# OP3BP_DKTP Komunikační technologie - přednáška

Vyhledávací techniky na Google.com

Nástroje Google.com

Google Bombs

# Vyhledávací techniky na Google.com

- Internetový vyhledávač
  - služba, která umožňuje na Internetu najít webové stránky, které obsahují požadované informace
  - uživatel zadává do vyhledávače klíčová slova
  - vyhledávač na základě své databáze vypisuje odkazy, které hledané informace obsahují
    - databáze udržována převážně automaticky na rozdíl od internetových katalogů
  - cílem je poskytnout při odpovědi na dotaz co nejrelevantnější informace
    - různá důležitost webových stránek - PageRank

# Vyhledávací techniky na Google.com

- Jak vyhledávač pracuje?
  - automaticky, využívá desítky až statisíce počítačů
  - kvalita vyhledávače = jak kvalitní dá odpovědi => uživatel najde hledanou informaci na prvních místech
  - měřit kvalitu stránek, které vyhledávač má ve své databázi (např. PageRank u Google, S-Rank u Seznamu, JyxoRank u Jyxo)
  - majitelé modifikací svých stránek dosáhnout co nejvyšší pozice ve výstupu vyhledávače (SEO)

# Vyhledávací techniky na Google.com

- výsledkem = vyhledávač musí své metody vylepšovat, aby vyhověl čím dál vyšším požadavkům svých návštěvníků a odstranil podvodníky
- většina internetových vyhledávačů pracuje ve třech krocích
  - procházení webových stránek
  - vytvoření databáze výskytu slov
  - indexování
  - poskytování odpovědí na dotazy

# Vyhledávací techniky na Google.com

- Databáze výskytu slov
  - data uložena na disk se kategorizují do databází
  - v databázi jsou uvedena všechna nalezená slova a adresy k nim, na kterých se tato slova vyskytují
  - protože sekvenční prohledání databáze by trvalo dlouho, následuje další krok, tzv. indexace

# Vyhledávací techniky na Google.com

- SEO (anglicky ***S***earch ***E***ngine ***O***ptimization)
  - aby se web umístil ve výsledcích vyhledávání co nejvýše
  - SEO techniky se rozlišují na *„povolené“* a *„zakázané“*
    - tzv. Black Hat SEO, které vyhledávače tvrdě postihují například vyřazením ze svého indexu
  - jakékoliv umělé zlepšování umístění ve výsledcích vyhledávání nežádoucí

# Vyhledávací techniky na Google.com

- ***«výraz»***
  - Prosté hledání, najde všechny stránky které obsahují «*výraz*», i skloňovaný
  - ***Př.: sluneční brýle***
- ***«výraz» -«výraz2»***
  - Prosté hledání, najde všechny stránky které obsahují «*výraz*» i skloňovaný a současně neobsahují «*výraz2*»
  - ***Př.: sluneční brýle -relax***
    - najde sluneční brýle mimo značky Relax

# Vyhledávací techniky na Google.com

- ***«výraz» + «výraz2»***
  - Najde všechny stránky které obsahují «*výraz*» i skloňovaný a současně obsahují «*výraz2*»
  - ***Př.: sluneční brýle +relax***
    - najde sluneční brýle vč. značky Relax
- ***«výraz» ****
  - Pro doplnění chybějících údajů sám stačí přidat znak hvězdičky (*) do věty či otázky, která má být doplněna
  - ***Př.: jaromír jágr vstřelil ****

# Vyhledávací techniky na Google.com

- ***"«výraz»"***
  - Google najde přesně «*výraz*» jak je mezi uvozovkami (i s chybami, ale sám je opravuje)
  - ***Př.: "sluneční brýle"***
- ***«výraz» site:cz***
  - hledání «*výraz*» pouze na českých stránkách
  - ***Př.: relax site:cz***

# Vyhledávací techniky na Google.com

- ***«výraz» site:auto.cz***
  - hledání «*výraz*» pouze na stránkách které mají ve své adrese auto.cz
  - ***Př.: relax site:www.maniashop.cz***
    - hledá brýle Relax na webu www.maniashop.cz
- ***define:«výraz»***
  - Hledá definici «*výraz*» ve slovníku (nejčastěji Wikipedie)
  - ***Př.: define:hmotnost***

# Vyhledávací techniky na Google.com

- ***link:«výraz»***
  - Hledání odkazu na «*výraz*»
  - ***Př.: link:www.muni.cz***
    - vypíše seznam všech webů které odkazují na www.muni.cz
- ***«výraz» filetype:pdf***
  - hledání dokumentů s názvem «*výraz*».doc
  - ***Př.: navod filetype:pdf***
    - hledá soubor navod.pdf

# Vyhledávací techniky na Google.com

- ***«výraz» «dolní_mez»..«horní_mez»***
  - hledá «*výraz*» v rozsahu hodnot od «*dolní_mez*» do «*horní_mez*»
  - ***Př.: lyže za 10000..15000***
    - hledá lyže v cenové relaci 10 000 až 15 000 Kč; ?Úskalí?
- ***«výraz1» v «výraz2»***
  - převod jednotek
  - ***Př.: 10.5 m v palcích***
    - převede jednotky metry na palce, i měny
  - ***Př.: 34.6 EUR in CZK***

# Vyhledávací techniky na Google.com

- Počítání
  - ***15+20-5*4/2 = 15 + 20 – ((5 * 4) / 2) = 25***
  - ***log 100 = 10***
  - ***ln 1000 = 6,90775528***
  - ***pi = 3,1415***
  - ***e = 2,71828183***
  - ***sin (pi/2) = 1***
  - ***5! = 120***
  - ***3^4 = 81*** nebo lze alternativně ***3**4***

# Vyhledávací techniky na Google.com

- ***27^(1/3) = 3*** nebo lze alternativně ***27**(1/3)***
  - ***4th root of 16*** – čtvrtá odmocnina z 16
  - ***cube root of 109*** – třetí odmocnina ze 109
  - ***square root of 42*** – druhá odmocnina ze 42
  - ***sqrt(42)*** – druhá odmocnina ze 42
    - ***15 mod 9*** nebo ***15 % 9 = 6*** – modulo (zbytek po dělení)
- Vykreslování funkčních závislostí
  - ***sin(x)*** ***2x + x**2*** ***ln(x)***
  - Více průběhů pomocí logické spojky AND
  - ***Př.: 5*sin(5x) AND sin(x+50) AND (x)**2 AND ln(x)***

# Vyhledávací techniky na Google.com

- Počítání s fyzikálními jednotkami
  - Práce: 150N * 50m = 7500 joules
    30 m/s * 45s = 1,35 kilometru
    80kg * 10N/kg = 800 newtons

# Google Bombs

- Princip objevil Adam Mathes v dubnu 2001
- V roce 2007 Google vývojáři přijali opatření
- Zneužití vyhledávače, aby vracel na dotaz jiné výsledky než je žádoucí
  - web která obsah dotazu neobsahuje, ale útočník na ni tak odkazuje
- Obvykle s humorným nebo politicko úmyslem
- Jsou spíše krátkodobé

# Google Bombs

- Jak se to dá udělat?
  - Googlu zajímá nejen k obsah stránek, ale i text který na ně odkazuje = jak se o webu píše na jiných webech
  - Pokud mnoho stránek (20 až 30 = použitím blogů) obsahuje odkaz na web s urážlivým textem (např. „děd Vševěd“) vrátí Google mezi výsledky i cílovou stránku
  - Je-li odkazů mnoho, cílová stránka vyjde na první místa ve výsledcích

# Google Bombs

- Výsledky hledání pak lze interpretovat takto:
  - „***Google se domnívá, že web pojednává o dědovi Vševědovi***“
- Správnější je ale:
  - „***Mnoho stránek uvádí, že tato stránka pojednává o dědovi Vševědovi***“

# Nástroje Google.com

- Vyhledávání
  - textů
  - textových souborů
  - obrázků
  - zdroj obrázků

- Překladač
- Počasí (např.: počasí Brno)

# Nástroje Google.com

- G-Mail *
- G-Docs *
- G-Calendar *
- Google Knihy (books.google.com)
  - Vyhledávání knih
  - Vyhledávání časopisů
- Počítání
- YouTube