

Project Βάσεις Δεδομένων

Αποστόλου Ιωάννης – 3190013

Καλανζιτής Ηλίας - 3190068

Εξήγηση Μέρους Α:

Για το μέρος Α είχαμε κάποια csv αρχεία που έπρεπε μέσω μιας προεργασίας να τα ανεβάσουμε στην βάση μας. Γι' αυτή την δουλειά χρησιμοποιήσαμε την γλώσσα προγραμματισμού [python](#) και την βιβλιοθήκη [pandas](#). Το script είναι σχεδόν τέρμα αυτοματοποιημένο, το μόνο που χρειάζεται είναι το να αλλάξεις τις μεταβλητές με το όνομα/ path των αρχείων. Αφαιρέσαμε τα διπλότυπα (Εκτός του rating & rating small), και συγκρίναμε σαν βάση μας το “*MoviesMetadata*” με τα άλλα αρχεία, ώστε να βρούμε ταινίες που δεν υπάρχουν στο “*MoviesMetadata*”.

Έπειτα πάλι με την χρήση script (όπως στις σειρές ασκήσεων) σκανάραμε τα αρχεία και δημιουργήσαμε τις εντολές του create table της sql.

Τέλος δημιουργήσαμε ένα sql αρχείο για το ανέβασμα των δεδομένων, ώστε να γίνει αυτοματοποιημένη η διεργασία και για λάθη κατά το ανέβασμα να μπορούμε να το ξανατρέξουμε γρήγορα. Στο τέλος του αρχείου κάνουμε convert το data type ενός column που χρειαζόμασταν σε json καθώς κατά το ανέβασμα δεν μπορούσαμε να το ανεβάσουμε σε json.

Έξτρα Πληροφορίες:

- Το αρχείο με το οποίο μπορείτε να συνδεθείτε στην βάση μέσω της **python** είναι το [connect.py](#), ενώ το αρχείο για την προεργασία ονομάζεται [check.py](#).
- Στο connect έχουμε ήδη τα credentials του εξεταστή, οπότε απλά χρειάζεται να τρέξει το script. Αλλιώς εάν επιθυμείτε να συνδεθείτε ως άλλος χρήστης μπορείτε να μπειτε στο αρχείο και να αλλάξετε τις αντίστοιχες μεταβλητές.
- Οι φωτογραφίες της οπτικοποίησης των queries είναι στον φάκελο photos.
- Το *ERM* είναι στο [erm.png](#).
- Στο [import_files.sql](#) βρίσκονται όλα τα queries για να δημιουργηθεί η βάση.