

# Vyhledávání dat

# Co hledáme?

- „hotová“ data
- Informaci

# Co hledám

- Téma
- Geografická lokace
- Časová lokace
- Úroveň – volič, obec, stát
- Hotový dataset x jednotlivá informace
- Frekvence dat

# Vyhledávání

- Google
  - Strategie záleží na tom co hledáme
    - Co nejpřesnější specifikace
    - správná klíčová slova
    - ne obecná slova a termíny bez jednoznačného významu
    - konkrétní popisná slova
    - Různý dotaz v různých zdrojích
    - Jazyk

# Vyhledávání Program obnovy venkova

- Uvozovky "Program obnovy venkova" – přesná shoda s danou kombinací slov v daném pořadí, lze označit jedno slovo
- Logické nebo "Program obnovy venkova" **OR** "Program rozvoje venkova" – vyhledá stránky, obsahující jeden, druhý, nebo oba vyhledávané výrazy. Místo operátoru OR lze použít také |.
- Logické A "Program obnovy venkova" **AND** "Program rozvoje venkova" – vyhledá stránky, obsahující oba výrazy.
- Mínus "Program obnovy venkova" -"Program rozvoje venkova" – mínus před klíčovým slovem ho vylučuje z vyhledávání. Namísto mínus lze také využít výraz NOT.
- Hvězdička "Program \* venkova" – hvězdička nahrazuje jedno nebo více slov. Vyhledá stránky, které mezi zadanými klíčovými slovy mají libovolná slova nebo věty.
- Závorky (Program (obnovy OR rozvoje) venkova) – závorky určují prioritu – při vyhledávání se jako první používá výraz ze závorek, teprve poté se bere ohled na další zadání.
- Dvě tečky Program obnovy venkova 2016..2020 – dvě tečky zastupují číselnou řadu, výsledky by měly obsahovat zadaný výraz a čísla 2016, 2017, 2018, 2019 nebo 2020
- Typ souboru "Program obnovy venkova" filetype:xls (další možnosti jsou pdf, doc, xlsx, docx)
- Konkrétní web "Program obnovy venkova" site:mmr.cz

# https://www.google.com/



# Google

Hledat Googlem

Zkusím štěstí

- Nastavení vyhledávání
- Rozšířené vyhledávání
- Vaše data ve Vyhledávání
- Historie vyhledávání
- Vyhledat v nápovědě
- Odeslat zpětnou vazbu
- Tmavý motiv: Vypnuto

Rozšířené vyhledávání

Kritéria vyhledávání stránek

všechna tato slova:

přesně toto slovo nebo slosovo:

kterékoli z těchto slov:

žádné z těchto slov:

čísla od:

do

Kritéria zúžení okruhu výsledků

jazyk:

kterýkoli jazyk

oblast:

všude

poslední aktualizace:

kdykoli

web nebo doména:

výskyt výrazů:

kdekoli na stránce

typ souboru:

libovolný formát

Nastavení

Česko

O společnosti Google Reklama Firma Jak funguje Vyhledávání

Třetí desetiletí našich opatření v oblasti ochrany klimatu

Ochrana soukromí Smluvní podmínky

# Příklad - vyhledání dotací v Jmk

Program obnovy venkova v Jihomoravském kraji

„Program obnovy venkova“ v Jihomoravském kraji

„Program obnovy venkova“ v Jihomoravském kraji filetype:xls

„Program obnovy venkova“ v Jihomoravském kraji filetype:xlsx

„Program obnovy venkova“ filetype:xls site: dotace.kr-jihomoravsky.cz

„Program obnovy venkova“ filetype:xlsx site: dotace.kr-jihomoravsky.cz

„Program obnovy venkova“ filetype:pdf site: dotace.kr-jihomoravsky.cz

# K vyzkoušení

- Excelové tabulky Programu obnovy venkova v Karlovarském kraji



# webarchiv

- <https://web.archive.org/>
- Obvykle je dostupné, jen to co bylo „vidět“
- Nefungují linky
- Nejsou dostupné soubory ke stažení
- Často neobsahuje všechny podúrovně

<https://datasetsearch.research.google.com/>

- Nová služba
- Spuštěno 2020
- Lepší se fungování
- Prochází hlavně repozitáře
  
- Vyzkoušejte:
  - volená strana
  - Party preference
  - Party preference Czech Republic

# Datové archivy

- re3data.org: Registry of Research Data Repositories
  - <https://www.re3data.org/>
- <https://dataverse.harvard.edu/>
- <https://data.mendeley.com/>
- <https://osf.io/>
  
- github: <https://github.com/search?q=datasets>
- Kaggle: <https://www.kaggle.com/datasets?search=elections>
- <https://data.europa.eu/>

# Otevřená data

- <https://data.europa.eu/en>
- <https://data.gov.cz/>
- <https://data.brno.cz/>

# Data vztažená k publikovaným textům

- Sekundární odkazy
  - Články v databázích/google scholar
    - Local elections candidates czech republic
  - wiki

# Příklad exit poll belgických lokálních voleb 2012

# Záludnosti nalezených dat

- Věrohodnost
  - Pravdivost
  - Srovnatelnost
  - Úplnost
- 
- Kontrola metadat: čas a způsob sběru, instituce provádějící sběr, někdy i zdroj financování
  - Srovnání s jinými daty na dané téma (v jiném čase nebo zemi)

- SVĚT EXISTUJE I MIMO INTERNET a někdy i mimo počítače (archivy, knihy, místa, lidé, ...)