu (pomineme-li obrázky). Přiřazením sémantiky například určíme, že nějaká část textu tvoří nadpis, jiná část odstavec daného článku, seznam nebo tabulku. Bez sémantiky by se jednalo o čitelná, ale velmi nepřehledná data. Jazyk HTML dělá to samé, jen v kontextu webové stránky. Umožňuje určit, co bude nadpis stránky, navigace, tabulka a tak dále. Přiřazení sémantiky je prováděno pomocí značek jazyka tak, že se obsah jednoduše umístí do konkrétní značky. Například obsah obalený značkou `<p>` bude považován za odstavec, obsah značky `<a>` za odkaz. Webové prohlížeče obsahují základní grafické prostředky, které dokáží obyčejné HTML v určité podobě zobrazit. Tato podoba však není jazykem HTML specifikována.

## SEO

Podíváme-li se na nabídku a poptávku po SEO, zjistíme, že majitelé webů jsou mnohdy ochotni za SEO zaplatit více než za vytvoření celého webu. SEO je trend moderního webdesignu, se kterým se dříve či později setká většina těch, kteří nějaký web vlastní nebo webové stránky vytváří. Slovo trend by mohlo naznačovat, že optimalizace pro webové vyhledávače jsou jen přechodně populární záležitosti. Skutečnost je ale jiná. První op-

timalizace přišly se zrodem webových vyhledávačů – lidé chtěli, aby jejich stránky byly za každou cenu první. Například vyhledávač společnosti Google byl založen již v roce 1998; tuto dobu lze považovat za webový pravěk. Dnešní podoba internetu se bez vyhledávačů neobejde a optimalizace jsou s vyhledávači úzce provázány. Nejedná se tedy o technologii, která rychle zastarává či pomine, jak je v oblasti WWW běžné, ale spíše o nadčasový princip.

Nyní se podrobněji podíváme, čím SEO ve skutečnosti je a jak se dělá. SEO je přizpůsobení nebo také optimalizace webové stránky pro internetové vyhledávače. Jeden z nejrozšířenějších omylů je, že SEO je o prvenství ve vyhledávačích. Tak tomu ale vůbec není.

Hlavním cílem SEO je získat kvalitní návštěvníky a k tomu je nutné si uvědomit dvě základní fakta. Zaprvé, webových vyhledávačů jsou tisíce a nemá smysl být první ve vyhledávači, který nikdo nepoužívá. Zadruhé, je nežádoucí být první ve vyhledávači na základě klíčových slov, která se daného webu vůbec netýkají. Zadá-li například uživatel do vyhledávače klíčové slovo „informatika“ a na prvním místě se objeví stránka nabízející prodej pečiva, je téměř jisté, že takovou stránku uživatel nenavštíví. Zde vidíme, že pozice ve vyhledávači skutečně není tím hlavním ukazatelem. V duchu výše popsaného můžeme říct, že SEO je o tom, aby byl web viděn tím, kdo ho vidět chce.

## První krok

Nejprve je třeba se rozhodnout, zda a do jaké míry budeme webové stránky optimalizovat. SEO může být a mnohdy také je velice nákladnou záležitostí po časové i finanční stránce. Dá se tedy říct, že v určitých situacích se pokročilá optimalizace stránek nemusí vyplatit. Na druhou stranu alespoň základní optimalizace by měla být naprostou samozřejmostí. Mnoho lidí argumentuje, že nemá smysl optimalizovat stránky, které nemusí být první ve vyhledávačích. Jak bylo řečeno výše, optimalizace stránek není jen o pozici ve vyhledávači. Rozhodneme-li se, že nechceme, aby byly naše stránky indexovány, můžeme použít celou řadu prostředků k tomu určených. Bránit indexaci nekvalitní či žádnou optimalizací je nejhorší možný způsob, který navíc nebude fungovat.

Dalším krokem je výběr vyhledávače, pro který budeme stránky optimalizovat. Ve skutečnosti je tato otázka poměrně nezajímavá, neboť základní optimalizace je společná téměř všem vyhledávačům. Pokud bychom přesto chtěli určit relevantní vyhledávače, v českém prostředí se jedná o již

zmíněné, nejčastěji používané vyhledávače od firem Google a Seznam.cz. Zásadní rozdíl mezi nimi spočívá v tom, že zatímco **seznam.cz** indexuje jen české webové stránky, vyhledávač od Googlu je globální. Za zmínku stojí také webové katalogy, které lze chápat jako specifickou variantu webových vyhledávačů. Příkladem jsou například portály **heureka.cz** nebo **zbozi.cz**. V neposlední řadě zde patří dnes velmi aktuální sociální sítě.

## Principy SEO

Základní a nejdůležitější pravidlo pro SEO spočívá v tom, že se nesnažíme optimalizovat stránky pro webové vyhledávače, ale pro lidi. Indexace webových stránek sice probíhá strojově, jejich návštěvníci jsou však lidé. Skutečnost, že uživatel ze stránky uteče během pár vteřin nevykompenzuje ani první pozice stránky ve výsledcích vyhledávání. Přirozeným vývojem jsme se dostali do situace, kdy nejlépe hodnocené stránky jsou ty nejvíce užitečné; tedy stránky, na kterých našel uživatel přesně to, co hledal. Kvalitní obsah stránek je prioritou, se kterou jde ruku v ruce i dobrý a důvěryhodný vzhled. Návštěvníkům se musí na stránkách líbit, musí se na nich snadno orientovat a pohodlně je ovládat, čehož lze dosáhnout kvalitní grafickou prezentací. Není třeba se však obávat, že pouze graficky dokonalé weby mají šanci. Uživatelé snadno odpustí grafické nedokonalosti, nabídneme-li jim kvalitní obsah.

Při vytváření obsahu je velice důležitý výběr tzv. klíčových slov, které náš web identifikují, neboť ta právě slouží pro nalezení našeho webu. Výběr klíčových slov není úplně jednoduchý a to především v přesycených oblastech. Klíčová slova nelze vytvářet uměle; musí být srozumitelná, věcná a především přirozená pro náš web. Technicky vzato je výběr klíčových slov součástí automatického procesu indexace. Tvůrce webu však může tento proces do značné míry ovlivnit.

## Základní SEO

Nyní se podíváme na základní SEO a zaměříme se na různé aspekty, které můžeme jako tvůrci webových stránek ovlivnit. Tím prvním je titulek stránky uvozený HTML značkou `<title>`. Nevyplněný titulek je tím nejhorším možným prohřeškem – webové vyhledávače na něj kladou velký důraz. Obecně by měl být titulek krátký, srozumitelný, smysluplný a výstižný. Titulek může být společný pro celý web, ale i jednotlivé podstránky webu mohou mít vlastní titulky, což se běžně jeví jako lepší varianta. Čas-

tou otázkou bývá, jaké znaky lze použít v titulku stránky. V dřívějších dobách bylo lepší se diakritice vyhnout, ale dnes již nejsou s kódováním češtiny takové problémy a vyhledávače si s ním snadno poradí. Volba oddělovačů, jako jsou pomlčka nebo svislá čára, nemá na celkový výsledek vůbec žádný vliv.

Druhým klíčovým prvkem je značka `<h1>`, která označuje nadpis první úrovně. Pokud se vrátíme k analogii s novinami, tak se jedná o hlavní nadpis. Každá webová stránka by měla tuto značku obsahovat. Dříve platilo, že na každé stránce musí být právě jedna značka tohoto typu. Dle standardu HTML 5 je možné uvádět na stránce více značek `<h1>`. Existují názory, že z pohledu SEO je více hlavních nadpisů nesmysl. Dnešním vyhledávačům to až tak moc nevadí.

Na webové stránce je možné použít celkem šest úrovní nadpisů. Pokud je uveden nadpis určité úrovně, měly by mu předcházet nadpisy předešlých úrovní. Například nadpis `<h3>` by měl být použit až za nadpisem `<h2>`. Uvědomme si, že na slova uvedená v nadpisech je kladen velký důraz a je tedy výhodné, aby se v nich vyskytovala klíčová slova.

Dalšími důležitými značkami jsou `<description>` a `<keywords>`. Jak napovídají jejich názvy, tyto tagy reprezentují popis stránky a výčet klíčových slov. Popisek stránky je některými vyhledávači zobrazen na stránce s výsledky vyhledávání a měl by tedy být krátký a dostatečně informativní. Tag `<keywords>` by měl obsahovat nejdůležitější klíčová slova. Tak lze uvádět s oddělovačem (například čárka), ale i bez něj. Mnoho tvůrců webu se domnívá, že oba tagy musí být stejné na všech podstránkách. Tak tomu ovšem není. Je vždy lepší, když je popis a seznam klíčových slov šitý každé podstránce na míru. Přestože již tyto značky nemají tak zásadní význam jako dřív, jejich použití je stále výhodné. V dalším díle tohoto článku se k nim tedy ještě vrátíme.

Velice užitečnou značkou je rovněž tag `<strong>` uvozující obsah se zdůrazněným významem. V internetovém prohlížeči se obvykle zobrazí obsah tohoto elementu tučným písmem a právě na něj pak webové vyhledávače kladou větší důraz. Ideální situace nastává, když je tímto obsahem klíčové slovo. Mnoho programátorů této vlastnosti využívá, uměle obaluje všechna klíčová slova tímto elementem a za pomoci CSS je prezentuje jako obyčejný text. Tyto praktiky však leží na hranici programátorské etiky.

Dalším ovlivnitelným aspektem je syntaktická správnost (odborně označována jako validita) webové stránky. Jazyk HTML poskytuje programátorům určitou volnost a pokud je stránka v něm zapsaná syntakticky

nesprávná, stále se může v prohlížečích zobrazovat korektně. Z programátorského hlediska by měla být každá webová stránka validní. Validitu je možné snadno ověřit pomocí webové služby dostupné na adrese <https://validator.w3.org/>. S validitou souvisí i kvalita a množství kódu – obecně platí, čím méně zdrojového kódu, tím lépe.

Chceme-li například nastavit určitou část stránky jako navigaci, použijeme k tomu účelu určenou značku `<nav>`. Stejně tak platí, že pro multi-mediální obsah – obrázky, video či hudbu – bychom měli vždy vložit i textovou informaci vhodně popisující daný prvek. Tato pravidla jsou přínosná nejen z pohledu SEO, ale zpřístupníme díky nim náš web i návštěvníkům se zrakovým postižením, kteří používají speciální webové prohlížeče.

Na závěr jen dodejme, že všechny výše popsané optimalizace zahrnující syntaxi webové stránky, by měly být naprosto samozřejmé a tvůrce stránek by se jimi měl vždy zabývat. Dá se říci, že se jedná o jakýsi etický kodex, který je bohužel ve spoustě případů opomíjen a porušován.

Dostáváme ke konci první části. Optimalizace webových stránek pro internetové vyhledávače je jedním z pilířů moderního webdesignu. Bohužel je mnohdy SEO nesprávně vnímáno pouze z pohledu pozice stránek ve vyhledávačích. Základní SEO se především týká správného použití jazyka HTML a kvalitního obsahu. Pokročilé SEO, kterému se budeme věnovat v dalším díle tohoto článku, je o chytrosti a technických detailech. Pokročilé SEO je navíc zakryto celou řadou mýtů, kterým se rovněž budeme, alespoň okrajově, věnovat.

**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 ke kvalitě textu.

## Literatura

- [1] *Hölscher Ch. – Strube G.*: Web search behavior of Internet experts and newbies. *Computer Networks*, roč. 33 (2000), č. 1–6, s. 337–346.
- [2] *Page L., Sergey B., Motwani R., Winograd T.*: The PageRank Citation Ranking: Bringing Order to the Web. Technical Report, Stanford InfoLab, 1999.
- [3] <https://moz.com/google-algorithm-change>.
- [4] <https://www.w3.org/TR/html5/>.