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)
);
```

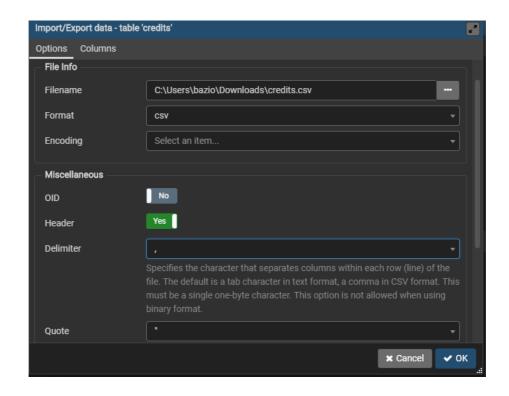
Επίσης, σε αυτό το σημείο προστέθηκαν και τα κατάλληλα keys (primary, foreign). Ενδεικτικό παράδειγμα:

```
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 τρέζαμε στην βάση μας, συγκεντρωμένα στο αρχείο "partA.sql"

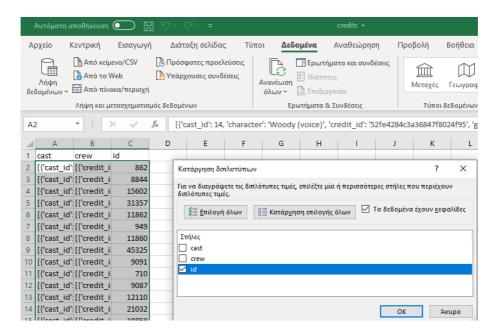
Εισαγωγή Δεδομένων

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



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)



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 εντολη:

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

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