

2^η ΕΡΓΑΣΙΑ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ 2022

ΜΠΙΛΙΑΣ ΓΕΩΡΓΙΟΣ - 3200278

ΜΠΟΥΡΑΖΑΝΑΣ ΠΑΝΑΓΙΩΤΗΣ - 3200268

Για την 2^η εργασία στις Βάσεις Δεδομένων αρχικά δημιουργήσαμε τα **.sql** αρχεία μέσω **cmd**, με την χρήση του **python** αρχείου που μας δόθηκε (**gen_ddl_python3**) και των **csv** αρχείων ή από μόνοι μας, και στην συνέχεια κάναμε τις κατάλληλες τροποποιήσεις (π.χ. θέσαμε τα **Primary Keys**). Για το **keywords.csv** με την χρήση της **python** δημιουργήσαμε ένα πρόγραμμα που έκανε διάσπαση των **json** και δημιουργούσε δυο νέα αρχεία (**keyword** και **movie_keywords**) με συμπληρωμένα τα κατάλληλα πεδία στο καθένα. Αναλυτικότερα, αυτό το **python** πρόγραμμα ουσιαστικά δημιουργούσε δυο νέα **csv** αρχεία για γραφή καθώς και άνοιγε το **keywords.csv** για ανάγνωση. Στη συνέχεια, με την χρήση του **csv.reader** διάβαζε αυτό το αρχείο και με την χρήση 2 **csv.writer** έγραφε τα κατάλληλα στοιχεία στο καθένα απ'τα 2 αρχεία που δημιούργησε ξεχωριστά μέχρι να γίνει ανάγνωση ολόκληρου του **keywords.csv** αρχείου. Επίσης, δημιουργήσαμε και ένα **tempo.sql** αρχείο που θα μας χρειαστεί για την αφαίρεση διπλοτύπων απ'το **keyword.csv**.

Αφού καταλήξαμε να έχουμε όλα τα **.sql** αρχεία που χρειαζόμασταν, στην συνέχεια, εισήγαμε ένα-ένα αυτά τα αρχεία στο **Query Tool** του **PgAdmin** και το τρέχαμε. Έτσι, καταλήξαμε να έχουμε όλα τα **tables** που θέλουμε. Μετά απλά κάναμε **import** με την χρήση της **Import/Export** του **PgAdmin** τα **.csv** αρχεία μέσα στα ανάλογα **tables** εκτός από το **keyword** όπου το κάναμε **import** μέσα στο **tempo** και στην συνέχεια μέσω του **Query Tool** κάναμε **run** το **create_keyword.sql** όπου ουσιαστικά έκανε αφαίρεση των διπλοτύπων. Αναλυτικότερα:

Για το **keyword** αρχικά δημιουργήσαμε το **tempo.sql** κάπως έτσι:

```
CREATE TABLE tempo(  
    id int,  
    name varchar(50)  
);
```

Στη συνέχεια, δημιουργούσαμε το **table** μέσα στο **PgAdmin** και κάναμε **Import** το **keyword.csv**.

Μετά απλά δημιουργούσαμε το **Keyword table** κάπως έτσι ώστε να έχει τα αρχεία του **keyword** χωρίς τα διπλότυπα:

```
CREATE TABLE Keyword AS  
SELECT DISTINCT id, name  
FROM tempo;
```

Και απλά μετά κάναμε διαγραφή του **tempo table** στο **PgAdmin** αφού μας ήταν πλέον άχρηστο.

Τέλος, δημιουργήσαμε ένα **alter_tables.sql** αρχείο όπου μέσα σ' αυτό βάζαμε τα ξένα κλειδιά στα **tables** και το **id** στο **Keyword** ως **Primary Key**.