



# Jak vyhledávat na Internetu

## Základní seznámení s Internetem

**Internet**<sup>1</sup> je celosvětová síť počítačových sítí založených na skupině protokolů [TCP/IP](#) (Transmission Control Protocol over Internet Protocol). Pro Internet je charakteristické, že nemá vlastníka ani žádné řídicí centrum, které by rozhodovalo co, kde a kdy bude v síti zveřejněno. Existují pouze organizace zabývající se provozem a dalším vývojem Internetu (např. [Internet Society - ISOC](#), [W3C - World Wide Web Consortium](#)).

Internet nabízí přístup k informacím bez bariér a omezení, bez ohledu na vzdálenost nebo hranice států. Na Internetu dochází k neustálým změnám a přesunům. Neustále vzniká velké množství nových zdrojů informací, ale také mnoho zdrojů zastarává a zaniká. Nejsou určena žádná závazná pravidla pro výběr zveřejňovaných informací a chování uživatelů v Internetu. Existují pouze nepsaná avšak obecně uznávaná pravidla slušného chování uživatelů, tzv. [Netiquette](#).

## Úskalí vyhledávání na Internetu

Díky snadnosti zveřejňování dokumentů a informací ve www prostředí Internetu dochází k obrovskému nárůstu informací a dokumentů, které často nedodržují pravidla a standardy pro elektronické publikování ani standardy jazyka HTML. To znesnadňuje jejich zpracování, **uspořádání a následné vyhledávání**.

Při hodnocení kvality a použitelnosti informací zveřejněných na Internetu je důležité uvědomit si, že:

- výběr zveřejněných informací není nikým odborně editován a kontrolován (Internet je demokratický a svobodný komunikační a informační nástroj)
- často naleznete dokument bez uvedení autora, je-li uvedeno autorovo jméno, jen zřídka je připojena informace o jeho odborné způsobilosti a účel za jakým byla informace zveřejněna
- problémem je datovat dokument (datum chybí; pokud text časový údaj obsahuje, nemusí být jasné, zda je to datum zveřejnění dokumentu, poslední revize...)
- je těžké určit, jedná-li se o informaci původní (primární) nebo převzatou (sekundární)
- elektronické publikování ztěžuje rozlišování skutečných seriózních informací od reklamních textů

## NEZAPOMEŇTE:

- Internet je jen jedno z prostředí, kde lze nalézt informace
- kombinujte zdroje na Internetu s tradičními informačními zdroji (tj. knihy, periodika, katalogy knihoven, databáze)
- skutečně kvalitní informační zdroje jsou na Internetu zpravidla placené

Dokumenty dostupné ve www prostředí lze částečně rozdělit na dva okruhy. První tzv. **viditelný web (visible web, surface web)** obsahuje dokumenty, které jsou běžně volně přístupné a snadno vyhledatelné vyhledávacími stroji. Jeho velikost se odhaduje okolo 2-3 miliard indexovatelných dokumentů, tj. dokumentů, které mohou být podchyceny v databázích vyhledávacích služeb založených na automatizovaném sběru dat.

Druhým okruhem je tzv. **neviditelný web (invisible web)**, někdy také nazývaný **hluboký web (deep web)**. Jde o oblast dokumentů, které jsou běžnými vyhledávacími stroji obtížně vyhledatelné (tj. jsou pro ně neviditelné). Podle studie Michaela Bergmana pro společnost BrightPlanet se dá odhadnout velikost neviditelného webu přibližně na 400- 500 násobek webu viditelného. Naleznete zde především informace uložené v **databázích** (54 % deep webu), **adresáře, specializované vyhledávače, dokumenty v jiných formátech** než HTML (např. PDF, PS, Macromedia Flash - v těchto formátech se často publikují výzkumné zprávy a studie, příspěvky z konferencí, oficiální zprávy a další významné dokumenty), stránky tzv. **samotáři**, z kterých není odkaz na jiné a na které není odkaz, **dynamicky generované stránky** (po základní interakci s uživatelem) - např. katalogy knihoven, kalkulátory a také **stránky chráněné heslem**.

**Jak poznáme věrohodné informace?** Aneb **čeho** bychom si měli všimnout při zobrazování výsledků vyhledávání? Hodně nám napoví:

- komu je zdroj určen (laikům x odborně zaměřené skupině uživatelů)
- údaje o autorovi - ať už jde o jeho podpis, údaj o jeho odborné způsobilosti, případně o odkaz na jeho kontakt (přece jenom více věříme informačnímu zdroji, pod který se někdo podepíše anebo zveřejní spojení na sebe)
- především u odborných zdrojů je to adresa zdroje, doporučené (fungující) odkazy a citovaná literatura
- údaje o aktualizaci
- gramatika, stylistika
- grafická úprava
- uživatelská přívětivost
- možnost diskuze s jinými uživateli (diskusní fórum)

## Možnosti vyhledávání na Internetu

Na Internetu najdeme téměř vše, co potřebujeme. Proto je důležité znát všechny možnosti vyhledávání na Internetu.

Nejdříve bychom měli začít tím, že v prostředí Internetu najdeme 2 základní druhy zdrojů:

1. veřejné (dostupné přímo, mají veřejný charakter, přístup k nim je bezplatný)
2. neveřejné (přístupné zprostředkovaně, jde především o zdroje placené, které poskytují profesionální a komerční databázová centra)

K tomu, abychom získali přístup k veřejným či neveřejným informačním zdrojům, nám slouží nejrůznější **vyhledávací nástroje**. Podle nabízených služeb je můžeme dělit na několik druhů. Měli bychom myslet na to, že každý vyhledávací nástroj poskytuje jiné služby (které také mohou zanikat nebo naopak mohou vznikat nové), prostředky vyhledávání, rozsah a obsah prohledávané databáze. Měli bychom si uvědomit, že pokud se pohybujeme v databázi vyhledávacího nástroje, tak prohlédneme pouze zdroje, které vyhledávací nástroj indexuje, a ne celý Internet!

Z tohoto důvodu jsou výsledky u jednotlivých vyhledávacích nástrojů odlišné.

Mezi **vyhledávací nástroje** patří:

- předmětové katalogy
- vyhledávací stroje
- metavyhledávače
- virtuální knihovny

## Vyhledávací nástroje

Pro vyhledávání informací na Internetu lze použít čtyři základní vyhledávací nástroje, přičemž každý z nich má své výhody i nevýhody. Záleží na tom, jaké informace uživatel požaduje a co od vyhledávání očekává. Je ale dobré vědět o všech vyjmenovaných možnostech, které lze různě kombinovat pro dosažení nejvíce efektivních výsledků.

Především předmětové katalogy a vyhledávací stroje se v současné době prolínají a nelze je striktně dělit. Některé vyhledávací stroje nabízejí i předmětový přístup a naopak. Rozdíl spočívá v tom, co bylo primární funkcí vyhledávací služby - zda předmětový katalog nebo vyhledávací stroj.

Chcete vědět, jaké vyhledávací nástroje existují? Zkuste praktické rozcestníky jako je [Beaucoup](#) nebo [Search Engine Colossus](#).

**Předmětové katalogy (subject trees)** fungují na principu hierarchie a jsou budovány ručně. Pokrývají proto zákonitě menší část webového prostoru než vyhledávací stroje. Internetové odkazy jsou zprostředkovány prostřednictvím organizovaných skupin a podskupin **od nejvšeobecnějších až po konkrétní webové stránky**. Předmětové katalogy je vhodné použít v případě, kdy přesně **nevíme co hledáme** nebo pokud **je záběr vyhledávání příliš široký** a potřebujeme se dopracovat k užšímu tématu. Vhodné jsou rovněž při vyhledávání informací o **firmách, výrobcích**, a také když se chceme vyhnout dokumentům s nízkou kvalitou obsahu, které často vracejí vyhledávací stroje. Kladem i záporem předmětových katalogů je hlavně to, že adresáře jsou uspořádávány ručně, tzn. zajištění

určité záruky důvěryhodnosti, ale také to, že pokud je výběr závislý na člověku, můžeme se setkat se subjektivním přístupem při hodnocení a zařazování informačních zdrojů.

Mezi nejznámější předmětové katalogy patří např.:

- [Yahoo!](#)
- [Seznam](#)
- [Atlas](#)
- [Centrum](#)
- [LookSmart](#)
- [Open Directory Project](#)

**Vyhledávací stroje (search engines)** umožňují automatizovaný sběr dat za pomoci robotů, kteří prostřednictvím hypertextových odkazů prohlížejí jednotlivé www stránky. Nalezeným informacím jsou poté přiřazena slova, která společně s www stránkami tvoří databázi možných odpovědí na uživatelský dotaz. Vyhledávací stroje umožňují vyhledávání, díky kterému nám stačí zadat správně naformulované klíčové slovo, které můžeme definovat také pomocí operátorů. Vyhledávací stroje můžeme použít hlavně v případě, pokud **máme konkrétní představu o tom, co chceme najít** a pokud **chceme projít co největší množství stránek**. Základním nedostatkem je (oproti předmětovým katalogům), že ne vždy nám výsledky vyhledávání přinesou relevantní či kvalitní odkazy. Sběr stránek probíhá automaticky prostřednictvím robotů, a proto se v databázi vyhledávacích strojů nachází mnoho stránek pochybné kvality.

Mezi vyhledávací stroje patří např.:

- [Google](#)
- [Jyxo](#)
- [Alta Vista](#)
- [AlltheWeb](#)
- [Lycos](#)
- [Excite](#)
- [Morfeo](#)

**Metavyhledávače (meta search engines)** představují nový trend ve vyhledávání informací na Internetu. Tyto systémy rozešlou uživatelský dotaz do několika heterogenních zdrojů. Každý z těchto zdrojů má svůj vlastní vyhledávací systém. Vyhledávání přitom probíhá napříč nejrůznějšími vyhledávacími jazyky. Výhodou metavyhledávačů je, že se nemusíme učit případné rozdílné způsoby při formulování dotazů u jednotlivých vyhledávacích systémů. Informace jsou nám zpřístupněny v jednotném formátu, často bývají seřazeny podle relevance a bez duplicitních položek.

- [Dogpile](#)
- [Globalsearch](#)
- [Metacrawler](#)

- [MetaGopher](#)
- [Search](#)
- [Clusty](#)
- [Kartoo](#)

### **Virtuální knihovny**

Mají podobnou funkci a strukturu jako předmětové adresáře či vyhledávací stroje. Jde o databáze zaměřující se na odbornou literaturu, vědecké informace atd. Jsou tvořeny informačními odborníky, což zajišťuje kvalitu informací.

- [Internet Public Library](#)
- [Scirus](#)

### **Jak prohledat neviditelný web**

Běžně používané vyhledávací služby jako Google, Yahoo! nebo Alta Vista do neviditelného webu neproniknou. Existují však vyhledávací systémy, které dokáží s neviditelným webem pracovat - např. odkazují na databáze. Neviditelný web mapují např.:

- [Direct Search](#) - vyhledávací nástroj neviditelného webu
- [CompletePlanet](#) - registruje přes 70 tis. databází a vyhledávacích strojů
- [Turbo 10](#) - metavyhledávací nástroj neviditelného webu, třídění výsledků vyhledávání podle kategorií, obsahuje vládní, univerzitní a obchodní online zdroje
- [Scirus](#) - přes 450 mil. vědecky zaměřených www stránek, specializovaný vyhledávací nástroj pro vědecké informace (články, výzkumné zprávy, patenty...)
- [Find Articles](#) - přístup k volně dostupným článkům i placeným zdrojům
- [Infomine](#) - akademický portál, obsahuje primární dokumenty z různých oblastí vědy

## **Kritéria úspěšnosti vyhledávání**

Jakým způsobem používáme Internet a jaké možnosti využíváme při vyhledávání, to se odvíjí od našich znalostí (jednak o vyhledávání informací obecně, ale také o vyhledávacích nástrojích, o tom, co nám nabízejí a jakým způsobem se v nich můžeme orientovat) a zkušeností (praxe s vyhledáváním).

Mezi hlavní příčiny našeho neúspěchu při hledání patří:

- chybné místo a způsob hledání (pokud dlouho nic nenacházíme i ve více zdrojích)
- špatná formulace dotazu (když nalezené informace neodpovídají našemu požadavku)

Proto, abychom byli schopní nalézt to, co potřebujeme, musíme především:

- mít správný odhad v tom, jestli je vůbec možné požadovanou informaci na Internetu najít
- zvolit takový způsob vyhledávání, který nám zaručí výsledek
- neztrácet naději, pokud se nám ze začátku nedaří nic relevantního nalézt

## Základní postup při vyhledávání

- ze všeho nejdříve si musíme ujasnit, co chceme najít, tzn. musíme popsat téma větou nebo více větami
- když už máme konkrétní představu o tom, co hledáme, měli bychom téma zformulovat do klíčových slov a také jejich synonym (dávat si pozor na pravopis)
- dalším krokem je výběr nejvhodnějšího vyhledávacího nástroje a zjištění, jakým způsobem pracuje a jaké techniky pro vyhledávání používá, a přizpůsobení dotazu jeho možnostem
- po zobrazení výsledků musíme vyselektovat pro nás relevantní informace a pokud nejsme s výsledky spokojeni, musíme změnit buď náš dotaz nebo vyhledávací nástroj

### Formulace dotazu

Pro většinu vyhledávacích nástrojů platí stejné způsoby zadávání informačního dotazu.

**Booleovské operátory** – určují vztahy mezi jednotlivými hesly

**AND** vyhledá dokumenty, které obsahují oba výrazy

"Jan Saudek" AND "Sára Saudková"

**OR** vyhledá dokumenty alespoň s jedním výrazem anebo s oběma, zvyšuje se tak množství výsledků vyhledávání

"Jan Saudek" OR "Sára Saudková"

**NOT** vyhledá dokumenty, které obsahují pouze první termín, ale už ne druhý (specifikujeme informační dotaz)

"Jan Saudek" NOT "Sára Saudková"

Morava NOT řeka

**Znaménka + a –**

Jde o obdobu operátorů AND a NOT, kde + znamená, že dokument musí obsahovat požadovaný výraz a naopak – jej obsahovat nesmí. Mezi znaménkem a termínem nesmí být mezera.

Morava +řeka

Morava -řeka

**Fráze**

Pokud chceme najít dokumenty, které obsahují výrazy v přesně stanoveném tvaru a pořadí vedle sebe, můžeme použít uzavřených dvojitých uvozovek. Nejčastěji se tato možnost využívá při vyhledávání jmen a názvů.

„Karel Čapek“

„sedmikráska obecná“

## Zajímavé funkce Googlu

Snad každý z nás dnes při vyhledávání informací používá v první řadě vyhledávací stroj Google. Tento vyhledávač je tak notoricky známý, že ho jistě není třeba podrobně představovat. Ale znáte opravdu všechny jeho funkce a služby? Představíme si pár triků, které vám mohou pomoci zefektivnit proces vyhledávání informací.

Příklady vyhledávání:

**george washington, George Washington, GeOrGe WaShInGtOn** - nezáleží na tom, zda zadáte velká či malá písmena, všechny tyto tři dotazy vám vrátí stejné výsledky

**dovolená Hawai** - 2 a více zapsaných pojmů je defaultně spojeno operátorem AND, tzn. vyhledají se stránky, kde figurují všechna zadaná hesla

Google ignoruje tzv. **stop slova** jako jsou předložky a spojky, samostatná písmena... Pokud však chcete zahrnout stop slovo do vyhledávacího dotazu, označte ho znamínkem +.  
např: **Star Wars Episode +I** (I musí být ve vyhledaných stránkách obsaženo).

Druhou možností je zadat dotaz jako frázi: **"Star Wars Episode I"** (najde přesnou frázi).

**masmédiá OR média** - najde stránky minimálně s jedním termínem

**Klaus -Santa** - znaménko mínus slouží pro vyloučení slov, která nechceme ve výsledcích vyhledávání

**"\* ze Žerotína"** - hvězdička nahradí libovolné slovo

**~copyright** - hledá uvedený termín včetně synonymních výrazů, např. i intellectual property. Zatím lze použít pouze pro slovní zásobu v angličtině

**DVD player \$50..\$100** - vyhledá stránky obsahující numerické údaje v zadaném rozsahu

**3 \* 5** - lze zadávat i početní úkony, Google vrátí výsledek výpočtu

V Googlu lze vyhledávat i formou určitých operátorů (příkazů). Způsob zápisu dotazu ve vyhledávacím okně je vždy ve formě příkaz:zadání dotazu (mezera) další termín

Příklady:

**filetype:xls ceník** - omezení jen na určitý formát dokumentu, např. pdf, xls, doc, php...

**intitle:CPU Athlon** - najde "CPU" v názvu stránky a "Athlon" kdekoli

**allintitle:domov koupelny** - najde "domov" a "koupelny" v názvu stránky (současně)

**link:www.mendelu.cz** - najde stránky obsahující odkaz na stránky uvedeného zdroje

**related:www.zaskolou.cz** - najde tématicky podobné stránky k uvedené adrese

**site:www.mendelu.cz "přijímací řízení"** - na uvedené www adrese prohledá zdroje na téma přijímací řízení

**inurl:dřevo** - najde uvedené slovo v URL adrese

**define:"global warming"** - najde definici zadaného pojmu, použitelné pouze v angličtině

**stocks:IBM** - najde burzovní, firemní informace o zadané firmě

**info:www.mendelu.cz** - kromě odkazu na stránku vyhledá další rozšiřující informace vztahující se k zadané stránce

**cache:www.mendelu.cz** - zobrazí předchozí verzi stránky uloženou v archivu Google

Další tipy pro použití Googlu:

Stejně funkce jako Google, ale v přívětivém obrázkovém prostředí, nabízí stránka

<http://www.soople.com>.

Zajímavým doplňkem pro vyhledávání je **Google Desktop**. Je to volně stažitelný program, který slouží k prohledávání lokálních souborů a adresářů v různých programech na vašem počítači.

**Google Scholar** - <http://scholar.google.com>: prohledává odborné informace - vědecké články, technické zprávy, vysokoškolské kvalifikační práce. Pokud není volně dostupný plný text, zobrazí se alespoň abstrakty.

**Google Book Search** - <http://books.google.com>/vyhledá základní informace o vyšlých knižních titulech (recenze, vydavatelské údaje, odkazy na online obchody), možnost prohlížení údajů o knize v knihovních katalozích, prohlížení úryvků z textů knih, pokud to umožňuje autorské právo.



---

## <sup>i</sup> Internet vs. internet

V textech se setkáváme jak se slovem Internet s velkým „I“, tak se slovem internet s malým „i“. Existuje buď jenom Internet, nebo jenom internet? Neexistují náhodou oba dva s tím, že otázka by měla spíše znít: Má smysl mezi nimi rozlišovat? Toto slovo může být chápáno dokonce ve třech významech:

1. **internet** = propojené počítačové sítě
2. **Internet** vl. jm. = Celosvětová informační a komunikační síť
3. **internet** = nové médium

### Vysvětlení toho, proč existuje Internet a internet

Internet s velkým "I" je skutečně vlastním jménem, a to vlastním jménem jedné konkrétní počítačové sítě, která původně vznikla v USA a dnes má již celoplanetární dosah. **Internet s velkým "I" je tedy jenom jeden.** Naproti tomu internet (zkrácením z anglického "internetwork", česky třeba "sousítí") s malým "i" je obecné označení pro jakékoli spojení několika počítačových sítí - dvou, tří, nebo třeba stovky sítí. Svůj vlastní internet, třeba jen ze dvou miniaturních sítí, si tedy může postavit kdokoli a jeho vlastník pak má plné právo tvrdit, že skutečně je jeho vlastníkem - zatímco vlastník Internetu neexistuje.

Faktický rozdíl mezi Internetem a internetem může být v některých situacích velmi důležitý a významný. Například, když někdo někde napíše, že "internet má x tolik a tolik uživatelů", pak je to nic neříkající informace, protože z ní není jasné, který z mnoha internetů má autor na mysli.

### Internet jako nové médium

Dnes se stále výrazněji formuje nové médium - vedle televize, rozhlasu, tištěných novin atd. - které má elektronickou podobu, dokáže oslovit čím dál tím více lidí, umožňuje vydavatelům publikovat jejich obsah atd. a které velmi úzce souvisí s Internetem, ale na druhé straně s ním není totožné. Jaký je tedy přesný vztah Internetu k tomuto novému médiu? Pro zjednodušení uveďme příklad vztahu mezi službou World Wide Web (jen jedna z mnoha služeb, které lze provozovat nad Internetem jako globální přenosovou infrastrukturou) a Internetem: *Internet je jedním z možných technických řešení, které umožňují šíření obsahu prostřednictvím služby WWW (tj. šíření WWW stránek).* Ve vztahu k Internetu jako k novému médiu se tedy skutečně ustálilo psaní s malým „i“.

## [POUŽITÉ ZDROJE](#)