Graphical user interface

Description automatically generatedΒάσεις Δεδομένων

2η εργασία

Μέλη: Ιωάννα Καλαντζή p3210057

Οδυσσέας Τζαμάκος p3210196

Στο αρχείο code.py πραγματοποιείται η διαδικασία διαχωρισμού του αρχείου keywords.csv στα δύο επιμέρους αρχεία, keyword.csv, και haskeyword.csv. Χρησιμοποιήσαμε regular expressions προκειμένου να αφαιρέσουμε τα non breaking spaces (temp1), τα double quotes (temp2), και τα single quotes (temp3) ώστε να μην έχουμε πρόβλημα αργότερα στο διάβασμα των στοιχείων. Αρχικοποιήσαμε ένα dictionary για το keyword, και μια λίστα για το haskeyword, και ανοίξαμε το αρχείο keywords.csv. Για κάθε γραμμή του αρχείου, αποθηκεύσαμε την πρώτη στήλη στο movieid και την δεύτερη στο keywordslist. Ύστερα εφαρμόσαμε τα παραπάνω regular expressions στο keywordslist, και μετατρέψαμε το αποτέλεσμα σε dictionary χρησιμοποιώντας το json.loads. Έπειτα συμπληρώσαμε τα keyword\_list και haskeyword\_list με τα κατάλληλα στοιχεία. Τέλος, γράψαμε τα αρχεία keyword.csv και haskeyword.csv, μεταφέροντας το dictionary keyword\_list, και τον πίνακα haskeyword\_list σε καθένα από τα αρχεία αντίστοιχα.

Στη συνέχεια, κάναμε import όλα τα αρχεία στον server μας, διορθώνοντας τον τύπο των μεταβλητών όπου ήταν απαραίτητο, και φτιάξαμε το αρχείο alter\_tables.sql με τα πρωτεύοντα και ξένα κλειδιά. Στο αρχείο movie.csv θέσαμε ως πρωτεύον κλειδί το id, στο genre.csv το id, στο productioncompany.csv το id, στο collection.csv το id, στο movie\_cast.csv ως πρωτεύον το id και ως ξένο κλειδί το movie\_id, στο movie\_crew.csv ως πρωτεύον το id και ως ξένο το movie\_id, στο keyword.csv το id. Στο belongstocollection.csv θέσαμε ως ξένο κλειδί το movie\_id, και το collection\_id, στο hasGenre.csv το movie\_id, και το genre\_id, στο hasProductionCompany.csv το pc\_id, και το movie\_id, στο haskeyword.csv το key\_id και το movie\_id, και στο ratings.csv το movie\_id.