Opsætning af SQL-database (Fakenews data)

Dette dokument tager udgangspunkt i fakenews datasættet

# Klargøring af data

Når man har datasættet i csv format skal man først opdele det i mindre dele og ryde op i data’et. Dette kan gøre med scriptet: [https://github.com/JonathanAhrenkiel-Frellsen/Milestone\_Datascience/blob/master/Organize\_Daniel\_crash.ipynb](https://github.com/JonathanAhrenkiel-Frellsen/Milestone_Datascience/blob/master/Organize2_Daniel.ipynb)

Dette er det script vi brugte i milestone opgaven.

# Opsætning af værktøjer

Efter det skal man opsætte sin database. Dette kan man gøre i pgadmin eller psql terminalen. Jeg tager udgangspunkt i psql! De credentials man skal bruge i psql er bare default og derefter din kode (koden behøver ikke at være lang, det er normalt at den bare er ”root”).

# Opstilling af databasen

Før man opstiller databasen er det i en god ide at skrive ”\d” i psql, dette giver alle tables. Hvis den er tom kan du forsætte, hvis du vil slætte et table skal man skrive ”DROP TABLE [table\_name] CASCADE;” [table\_name] er navnet på det table man vil slætte, CASCADE er ikke nødvendigt, kun hvis man slætter et table som har relation til andre tables eller omvent. HUSK SIMICOLON jeg glemmer det altid 😉

Når eller hvis du har slættet de tables som ikke skal være der kan du skrive komandoen ”\i [path\_to\_sql\_file].sql;” eksempel: ”\i C:/Users/jola1/Desktop/ERDreal.sql;” tor det er vigtigt at skrive ”/” og ikke ”\”, men er ikke sikker. Efter det kan du skrive ”\d” igen for at se om dine tables er lavet.

# Populate database

Nu har du opsættet csv filerne fra afsnittet ”Klargøring af data” og opsættet databasen i afsnittet ”Opstilling af databasen” nu kan du overfør dataet fra csv filerne til databasen med psql. Dette gør man med komandoen: “\COPY [table\_name] FROM '[csv\_file\_position]’ DELIMITER ',' CSV HEADER;

Her er et eksempel:

\COPY article(id,domain\_id,type\_id,url,content,title,meta\_description,scraped\_at,updated\_at,inserted\_at) FROM 'C:/Users/jola1/Desktop/Milestone\_Datascience-master/article\_clean.csv' DELIMITER ',' CSV HEADER;

eller

\COPY article FROM 'C:/Users/jola1/Desktop/Milestone\_Datascience-master/article\_clean.csv' DELIMITER ',' CSV HEADER;

Den sql fil som jeg bruger kan downloades fra her:

<https://discordapp.com/channels/@me/690271126944874496/714802715682668585>