

Vyhledávání dat

Doplňěk k minulé hodině

- Jiný seznam datových archivů
 - <https://ukdataservice.ac.uk/help/other-data-providers/data-archives/european-data-archives/>
- Obsahuje i Španělsko

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

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: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

Datové archivy – přidat openscience, nachystat ukázkou

- 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/>

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, ...)