

## Σχεσιακή αλγεβρα για το πρωτο σετ ασκησεων στα Συνθετα Δεδομενα

Λαμβανοντας υποψη ότι στην σχεσιακή 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 προχωρω μεσα στο basics2 μεχρι τα 2 αρχεια να εχουν ιδιο tconst και αν εχουν ιδιο γραφω στο αρχειο το primary\_title και τους directors.

**T1<- (title.crew)**

**T2<- (title.crew)**

**B1<- (title.basics)**

**P1 <- σ T1.tconst = T2.tconst (T1 X T2)**

**A <- σ P1.tconst = B1.tconst ( B1 X P1)**

**Π primaryTitle, directors ( A )**

### Q1.2)

Αρχειο suntheta\_1,2.py

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

**E1<- (title.episode)**

**B1<- (title.basics)**

**P1= σ E1.episodeNumber=1 (E1)**

**A<- σ E1.tconst =P1.tconst (E1XP1)**

**Π primaryTitle, parentTconst, seasonNumber (A)**

### Q1.3)

Αρχείο suntheta\_1,3.py

Ανοίγω τα αρχεία title.ratings2, title.basics2 προσπελάζω τα basics2, ratings2 και αν βρω το ίδιο tconst προχωρώ τους μετρητές και των 2 αρχείων στην επόμενη γραμμή διαφορετικά γράφω στο αρχείο το primaryTitle και προχωρώ μόνο τον μετρητή του basics2.

**B1<- (title.basics)**

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

**Σ1 <-(B1 X R1)**

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

**A<-Σ1-P1**

**Π B1.primaryTitle (A)**

### Q2.1)

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

Hashing version:

Ανοίγω τα αρχεία title.ratings2 κρατώντας έναν πίνακα 10 θέσεων (hash\_list) και προσπελάζω το αρχείο. Περνώ το average\_rating του προσθέτω το 0.9 και περνώ το ακέραιο μέρος και προσυζηνώ κατά 1 τη θέση average\_rating-1 του πίνακα hash\_list. Στο τέλος κάνω την καταλληλή εκτύπωση χρησιμοποιώντας τον πίνακα hash\_list.

Sorting version

Ταξινομώ το αρχείο title.ratings στο title.ratings3 το ανοίγω και το προσπελάζω το αρχείο κρατώντας 2 μεταβλητές μια για κατωφλί (katwfli) για να καταλαβαίνω ποτέ αλλάζει το average\_rating διαστήμα και μια για το πλήθος των average\_rating στο διάστημα που βρισκόμαστε (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)**

Αρχείο suntheta\_1,3.py

Ανοίγω τα αρχεία title.ratings2, title.basics2 τα προσπελαζω κρατώντας 2 λεξικά ένα για το άθροισμα των average\_rating και ένα για το count των βαθμολογιών για κάθε χρονιά, αν βρω το ίδιο tconst ελέγχω στα λεξικά αν έχω ήδη αυτή τη χρονιά και αν ναι τα ενημερώνω αναλόγα αν όχι το προσθέτω στο λεξικό με αναλογές αρεχικές τιμές και μόλις βγω από το δεύτερο if προχωράω τους μετρητές και των 2 αρχείων στην επόμενη γραμμή διαφορετικά προχωράω μόνο τον μετρητή του basics2. Στο τέλος κάνω τη εκτυπώσεις βάση τα 2 λεξικά.

**B1<- (title.basics)**

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

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

**A <- startYear ⋈ avg(averageRating)    (T1)**

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