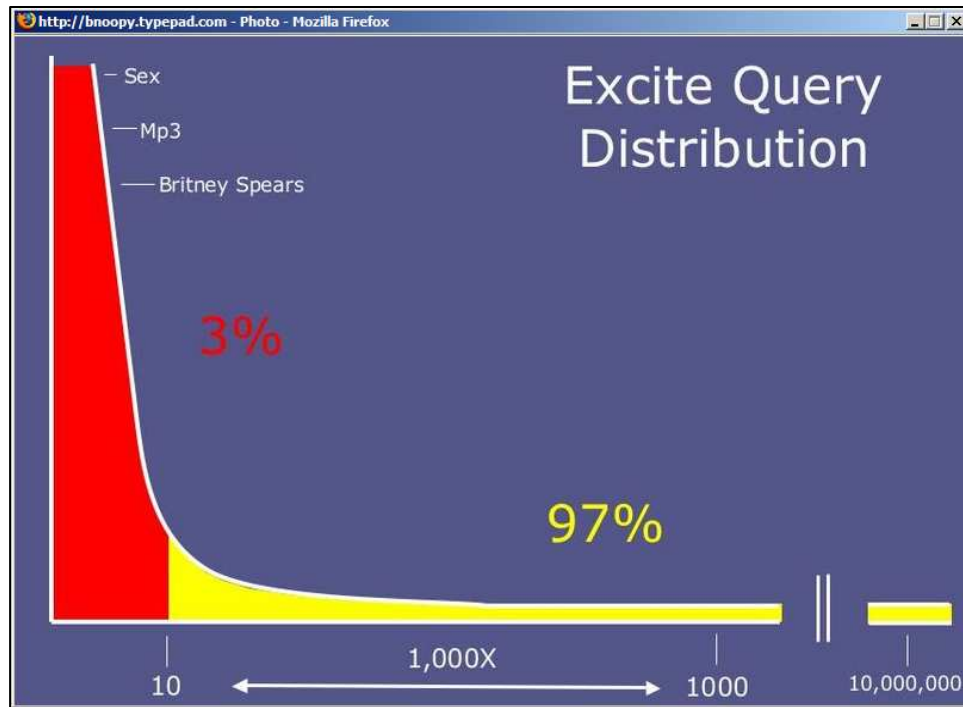


## Vyhľadávanie

Štatistika z miliárd uskutočnených vyhľadanií ukazuje, že ľudia zvyčajne kladú jednoduché otázky. 20% otázok obsahuje iba jedno slovo, ďalších 50% sa skladá z dvoch až troch slov. 95% ľudí nikdy nepoužije zložitejšie formy vyhľadávania.

Pritom sa nedá povedať, že by ľudia kladli vyhľadávačom tie isté otázky. Keď si pozrieme graf závislosti frekvencia výskytu otázky (vertikálna os) a celkového počtu otázok (horizontálna os), uvidíme, že väčšina otázok sa nikdy neopakuje. Otázky, ktoré kladie množstvo ľudí, reprezentujú iba 3% všetkých otázok.



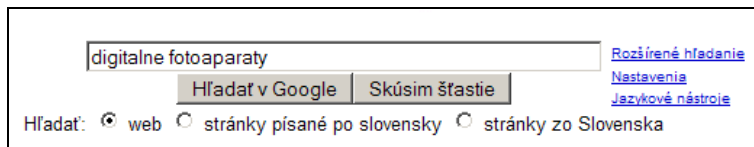
[http://bnoopy.typepad.com/bnoopy/2005/03/the\\_long\\_tail\\_o.html](http://bnoopy.typepad.com/bnoopy/2005/03/the_long_tail_o.html)

Pre cvičenie techník vyhľadávania sme vybrali vyhľadávač firmy Google. Ide v súčasnosti o najznámejší a celosvetovo najpoužívanejší vyhľadávač. Základnú vstupnú stránku vyhľadávača Google používajú stovky miliónov ľudí na celej planéte. Už dizajn tejto stránky spôsobil malú revolúciu. Neobsahuje nič iné, ako logo firmy Google a jedno malé jednoriadkové okienko pre zápis otázky.



### Základná logika vyhľadávania

Predstavme si, že si chceme kúpiť digitálny fotoaparát a hľadáme, aké výrobky sú na trhu. Otvoríme si vyhľadávač Google ([www.google.sk](http://www.google.sk)) a do okienka napíšeme svoju otázku.



Google za menej ako pol sekundy nájde temer 5 miliónov stránok a zobrazí linky na prvých niekoľko s krátkou textovou anotáciou. Okrem nájdených stránok ponúkne aj stránky, ktoré sú síce tiež k téme, ale sú platenou reklamou firiem (sponsored links).

The screenshot shows a Google search for "digitálne fotoaparaty". The search bar contains the text "digitálne fotoaparaty" and the "Hľadať" button is visible. Below the search bar, there are radio buttons for "web", "stránky písané po slovensky", and "stránky zo Slovenska". The search results are displayed in a grid format, with several sponsored links and organic results. The sponsored links include "Digitálne Fotoaparaty" from pixmania.sk, "Najlepší fotoaparát?" from Fotolab.sk, "Fotoaparáty na Datart.sk", "Megafoto - digitálne fotoaparáty a príslušenstvo", and "Digitálne fotoaparát Beng". The organic results include "Digitálne fotoaparáty" from talkive.sk, "Zrkadlovka Canon EOS 500D", and "Digitálne fotoaparáty" from FotoExpert.sk.

Ak zadám hľadané slová v úvodzovkách:

The screenshot shows a Google search bar with the text "digitálne fotoaparaty" enclosed in double quotes. The "Hľadať" button is visible, and the search options below are the same as in the previous screenshot.

vyhľadávač nájde iba desatinu stránok. Dôvod odlišného výsledku je v tom, že v prvom prípade nájde také stránky, na ktorých sa vyskytujú hľadané slová kdekoľvek na stránke. V druhom prípade hľadá iba stránky, kde sa vyskytuje slovné spojenie „digitálne fotoaparáty“.

Vyhľadávače umožňujú vytvárať rôzne logické výrazy, podľa ktorých pracujú. Nasledujúca tabuľka popisuje najzákladnejšie typy týchto logických výrazov aj s ich interpretáciami:

Otázka	Interpretácia
<i>digitálne fotoaparáty</i>	Hľadá stránky, kde sa vyskytujú hľadané slová kdekoľvek na stránke
<i>“digitálne fotoaparáty“</i>	Hľadá stránky, kde sa vyskytuje presné slovné spojenie <i>“digitálne fotoaparáty“</i>
<i>digitálne OR fotoaparáty</i>	OR -> aspoň jeden z Hľadá stránky, kde sa vyskytujú aspoň jedno zo slov
<i>digitálne fotoaparáty (OLYMPUS OR SONY)</i>	(... OR ...) -> aspoň jeden v zátvorke s ostatnými slovami Hľadá stránky, kde sa vyskytujú obe slová <i>digitálne fotoaparáty</i> a aspoň jedno zo slov v zátvorke kdekoľvek na stránke.

digitálne fotoaparáty -OLYMPUS	- -> negácia Hľadá stránky, kde sa vyskytujú obe slová <i>digitálne fotoaparáty</i> a nevyskytuje sa slovo OLYMPUS (- predstavuje negáciu)
-ape	~ -> hľadá všetky synonymá Hľadá výskyt všetkých synonym anglického slova ape (monkey, gorilla, chimpanze). Synonymá vypisuje tučným písmom. Google vyhľadáva synonymá algoritmicky, čo môže byť niekedy prekvapujúce.
"digitalne fotoaparaty" 5..8 Mpx	5..8 -> číselný interval od 5 po 8 Hľadá stránky, kde sa nachádzajú <i>digitálne fotoaparáty</i> s rozlíšením v intervale 5 až 8 megapixelov.
"digitalne fotoaparaty" ..8 Mpx "digitalne fotoaparaty" 5.. Mpx	..8 -> menší ako 8 5.. -> väčší ako 5 Hľadá stránky, kde sa nachádzajú <i>digitálne fotoaparáty</i> s rozlíšením <b>menším, ako 8</b> megapixelov (väčším, ako 5 megapixelov)
"digitalne * fotoaparaty "	A * B -> ľubovoľné jedno slovo medzi A a B A ** B -> ľubovoľné dve slová medzi A a B A *** B -> ľubovoľné tri slová medzi A a B <i>"digitalne * fotoaparaty"</i> - hľadá stránky, kde sa nachádza slovné spojenie: <i>digitálne XXXX fotoaparaty</i> .

**Úloha:** Vyskúšajte si syntax vyhľadávania z predchádzajúcej tabuľky na vlastnom príklade a pozrite si počty nájdených stránok.

### Špeciálna syntax vo vyhľadávaní

Často by sme radi obmedzili nájdené stránky aj podľa ďalších kritérií, napríklad podľa typu súborov (textový, obrázkový, hudobný). Aj toto môžeme ovplyvňovať priamo pri zadaní otázky vyhľadávaču tak, že do otázky vložíme jedno alebo aj viac špeciálnych obmedzení. Zoznam najznámejších obmedzení popisuje nasledujúca tabuľka.

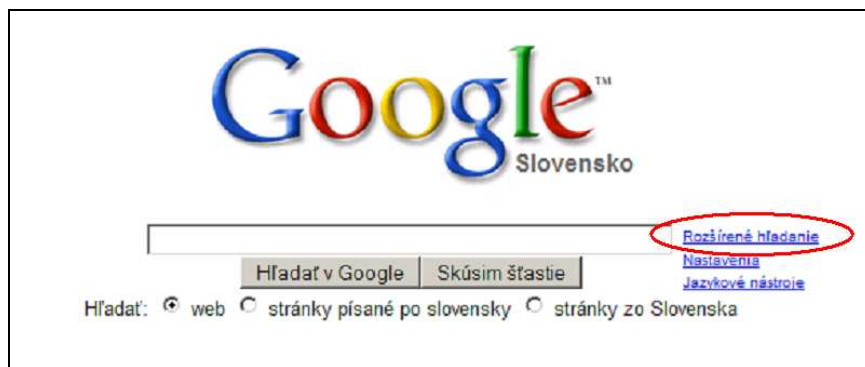
Obmedzenie	Interpretácia
intitle:"digitalne fotoaparaty"	intitle:aaa -> <i>aaa</i> sa musí nachádzať v titulku V titulku stránky sa vyskytuje spojenie <i>digitalne fotoaparaty</i>

intext:"ked vam pride email z adresy"	intext:aaa -> <i>aaa</i> sa nachádza v texte V texte stránky sa vyskytuje slovné spojenie " <i>ked vam pride email z adresy</i> "
inanchor:"vystúpenie v Bratislave"	inanchor:aaa -> <i>aaa</i> sa musí nachádzať v texte v hyperlinku V texte hyperlinku sa vyskytuje slovné spojenie " <i>vystúpenie v Bratislave</i> "
site:sme.sk	site:aaa -> <i>aaa</i> sa byt' adresa Adresa hľadanej stránky musí končiť na <i>sme.sk</i>
link:fmph.uniba.sk	link:aaa -> stránky, ktoré obsahujú link na stránku s URL <i>aaa</i> Stránky, ktoré obsahujú link na stránku s <i>fmph.uniba.sk</i>
cache:www.inzine.sk	cache:aaa -> nájde kópiu stránky <i>aaa</i> uloženú v Google, aj keď skutočná stránka <i>aaa</i> môže byť už zmenená alebo zrušená
filetype:doc	filetype:xxx -> vyhľadávanie sa obmedzí iba na súbory s príponou <i>xxx</i>
related:sme.sk	related:aaa -> stránky, ktoré sú v nejakom vzťahu k stránke <i>aaa</i>
define:C++	define:aaa -> hľadá definície slova <i>aaa</i>
movie:Predcitac	movie:aaa -> hľadá recenzie filmu s názvom <i>aaa</i> Nielen nachádza recenzie filmu Predčítavač, ale aj ponúka premietacie časy pre miesto, kde sa nachádzate.
music:Kate Bush	music:aaa -> hľadá akékoľvek k hudbe sa vzťahujúce informácie (autorov, interpretov, skladby, albumy, texty a pod.)

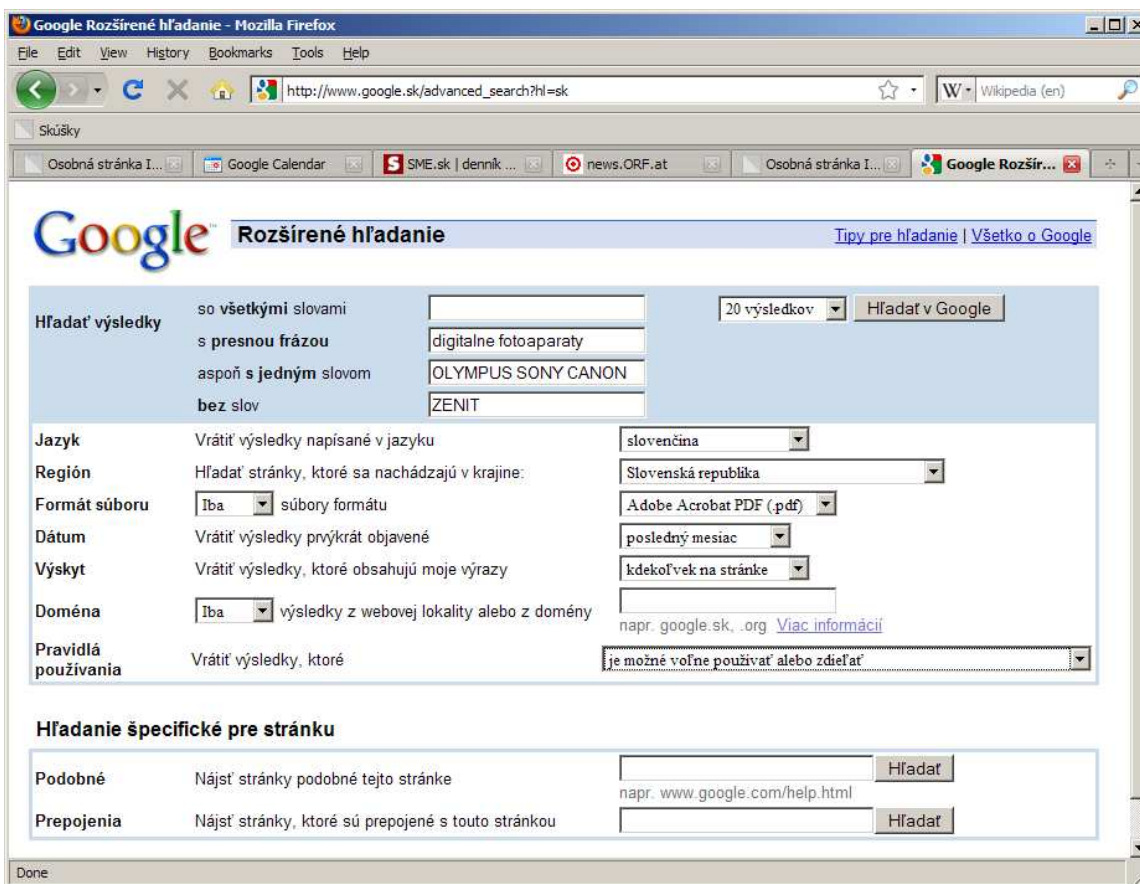
**Úloha:** Vyskúšajte si obmedzenia z predchádzajúcej tabuľky na vlastných príkladoch a porovnajte si rovnaké vyhľadávania bez obmedzení a s obmedzeniami.

### **Rozšírené hľadanie**

Google umožňuje aj detailnejšie špecifikovanie toho, čo sa snažím nájsť. Tomu slúži prechod na rozšírené hľadanie.



Rozšírené hľadanie nám umožňuje detailne špecifikovať, čo chcem hľadať, s akými ďalšími kritériami bez toho, aby som musel ovládať špeciálnu syntax:



Úloha: Vyskúšajte si predchádzajúce úlohy za pomoci formuláru pre Rozšírené hľadanie.

### Skúsím šťastie

Ak sa rozhodnete, že nechcete vidieť zoznam nájdených stránok, ale chcete ísť rovno na najvyššie vyhodnotenú stránku, urobíte to stlačením tlačítka Skúsím šťastie.



Úloha: Vyskúšajte si to pre niekoľko rôznych prípadov.

### Špeciálne služby vyhľadávača Google

Vyhľadávač Google je možné použiť aj pre niektoré špeciálne služby, ktoré umožňujú používať ho, ako jednoduchú kalkulačku, prekladač z jazyku do jazyku alebo napr. sledovanie najpopulárnejších trendov vo vyhľadávaní.

#### Kalkulačka

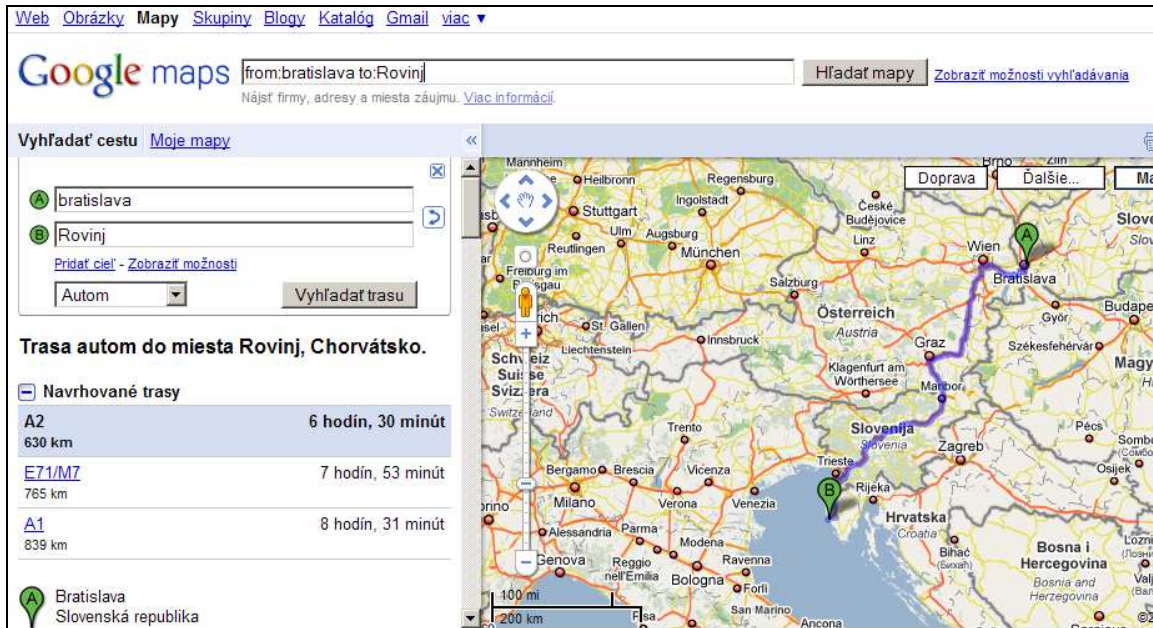


Kalkulačku je možné použiť nielen na jednoduché numerické výpočty, ale aj na rôzne konverzie, napr. rôznych mier, váh, kurzov a pod.

Zadanie	Výsledok
$60*60*24*365$	$60 * 60 * 24 * 365 = 31\,536\,000$
15 km in miles	15 kilometers = 9.32056788 miles
$25*9+(\text{sqrt } 10)^3$	$(25 * 9) + (\text{sqrt}(10)^3) = 256.622777$
14 Euros in czech koruny	14 Euros = 362.899803 Czech koruny

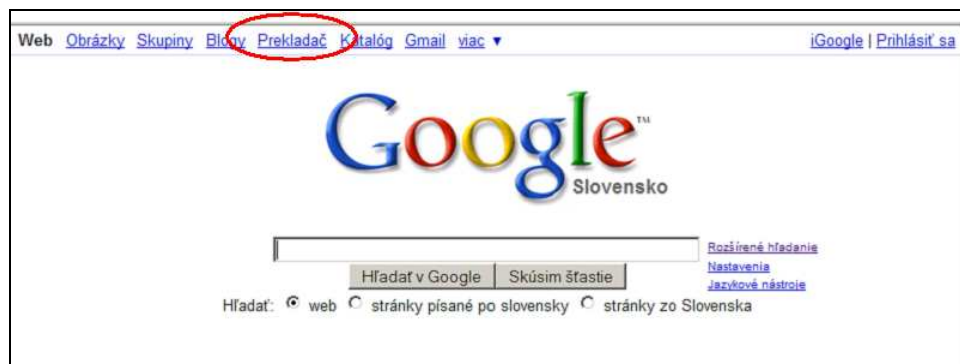
#### Mapy

Aplikácia Google Maps (<http://maps.google.com/maps?hl=sk>) umožňuje mimo mnohého iného nájsť si trasu cesty z miesta na miesto.

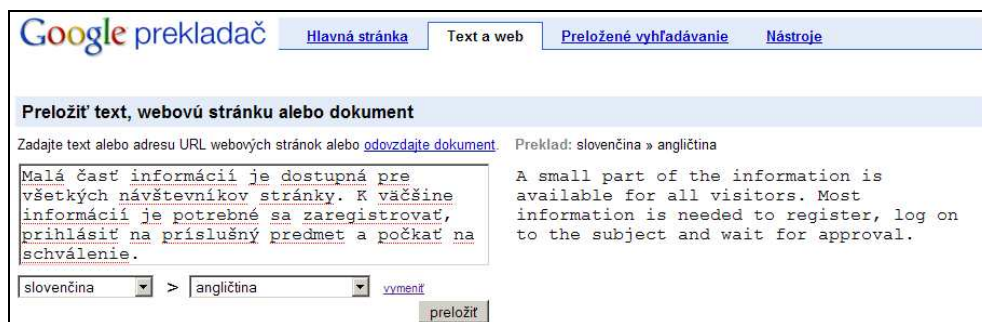


## Prekladač

Google umožňuje dať si preložiť ľubovoľný text z vyše 50 jazykov do iného jazyku.



Po výbere tejto možnosti sa dostaneme do prekladateľského formuláru, kde môžeme vložiť text a vybrať si, z akého jazyku do akého jazyku ho chceme preložiť.



## Zeitgeist

Google umožňuje prezerat' si najčastejšie otázky, ktoré ľudia kladú. Je možné si pozerat' aktuálne výsledky, aj porovnávat' historické trendy.

(<http://www.google.com/insights/search/#>)



Google Insights for Search Pomocník | Prihlásiť sa | Slovenčina ▾  
beta

Porovnať podľa:	Vyhľadávané výrazy	Filter
<input checked="" type="radio"/> Vyhľadávané výrazy <input type="radio"/> Polohy <input type="radio"/> Časové obdobia	Tip: Parameter ALEBO vyjadrite pomocou znaku plus (príklad: tenis + squash) <input type="text" value="Všetky vyhľadávané výrazy"/> <a href="#">+ Pridať vyhľadávaný výraz</a>	<input type="text" value="Vyhľadávanie na webe"/> <input type="text" value="Na celom svete"/> <input type="text" value="2004 - súčasnosť"/> <input type="text" value="Všetky kategórie"/>
		<input type="button" value="Hľadať"/>

Zistite, čo svet vyhľadáva.

Pomocou nástroja Google Insights for Search môžete porovnať modely počtu vyhľadávaní naprieč konkrétnymi regiónmi, kategóriami, časovými intervalmi a produktmi. Pozrite si [príklady](#) používania nástroja Google Insights for Search.

**Kategórie**  
 Spresnite údaje podľa určitých kategórií, napríklad financie, zdravie alebo šport.  
 Príklady: [Kategória Zábava v USA \(za posledných 30 dní\)](#) | [Najslávnejšie celebrity v New Yorku \(za posledných 30 dní\)](#)

**Sezónnosť**  
 Predvídajte dopyt po vašich produktoch, aby ste mohli vhodne vytvoriť rozpočet a plánovať.  
 Príklady: [Lyžovanie verus surfovanie](#) | [Hotel v rokoch 2009, 2008, 2007, 2006](#)

**Geografická distribúcia**  
 Zistite, kde nájdete svojich zákazníkov. Pozrite si, ako sa objem vyhľadávania distribuuje v rôznych regiónoch a mestách.  
 Príklady: [Futbal v Brazílii, Taliansku, Nemecku, Spojenom kráľovstve](#) | [Lyžovanie v Austrálii verus vo Švajčiarsku](#)

**Vlastnosti**  
 Pozrite si modely vyhľadávania na iných stránkach vo vlastníctve spoločnosti Google.  
 Príklady: [Stúpajúce vyhľadávania obrázkov vo Francúzsku \(za posledných 30 dní\)](#) | [Štípatka verus mačiatka vo vyhľadávaní obrázkov v USA](#)

**Viac príkladov**

[Shiraz, Syrah, Merlot](#)  
[Beatles, Elvis](#)  
[Chrípka](#)  
[Čokoládová zmrzlina, vanilková zmrzlina](#)  
[Firefox, Internet Explorer, Safari, Opera, Chrome](#)  
[Facebook, Orkut, Myspace, hi5](#)  
[Modrá, červená, žltá, zelená](#)  
[Blog, noviny](#)  
[Hurikán Katrina, ceny pohonných hmôt](#)  
[perl, python, ruby, php](#)  
[Zemetrasenie](#)

Tieto analýzy je možné si pozrieť špecificky: pre krajiny, regióny, časové trendy, témy, jednotlivé domény a pod.

**Úloha:** Vyskúšajte si kalkulačku. Nechajte si preložiť nejaký svoj vlastný text zo slovenčiny do jazyku, ktorý ovládáte a naspäť. Nájdite si cestu z Vášho bydliska na pracovisko (do školy). Pozrite si najhladanejšie slová za posledný mesiac.

## Výstup

**Záverečná úloha:** Do **30.11.** publikujte na vašom blogu článok, v ktorom zhrniete, ako ste pri riešení jednotlivých úloh postupovali a čo ste sa pritom naučili.