

# Database Project Report

---

*authors: Χρήστος Αργυρόπουλος(3170010), Σπυρίδων Μπάζιος(3170113)*

## Δημιουργία σχήματος βάσης

Σε αυτό το αρχικό στάδιο, εξετάσαμε τα δεδομένα csv και γράψαμε τις κατάλληλες εντολές "create table" για την δημιουργία του σχήματος της βάσης. Ενδεικτικό παράδειγμα:

```
CREATE TABLE keywords(  
    id int,  
    keywords varchar(8000)  
);
```

Επίσης, σε αυτό το σημείο προστέθηκαν και τα κατάλληλα keys (primary, foreign). Ενδεικτικό παράδειγμα:

```
alter table credits add constraint pk_credits primary key  
(id);  
alter table credits add constraint fk_credits_movie  
foreign key (id) references movies_metadata (id);
```

*Μπορείτε να δείτε αναλυτικά όλες τα queries που τρέξαμε στην βάση μας, συγκεντρωμένα στο αρχείο "partA.sql"*

## Εισαγωγή Δεδομένων

Η εισαγωγή των δεδομένων έγινε με την χρήση του λογισμικού 'pgAdmin'

Import/Export data - table 'credits'

Options Columns

**File Info**

Filename: C:\Users\bazio\Downloads\credits.csv

Format: CSV

Encoding: Select an item...

**Miscellaneous**

OID: No

Header: Yes

Delimiter: ,

Quote: "

Specifies the character that separates columns within each row (line) of the file. The default is a tab character in text format, a comma in CSV format. This must be a single one-byte character. This option is not allowed when using binary format.

Cancel OK

## Διαγραφή διπλοτύπων

Η διαγραφή των διπλοτύπων για κάθε έναν από τους πίνακες έγινε με την βοήθεια του εργαλείου διαγραφής διπλοτύπων του Exel (εκτός του πίνακα credits, που έγινε με query, λόγω προβλημάτων με το Exel)

Αυτόματη αποθήκευση credits

Αρχείο Κεντρική Εισαγωγή Διάταξη σελίδας Τύποι **Δεδομένα** Αναθεώρηση Προβολή Βοήθεια

Λήψη δεδομένων: Από κείμενο/CSV, Από το Web, Από πίνακα/περιοχή

Ερωτήματα και συνδέσεις: Ερωτήματα και συνδέσεις, Ιδιότητες, Επεξεργασία

Ερωτήματα & Συνδέσεις

Τύποι δεδομένων

Α2: [{"cast\_id": 14, "character": "Woody (voice)", "credit\_id": "52fe4284c3a36847f8024f95", "g

	A	B	C	D	E	F	G	H	I	J	K	L
1	cast	crew	id									
2	[{"cast_id": 14, "character": "Woody (voice)", "credit_id": "52fe4284c3a36847f8024f95", "g		862									
3	[{"cast_id": 14, "character": "Woody (voice)", "credit_id": "52fe4284c3a36847f8024f95", "g		8844									
4	[{"cast_id": 14, "character": "Woody (voice)", "credit_id": "52fe4284c3a36847f8024f95", "g		15602									
5	[{"cast_id": 14, "character": "Woody (voice)", "credit_id": "52fe4284c3a36847f8024f95", "g		31357									
6	[{"cast_id": 14, "character": "Woody (voice)", "credit_id": "52fe4284c3a36847f8024f95", "g		11862									
7	[{"cast_id": 14, "character": "Woody (voice)", "credit_id": "52fe4284c3a36847f8024f95", "g		949									
8	[{"cast_id": 14, "character": "Woody (voice)", "credit_id": "52fe4284c3a36847f8024f95", "g		11860									
9	[{"cast_id": 14, "character": "Woody (voice)", "credit_id": "52fe4284c3a36847f8024f95", "g		45325									
10	[{"cast_id": 14, "character": "Woody (voice)", "credit_id": "52fe4284c3a36847f8024f95", "g		9091									
11	[{"cast_id": 14, "character": "Woody (voice)", "credit_id": "52fe4284c3a36847f8024f95", "g		710									
12	[{"cast_id": 14, "character": "Woody (voice)", "credit_id": "52fe4284c3a36847f8024f95", "g		9087									
13	[{"cast_id": 14, "character": "Woody (voice)", "credit_id": "52fe4284c3a36847f8024f95", "g		12110									
14	[{"cast_id": 14, "character": "Woody (voice)", "credit_id": "52fe4284c3a36847f8024f95", "g		21032									
15	[{"cast_id": 14, "character": "Woody (voice)", "credit_id": "52fe4284c3a36847f8024f95", "g		10858									

Κατάργηση διπλοτύπων

Για να διαγράψετε τις διπλότυπες τιμές, επιλέξτε μία ή περισσότερες στήλες που περιέχουν διπλότυπες τιμές.

Επιλογή όλων Κατάργηση επιλογής όλων Τα δεδομένα έχουν κεφαλίδες

Στήλες

☐ cast

☐ crew

☒ id

OK Άκυρο

## Διαγραφή δεδομένων που δεν υπάρχουν στον movies\_metadata

Η διαγραφή δεδομένων που δεν υπάρχουν στον movies\_metadata απο άλλους πίνακες έγινε με την παρακάτω ενδικτική εντολή:

```
delete from credits
where id not in (
    select id from movies_metadata
);
```

*Μπορείτε να δείτε αναλυτικά όλες τα queries που τρέξαμε στην βάση μας, συγκεντρωμένα στο αρχείο "partA.sql"*

## Δημιουργία του πίνακα genres

Αναγκαία κρίθηκε η δημιουργία του πίνακα genres, του οποίου τα δεδομένα εξαγάγαμε από τον πίνακα movies\_metadata. Για την δημιουργία του αρχικά χρησιμοποιήθηκε ένας βοηθητικός πίνακας temp και κατάλληλες εντολές replace όπως ενδεικτικά φαίνονται στο παρακάτω απόσπασμα από το αρχείο partA.sql:

```
create table temp as
select distinct (regexp_split_to_table(genres, '},{'))
from movies_metadata;

update temp
set regexp_split_to_table = replace(regexp_split_to_table,
['{','}');
```

## Δημιουργία του πίνακα movies\_connect\_genres

Η δημιουργία του πίνακα movies\_connect\_genres έγινε με κύριο σκοπό την ένωση του πίνακα genres με τον movies\_metadata. Η ένωση έγινε με την χρήση cross join. Απόσπασμα της υλοποίησης μπορείτε να δείτε παρακάτω.

```
create table movies_connect_genres as
select movies_metadata.id as movie_id, genres.id as
genre_id
from movies_metadata cross join genres
where movies_metadata.genres like '%' || genres.id || '%';
```

*Μπορείτε να δείτε αναλυτικά όλες τα queries που τρέξαμε στην βάση μας, συγκεντρωμένα στο αρχείο "partA.sql"*