

# Svět podle Googlu

Ondřej Pečený



Google založili dva stanfordští studenti Sergey Brin & Larry Page v roce 1998. Již v tom samém roce vykazoval 10 000 dotazů denně. Na konci roku 1999 to byly 3 000 000 dotazů, v květnu 2002 150 000 000 a dnes více než 200 000 000 dotazů za den. Databáze je umístěna na více než 10 000 linuxových serverech. V současnosti má firma Google Inc. více než 500 zaměstnanců a je jednou z mála čistě internetových firem, která je zisková.

---

Google se stal alespoň pro rok 2002 značkou číslo jedna. Vyplynulo to z průzkumu serveru Brandchannel.com ([http://www.brandchannel.com/features\\_effect.asp?id=143](http://www.brandchannel.com/features_effect.asp?id=143)) a za zajímavost považují to, že tak předstihl značky jako Apple, Coca-Cola či Ikea. Jak je však tedy možné, že se stále setkáváme s lidmi, kteří Google neznají? Jsou to pravděpodobně ti, kdo internet nepoužívají, pohybují se tam málo nebo si vystačí s několika oblíbenými servery a na internetu nevyhledávají pokročilejším způsobem (např. nepoužívají více vyhledávacích serverů). Mohou také požívat pouze regionální vyhledávací službu. Lze samozřejmě namítnout, že průzkum byl prováděn na internetu samotném, tudíž jsou jeho výsledky zkreslené. Informační pracovníci v komerčních i nekomerčních institucích však Google většinou znají a používají.

Povězme si něco o tom, co vede k obrovské popularitě Googlu mezi fulltextovými vyhledávacími službami. Bezesporu je to především obrovská velikost databáze indexovaných i neindexovaných dokumentů na Webu. Ke 14.4.2003 jich byly podle úvodní stránky 3 083 324 652. Uživatelé rovněž oceňují jednoduché uživatelsky přívětivé rozhraní, za kterým se skrývá technologicky sofistikovaný vyhledávací systém a absenci agresivních reklam, kterými jsou obšťastňováni jinde (srovnejte například s některými českými vyhledávači). Google ale není zcela bez reklamy – provozuje vlastní reklamní systém sponzorovaných odkazů Google AdWords, který však uživatelé snáze akceptují. Je možno zde vyhledávat soubory více typů, což neumožňuje každý server podobného zaměření a pozice jednotlivých odkazů je posuzována na základě systému jejich vzájemného provázání a popularity (Google PageRank). Jednoduše řečeno, odkazuje-li na nějakou stránku mnoho stránek jiných, bude v Googlu pravděpodobně zařazena výše. Ve skutečnosti je tento systém o něco složitější. Uživatelé rovněž oceňují možnost zobrazení indexované stránky tak, jak byla Googlem zaznamenána, i když se třeba od té doby změnila (odkaz Archiv za URL nalezené stránky). O popularitě Googlu svědčí například i to, že se v anglicky mluvícím světě začíná již běžně používat „google“ jako sloveso značící používání stejnojmenné služby pro vyhledání nějakých informací („to google something“). V češtině jsem se zatím s podobnými odvozeninami nesetkal. Jistě jde tak trochu i o módu. Dříve jsme se setkávali s otázkami typu: „Ty ještě nemáš e-mail?“ „Ty ještě nemáš stránky na internetu?“ Dnes můžeme slyšet udivené: „Ty ještě neznáš (nepoužíváš) Google?“

Ti, kdo se na internetu nějak prezentují, v naprosté většině také chtějí, aby jejich stránky navštívilo co nejvíce lidí. Někdy se může jednat jen o návštěvníky se specifickými zájmy. Ale v obou případech je snahou provozovatelů internetových prezentací nějak potenciálním návštěvníkům hledání usnadnit. Kromě volby vhodné internetové domény či domén (pokud taková možnost existuje) je pak pro provozovatele prezentací velkou starostí přední umístění stránek na vyhledávacích serverech. V situaci, kdyby většina populace online používala jen jeden z nich, by pak stačilo usilovat o dobrou pozici v seznamu vyhledaných stránek jen tam. Následuje logická otázka: Používá pro vyhledávání většina uživatelů internetu Google?

Odpověď nebude jednoznačná. Existují i jiné vyhledávací služby zaměřené buď na celosvětový internet nebo regionálně či omezené jazykově, které Googlu mohou velice konkurovat nebo se s ním mohou doplňovat. Navíc se v poslední době setkáváme s tím, že některé vyhledávače a portály využívají databáze a vyhledávací technologie Googlu (AOL, Netscape Search, Yahoo, z českých např. Seznam). Mým názorem je, že většina

těch, kteří internet používají již delší dobu a lze je považovat za jeho zkušenější uživatele, používá kromě ostatních vyhledávacích služeb stále častěji Google. To lze, myslím, prohlásit globálně. Přibližme se tedy zpět k odpovědi na položenou otázku. Do průzkumu preferencí na poli světových vyhledávacích služeb by se muselo zapojit velké množství lidí ze všech koutů světa, musely by zde odpovídat poměry pohlaví, věku, vzdělání a podobných parametrů celkovému rozložení v populaci online. To je úkol i pro sociology a pokud vím, nebylo něco takového globálně realizováno. Proto mohu pouze subjektivně na základě svých znalostí a zkušeností tvrdit, že Google považuji za nejdůležitější fulltextový vyhledávací server se všeobecným zaměřením. Co mě k tomuto názoru vede a zda tu existuje nějaký háček se dozvíte dále. Pokud očekáváte alespoň nějaká objektivnější data, tak odkazuji na výsledky měření firmy Nielsen/NetRatings zveřejněné na <http://searchenginewatch.com/reports/netratings.html>, kde se dozvíte o očividné převaze Googlu mezi vzorkem 60 000 Američanů v počtu hodin strávených vyhledáváním.

Vrátím-li se k tomu, co jsem zde již naznačil, většina provozovatelů internetových prezentací (bráno celosvětově) usiluje o dobrou pozici odkazu na jejich stránky v Googlu, kterému dává přednost před ostatními vyhledávači. Google má tedy mezi vyhledávači jakýsi monopol. Do určité míry lze webovou prezentaci pro Google optimalizovat a ve světě se o to pokouší velké množství lidí. Číhá zde však následující nebezpečí. Někaký čas může být odkaz při určité kombinaci vyhledávaných slov mezi prvními a najednou se může z různých důvodů (známých pouze Googlu) propadnout nebo být dokonce vyřazen. Známy je případ společnosti Data Recovery Group, která se loni v Googlu umísťovala okolo čtvrtého místa a když byla na začátku letošního roku z databáze vyřazena, postihl ji třicetiprocentní propad tržeb (viz <http://news.com.com/2102-1023-963618.html>). Jak je Google používán česky mluvícími uživateli internetu?

Velkou předností pro česky mluvící uživatele internetu je skutečnost, že Google používá v závislosti na jazykovém nastavení operačního systému české rozhraní pro vyhledávání. Komu by nevyhovovalo, může si zvolit z dalších 87 jazyků, mezi kterými nechybí např. esperanto, latina, nepálština či hackerština. Tlačítko „Zkusím štěstí“ (anglické „I am feeling lucky“ překládané dříve úsměvným „Cítím se šťastně“) v jednoduchém vyhledávání přenesou uživatele rovnou na první vyhledanou stránku. Pod ním je možnost zaškrtnout „Vyhledávání stránek česky“. Tato funkce však nefunguje na sto procent a osobně bych doporučoval ji pokud možno nepoužívat. Kromě češtiny můžete vyhledávat stránky ve 34 dalších jazycích.

Ve srovnání s českými vyhledávači Google plně obstojí, i když naprostá většina českých uživatelů internetu používá podle mých zkušeností Seznam. Seznam také obsahuje lepší hierarchicky tříděný katalog stránek pro českou oblast internetu. Používá však částečně technologii Googlu. Google využívá katalogu Open Directory Project ([www.dmoz.org](http://www.dmoz.org)), kterého využívají i další vyhledávače, ale ten je českými uživateli teprve objevován. Setkávám se se stále větším počtem lidí, kteří tvrdí, že pomocí Googlu naleznou více informací než v českých vyhledávacích jako je Seznam, Centrum a Atlas. Google rozlišuje mezi slovy zadávanými s diakritikou nebo bez ní a pro nalezení většího počtu stránek je někdy nutno zadávat slova skloňovaná nebo časovaná. I v angličtině je nalezen jiný počet stránek např. při zadání slova „airline“ a „airlines“.

Informační pracovníci v komerčních i nekomerčních institucích by rozhodně měli využívat všech možností, které jim nabízí pokročilé (rozšířené) vyhledávání na Googlu. Kromě běžných logických operací zde můžeme vyhledávat přesné fráze, volit si stránky napsané v konkrétním jazyce či omezovat vyhledávání pouze na určitý formát souboru. V současnosti je k dispozici vyhledávání souborů typu PDF, PS, DOC, XLS, PPT a RTF, ale pomocí příkazu „filetype“ je možno vyhledávat i soubory typu TXT, HTML, WPD, WML atd. Syntaxi příkazu naleznete níže. Vyhledané dokumenty nám Google pomocí odkazu „View HTML“ umožní alespoň v základní podobě zobrazit aniž bychom k tomu potřebovali příslušný program. Tento odkaz může sloužit i pro rychlý náhled na dokument bez nutnosti jej stahovat v originálním formátu. Vyhledávané termíny jsou v tomto případě v zobrazeném dokumentu pro lepší orientaci barevně odlišeny. Google v současnosti obsahuje odkazy na zhruba 35 000 000 dokumentů jiného formátu než HTML, což je

oblast nazývaná mnohými „neviditelný web“, v angličtině „invisible“, „deep“ či „dark web“. Více se tímto tématem ve svém příspěvku zabýval např. Martin Lhoták na Inforu 2002 (<http://www.inforum.cz/inforum2002/prednaska4.htm>). Sám považuji „neviditelný web“ za dosti subjektivní termín odvíjející se například právě od znalosti pokročilého vyhledávání. Mám tím na mysli to, že míra „neviditelnosti“ dokumentů jiných než HTML může být u každého dána schopností je nalézt (třeba pomocí Googlu). „Neviditelný web“ se však netýká jen těchto dokumentů.

Kromě toho je možno na stránce pokročilého vyhledávání zvolit datum poslední aktualizace dokumentu na webu (3, 6 nebo 12 měsíců) a zvolit místo výskytu vyhledávaných termínů. Je možno je vyhledávat kdekoli na stránce, v jejím názvu, těle, v adrese URL nebo v odkazech na stránce. Vyhledávání můžeme omezit jen na doménu druhé nebo nejvyšší úrovně (např. inforum.cz, cz, com nebo org). Můžeme se pokusit najít stránky podobné nějaké zadané stránce nebo ty, které na ni odkazují. Stránka pokročilého vyhledávání nabízí možnost využívání všech těchto funkcí, pokud si nechcete pamatovat konkrétní příkazy zadávané do políčka jednoduchého vyhledávání (ty mohou ale práci zrychlit). Následuje seznam použitelných příkazů s příklady:

Star Wars Episode +I	- „I“ musí být ve vyhledaných stránkách obsaženo (patří mezi takzvaná „stop words“)
„informační věda“	- najde přesnou frázi
„jak pracovat s * překladem“	- najde přesnou frázi s různými slovy místo hvězdičky
cembalo OR virginal	- najde stránky s minimálně jedním termínem
Klaus -Santa	- „Santa“ nesmí být ve vyhled. stránkách obsaženo
rozpočet filetype:xls	- omezení jen na určitý formát dokumentu (XLS)
intitle:CPU Athlon	- najde „CPU“ v názvu stránky a „Athlon“ kdekoli
allintitle:idnes galerie	- najde „idnes“ a „galerie“ v názvu stránky (současně)
inurl:chemie pokusy	- najde slovo „chemie“ v URL a „pokusy“ kdekoli
allinurl:adsl praha	- najde slovo „adsl“ a „praha“ v URL (současně)
link:www.inforum.cz	- najde stránky obsahující odkaz na stránky Infora
albertina site:cas.cz	- omezí na doménu „cas.cz“ a najde „albertina“
allinanchor:Sydney	- najde stránky obsahující „Sydney“ v odkazu
related:www.nkp.cz	- najde stránky podobné stránce Národní knihovny ČR
info:www.learning.cz	- zobrazí informace o dané stránce (cache, related, link, kde se vyskytuje „www.learning.cz“)
cache:www.prace.cz nabídky	- zobrazí danou stránku tak, jak byla indexována a zvýrazní slovo „nabídky“
stock:intc ihoo	- burzovní informace o Intelu a Yahoo

Pokud bychom se podívali na pokročilé (rozšířené) hledání obrázků, budeme již kromě tří kolonek na této stránce znát z pokročilého vyhledávání všechny ostatní. Je zde totiž možno zvolit velikost hledaného obrázku (jako ikona, malý, střední, velký, velmi velký, jako wallpaper), jeho typ (JPG, GIF, PNG) a zbarvení (libovolné, černobílé, šedé, barevné).

Google vyvíjí program pro návrh pravopisu jednotlivých vyhledávaných termínů, který by měl na základě kontextu zobrazovat alternativy pro slova napsaná s překlepy. Měl by být údajně efektivnější než ostatní komerčně dostupné programy pro kontrolu pravopisu. Již nyní se můžete při vyhledávání setkat s tím, že je vám nabídnut jiný obvyklejší termín. Kdybychom chtěli například najít stránky výrobce praček Whirlpool a napsali omylem „whirpool“, napíše Google: „Do you mean whirlpool?“

Zmiňoval jsem se zatím o pozitivních vlastnostech Googlu, ale je třeba zmínit i ty negativní. Má omezené možnosti vyhledávání co do pravostranného rozšiřování (případně nahrazování jednoho a více znaků), neumožňuje používat závorky – skládání termínů, neumožňuje používání proximitních operátorů, indexuje pouze prvních 101 kB webové stránky a 120 kB dokumentu typu PDF. Google nerozlišuje mezi malými a velkými písmeny. Nepodařilo se mi nalézt možnost vyhledávání souborů formátu video, audio nebo pomocí protokolu FTP, kterou např. AlltheWeb poskytuje. Oproti vyhledávačům AlltheWeb nebo Teoma poskytuje však Google možnost vyhledávání termínů v již vyhledaných stránkách (Search within results). Jako Alta Vista pak umožňuje překlad

některých stránek do jiných jazyků. Do budoucna by Google určitě mohl počítat s integrací databází jako je např. ProQuest do svého systému, aby se tak pro předplatitele této služby ještě více zjednodušilo vyhledávání a Googlu se dostalo větší popularity.

Google nabízí svým uživatelům službu SafeSearch, tj. podle svého tvrzení vyhledávání pouze mimo stránky „s pornografickým a explicitně sexuálním obsahem“, ale testování této služby Benjaminem Edelmanem z Berkman Center for Internet and Society na Harvard Law School ukázalo, že jsou blokovány desítky tisíc webových stránek bez zjevné příčiny. Byly to stránky vzdělávacích a nevýdělečných institucí, zpravodajských médií a národních či místních zastupitelstev. Uživatelé spoléhající se tak na SafeSearch mohou tak systematicky přicházet o informace díky chybám této služby (více na <http://cyber.law.harvard.edu/people/edelman/google-safesearch>).

Dostáváme se k otázce bezpečnosti. Některým lidem (zvláště webmasterům) se u Googlu nelíbí funkce zobrazování stránek v podobě, jakou měly při návštěvě robota Googlu neboli Google Cache. Takovou funkci v širším měřítku nalezneme ještě např. u známého serveru Internet Archive ([www.archive.org](http://www.archive.org)) a tvůrci stránek někdy namítají, že když stránky z nějakého důvodu z internetu stáhli, tak by neměly být někde stále dostupné. Sice existuje možnost, jak Googlu indexování stránek pro Google Cache zakázat (princip opt-out), ale je namítáno, že by spíše mělo být ručně povolováno (a implicitně zakázáno – princip opt-in).

Google by někteří rádi nominovali na titul Velkého Bratra roku a uvádějí hned několik důvodů (<http://www.google-watch.org/bigbro.html>): Jednak je to již zmiňovaná archivace stránek v systému Google Cache. Dalším důvodem je používání cookies, které mají platnost až do roku 2038. Každý, kdo navštíví Google, obdrží cookie (krátký textový soubor) do svého počítače a Google jej tak může při dalších návštěvách identifikovat. Pro Google je tedy pomocí cookie možné zjistit, co jste zde všechno hledali, kdy to přesně bylo, jaký jste použili prohlížeč nebo jakou jste měli IP adresu. Google k tomu dodává na stránce <http://www.google.com/privacy.html>, že zasílá cookies kvůli zlepšení kvality služby a lepšímu porozumění uživatelům. Neposkytne údajně získané informace třetí straně vyjma situace právního řízení jako je vydání zatykače, předvolání k soudu, zákonné opatření nebo soudní příkaz. Dalším problémem je podle odpůrců Googlu softwarový nástroj Google Toolbar, který si můžete stáhnout ze stránek Googlu a nainstalovat do Internet Exploreru. Měl by jako integrální součást prohlížeče urychlovat přístup k vyhledávání na Googlu, protože obsahuje vyhledávací políčko, které je tak při procházení různých stránek stále k dispozici, ale může také Googlu zasílat poslední vyhledávané termíny (opět za pomoci cookies). Navíc se aktualizuje automaticky bez upozornění.

Google již není jen vyhledávač. Na <http://www.google.com/intl/cs/options/index.html> je možno se seznámit s dalšími službami a nástroji:

- Froogle – vyhledávání cen různých produktů na webu
- Google Answers – fórum, kde na rozličné otázky odpovídají za poplatek odborníci
- Google Catalogs – možnost procházet tištěné reklamní katalogy
- Google Groups – možnost komunikovat v diskusních skupinách Usenet
- Google Labs – prototypy a projekty inženýrů Google ve vývoji
- Google News – možnost prohledávat více než 4500 zpravodajských zdrojů
- Google Special Searches – možnost vyhledávat na téma Apple, Microsoft atd.
- Google University Search – vyhledává na stránkách konkrétní university
- Google Web Directory – hierarchický katalog organizovaný do témat
- Google Wireless – možnost přístupu ke Googlu z PDA zařízení
- Google Browser Buttons – přidání tlačítek odkazujících na Google do prohlížeče
- Google in Your Language – dobrovolný překlad vyhledávacího rozhraní
- Google Toolbar – přidání vyhledávacího políčka do Internet Exploreru
- Google Translate Tool – překlad textu nebo celých stránek (v anglickém rozhraní)
- Google Web APIs – nástroj pro použití Googlu určený programátorům

