Σχεσιακη αλγεβρα για το πρωτο σετ ασκησεων στα Συνθετα Δεδομενα

Λαμβανοντας υποψη ότι στην σχεσιακη αλγεβρα εχουμε μια γραμμη για κάθε στοιχειο ενός πινακα.

**Q1.1)**

Αρχειο suntheta\_1,1.py

**title.crew2 είναι το ταξινομημενο αρχειο από το title.crew (δημιουργηθηκε από το suntheta\_sorting\_files.py ) και τα αλλα αρχεια αντιστοιχα.**

Ανοιγω τα αρχεια title.crew2, title.basics2 προσπελαζω το crew2 και αν υπαρχουν παραπανω από ενας director προχωραω μεσα στο basics2 μεχρι τα 2 αρχεια να εχουν ιδιο tconst και αν εχουν ιδιο γραφω στο αρχειο το primary\_title και τους directors.

Τ1<- (**title.crew)**

Τ2<- (**title.crew)**

**Β1<**- **(title.basics)**

**Ρ1 <**-σ Τ1.tconst = Τ2.tconst  (Τ1 X Τ2)

**Α <**-σ Ρ1.tconst = **Β1**.tconst  ( Β1 Χ Ρ1)

π **primaryTitle, directors** ( Α )

**Q1.2)**

Αρχειο suntheta\_1,2.py

Ανοιγω τα αρχεια title.episode2, title.basics2 προσπελαζω το episode2 και αν βρω το seasonNumber =1 προχωραω μεσα στο basics2 μεχρι τα 2 αρχεια να εχουν ιδιο tconst και αν εχουν ιδιο γραφω στο αρχειο το primaryTitle, parentTconst,seasonNumber**.**

**Ε1<**- **(title.episode)**

**Β1<**- **(title.basics)**

**Ρ1= σ E1.**episodeNumber=1 **(E1)**

**Α<- σ Ε1.**tconst  **=Ρ1.**tconst  **(Ε1ΧΡ1)**

**π primaryTitle, parentTconst, seasonNumber (Α)**

**Q1.3)**

Αρχειο suntheta\_1,3.py

Ανοιγω τα αρχεια title.ratings2, title.basics2 προσπελαζω τα basics2, ratings2 και αν βρω το ιδιο tconst προχωραω τους μετρητες και των 2 αρχειων στην επομενη γραμμη διαφορετικα γραφω στο αρχειο το primaryTitle και προχωραω μονο τον μετρητη του basics2.

**Β1<**- **(title.basics)**

**R1<-(title.ratings)**

**Σ1 <-(Β1 Χ R1)**

**P1 <- σ B1.**tconst = R1.tconst  **(Β1 Χ R1)**

**Α<-Σ1-P1**

**Π Β1.primaryTitle (Α)**

**Q2.1)**

Sorting version, Hashing version στο αρχειο suntheta\_2,1.py

Hashing version:

Ανοιγω τα αρχεια title.ratings2 κρατώντας έναν πινακα 10 θεσεων (hash\_list) και προσπελαζω το αρχειο.Περνω το average\_rating του προσθετω το 0.9 και περνω το ακεραιο μερο και προσαυξανω κατά 1 τη θεση average\_rating-1 του πινακα hash\_list. Στο τελος κανω την καταλληλη εκτυπωση χρησιμοποιωντας τον πινακα hash\_list.

Sorting version

Ταξινομω το αρχειο title.ratings στο title.ratings3 το ανοιγωκαι το προσπελαζω το αρχειο κρατώντας 2 μεταβλητες μια για κατωφλι (katwfli ) για να καταλαβαινω ποτε αλλαζει το average\_rating διαστημα και μια για το πληθος των average\_rating στο διαστημα που βρισκομαστε (katwfli\_sum).

R1<- σ averageRating<=1.0 (**title.ratings)**

R2<- σ averageRating<=2.0 and averageRating >1.0 (**title.ratings)**

R3<- σ averageRating<=3.0 and averageRating >2.0 (**title.ratings)**

R4<- σ averageRating<=4.0 and averageRating >3.0 (**title.ratings)**

R5<- σ averageRating<=5.0 and averageRating >4.0 (**title.ratings)**

R6<- σ averageRating<=6.0 and averageRating >5.0 (**title.ratings)**

R7<- σ averageRating<=7.0 and averageRating >6.0 (**title.ratings)**

R8<- σ averageRating<=8.0 and averageRating >7.0 (**title.ratings)**

R9<- σ averageRating<=9.0 and averageRating >8.0 (**title.ratings)**

R9<- σ averageRating<=10.0 and averageRating >9.0 (**title.ratings)**

**Π COUNT(R1) ( g count(R1) )**

**Π COUNT(R2) ( g count(R2) )**

**Π COUNT(R3) ( g count(R3) )**

**Π COUNT(R4) ( g count(R4) )**

**Π COUNT(R5) ( g count(R5) )**

**Π COUNT(R6) ( g count(R6) )**

**Π COUNT(R7) ( g count(R7) )**

**Π COUNT(R8) ( g count(R8) )**

**Π COUNT(R9) ( g count(R9) )**

**Π COUNT(R10) ( g count(R10) )**

**Q2.2)**

Αρχειο suntheta\_1,3.py

Ανοιγω τα αρχεια title.ratings2, title.basics2 τα προσπελαζω κρατώντας 2 λεξικα ένα για το αθροισμα των average\_rating και ένα για το count των βαθμολογιων για κάθε χρονια ,αν βρω το ιδιο tconst ελενχω στα λεξικα αν εχω ηδη αυτή τη χρονια και αν ναι τα ενημερωνω αναλογα αν όχι το προσθετο στο λεξικο με αναλογες αρεχικες τιμες και μολις βγω από το δευτερο if προχωραω τους μετρητες και των 2 αρχειων στην επομενη γραμμη διαφορετικα προχωραω μονο τον μετρητη του basics2. Στο τελος κανω τη εκτυπωσεις βαση τα 2 λεξικα.

**Β1<**- **(title.basics)**

**R1<-(title.ratings)**

**T1 <- σ Β1.**tconst = **R1.**tconst  **(B1 X R1)**

**A <-** startYear **g avg(**averageRating**) (Τ1)**

**Π** startYear , **avg(**averageRating**) (Α)**