

## ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ

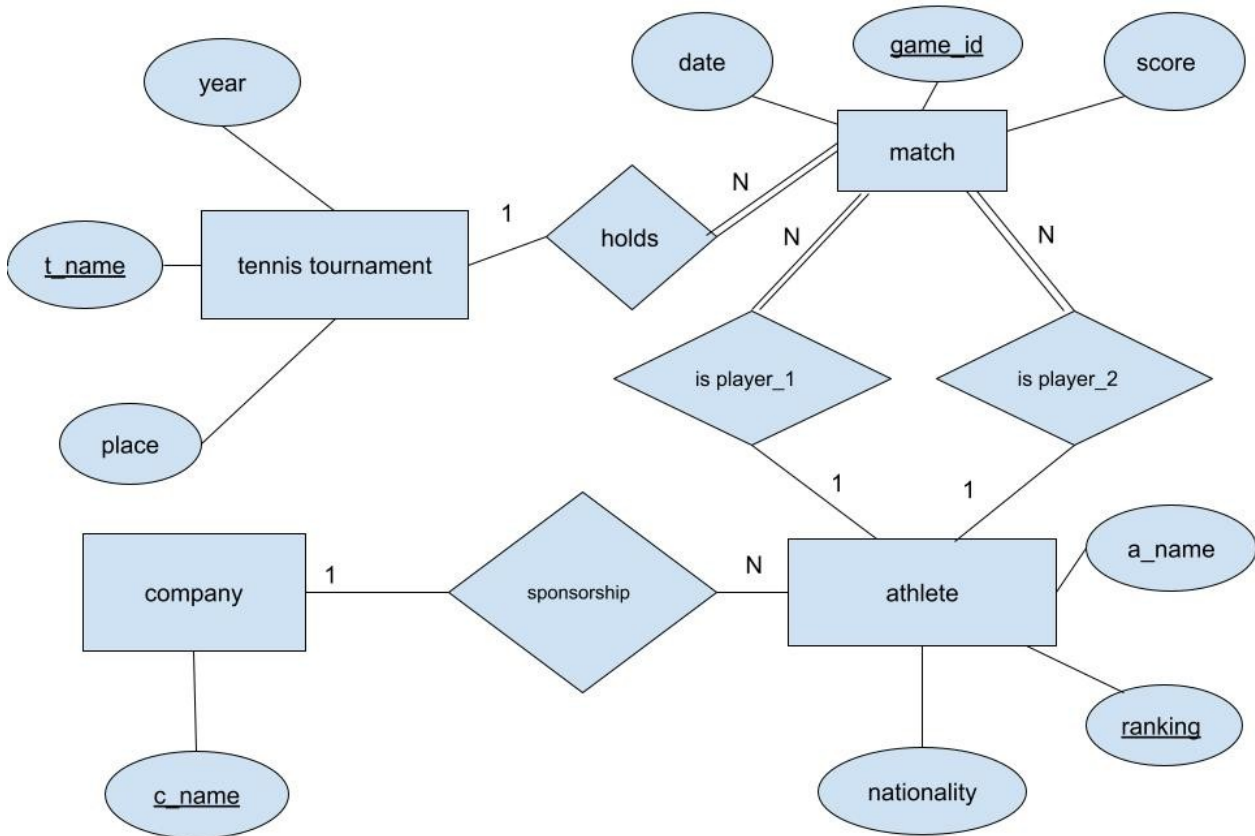
GroupsForExercises116

ΟΝΟΜΑ:	Ρουφογάλη Αρχοντία	Τσαρτσαράκος Κυριάκος	Γκίγκας Αγγελος
Αριθμός Μητρώου:	03118114	03118054	03118218

### 1η ΣΕΙΡΑ ΑΣΚΗΣΕΩΝ

#### ♦ Άσκηση 1

A.



Επιλέξαμε να δημιουργήσουμε σεντ οντοτήτων για τα τουρνουα, τους αγώνες, τους αθλητές και τις εταιρείες, καθώς αυτό συμπεραίνεται από την εκφώνηση. Σαν key attributes επιλέξαμε τα t\_name, game\_id, ranking και c\_name αντίστοιχα, καθώς είναι χαρακτηριστικά και μοναδικά για κάθε οντότητα. Το a\_name δεν θα μπορούσε να γίνει key attribute στο athlete καθώς δύο αθλητές θα μπορούσαν να έχουν ίδιο όνομα. Για αυτό, έχοντας να επιλέξουμε ανάμεσα στα υπόλοιπα χαρακτηριστικά, επιλέξαμε το ranking καθώς είναι το μόνο το οποίο διαφέρει για κάθε αθλητή. Σε κάθε match υποχρεωτικά συμμετέχουν 1 πρώτος παίκτης, 1 δεύτερος

παίκτης και 1 τουρνουα, και για αυτό βάλαμε διπλές γραμμές, ενώ επειδή οι σχέσεις είναι one to many (αφού σε προσθήκη στα παραπάνω ένας αθλητής και ένα τουρνουά μπορούν να δημιουργούν σχέση με πολλά matches) βάλαμε N από τη μεριά του match και 1 από τη μεριά των υπολοίπων. Η σχέση sponsorship είναι one to many, καθώς μία εταιρεία μπορεί να χορηγεί πολλούς αθλητές, αλλά κάθε αθλητής μπορεί να χορηγείται από μόνο μία εταιρεία. Για αυτό από τη μεριά των εταιρειών έχουμε 1 και από τη μεριά των αθλητών N. Επίσης δεν είναι υποχρεωτικό κάθε εταιρεία να χορηγεί κάποιο αθλητή, ούτε κάθε αθλητής να χορηγείται από κάποια εταιρεία, για αυτό σε αυτή τη σχέση δεν υπάρχει κάπου διπλή γραμμή, όπως στην πρώτη.

B. tennis tournament (t\_name, year, place)

match (game\_id, date, score)

athlete (ranking, a\_name, nationality)

company (c\_name)

holds (t\_name, game\_id, year, place, date, score)

is\_player\_1 (ranking, game\_id, a\_name, nationality, date, score)

is\_player\_2 (ranking, game\_id, a\_name, nationality, date, score)

sponsorship (ranking, c\_name, a\_name, nationality)

Τα 4 πρώτα σχήματα αναφέρονται στα σετ οντοτήτων του ER, και έχουν τα υπογραμμισμένα σαν primary keys. Τα 4 τελευταία σχήματα αναφέρονται στις σχέσεις του ER, και έχουν τα υπογραμμισμένα σαν foreign keys. Ουσιαστικά με το relational schema καταγράψαμε τα σετ οντοτήτων με τα χαρακτηριστικά τους, αλλά και τις σχέσεις, με τα χαρακτηριστικά των σετ οντοτήτων που συμμετέχουν σε αυτές.

## ◆ Άσκηση 2

Q1.)  $\pi_{pid}[(Person \bowtie_{cid = cid \wedge scompanyname = GoogleCompany}) \bowtie_{pid = pid \wedge (scompanyname = FacebookCompany \wedge (cid = cid \wedge sharenum > 500))} Shares)]$

Η παραπάνω έκφραση αποτελείται από δύο κομμάτια. Στο πρώτο κομμάτι ουσιαστικά επιλέγονται τα άτομα τα οποία δουλεύουν στην Google και στο δεύτερο κομμάτι επιλέγονται τα άτομα, τα οποία κατέχουν περισσότερες από 500 μετοχές της εταιρείας Facebook. Επομένως, με το κεντρικό join εξακριβώνουμε τα άτομα τα οποία είναι υπάλληλοι της εταιρείας Google και παράλληλα διαθέτουν περισσότερες από 500 μετοχές της εταιρείας Facebook. Τέλος, με το project out του pid των υπαλλήλων, βρίσκουμε τα pid των υπαλλήλων οι οποίοι ικανοποιούν το παραπάνω συμπέρασμα.

Q2.)  $\pi_{pid}(Person \bowtie_{(managerid = pid \wedge cid = cid)} Shares)$

Με το παραπάνω join ουσιαστικά θέτοντας manager id = pid βρίσκουμε τους μάνατζερς οι οποίοι κατέχουν μετοχές και επίσης θέτοντας cid = cid διευκρινίζουμε ότι οι μετοχές που κατέχουν αυτοί οι μάνατζερς πρέπει να είναι από την εταιρεία, στην οποία δουλεύει ο αντίστοιχος υπάλληλος στον οποίο αυτό το άτομο είναι μάνατζερ. Επίσης, με το project out στο pid βρίσκουμε το pid των υπαλλήλων των οποίων οι μάνατζερς κατέχουν μετοχές από τις εταιρείες στις οποίες δουλεύουν.

Q3.)  $\pi_{pid}[(\sigma_{count \geq 3})(pid \ gcount(cid)(Person \bowtie_{pid = pid} Shares))]$

Το παραπάνω ερώτημα υλοποιείται με aggregation, στο οποίο βρίσκουμε όλους τους υπαλλήλους, οι οποίοι έχουν μετοχές από κάποια εταιρία και μέσω της ρουτίνας count, για κάθε υπάλληλο μετράμε το πλήθος των cid από shares που αντιστοιχούν στον κάθε ένα και μέσω του select επιλέγουμε τους υπαλλήλους με πλήθος cid μεγαλύτερο ή ίσο του 3 και από αυτούς του υπαλλήλους που επιλέξαμε κάνουμε project out το pid τους. Έτσι βρίσκονται όσοι υπάλληλοι κατέχουν μετοχές από τουλάχιστον τρεις εταιρείες.

Q4.)[(ppid, cidShares / pcidCompany)⋈pid = pidPerson]

Με την διαίρεση των παραπάνω δύο παραγόντων, ουσιαστικά στον “παρονομαστή” βρίσκουμε όλα τα cid των εταιρειών και με την χρήση της διαίρεσης με τον “αριθμητή” ο οποίος ουσιαστικά είναι οι μετοχές οι οποίες αντιστοιχούν σε κάθε υπάλληλο, βρίσκουμε τα pid που συνδυάζονται με όλα τα cid των εταιρειών και κάνοντας ένα join με την συνθήκη pid = pid βρίσκουμε τους υπαλλήλους, οι οποίοι κατέχουν μετοχές από όλες τις εταιρείες.

### ◆ Άσκηση 3

A)

Q1:

SELECT storeid, sname

FROM store

WHERE (employee\_number <= 100) or (city = “Αθήνα”);

Q2:

SELECT store.sname

FROM

(

SELECT \*

FROM

( SELECT \*

FROM Goods

WHERE gname = “μολύβι”

) as my\_pencil

```
INNER JOIN Supply ON Supply.gid = my_pencil.gid

) as pencil_supply

INNER JOIN Store ON store.storeid = pencil_supply.storeid;
```

**Q3:**

```
SELECT Store.sname, Store.city

FROM

(

SELECT *

FROM Supply as supply_1

WHERE NOT EXISTS (

(SELECT supply_2.gid FROM Supply as supply_2 WHERE supply_2.storeid = "0808")

EXCEPT

(SELECT supply_3.gid FROM Supply as supply_3 WHERE supply_3.storeid = supply_1.storeid)

)

) as division

INNER JOIN Store ON Store.storeid = division.storeid;
```

B)

**Q4:**

```
SELECT Store.sname

FROM

(

SELECT COUNT(gid), storeid

FROM Supply
```

```
GROUP BY storeid

ORDER BY COUNT(gid) DESC

LIMIT 5

) as best_5_stores

INNER JOIN Store ON Store.storeid = best_5_stores.storeid;
```

**Q5:**

```
SELECT Store.city

FROM

(

SELECT *

FROM Supply

INNER JOIN Goods ON Supply.gid = Goods.gid

) as expensive

INNER JOIN Store ON Store.storeid = expensive.storeid

GROUP BY Store.city

HAVING (MAX(expensive.price) > 200);
```

**Q6:**

```
SELECT athens.gid

FROM

(

SELECT Supply.storeid, Supply.gid

FROM Supply, Store

WHERE Supply.storeid = Store.storeid AND Store.city = "Αθήνα"
```

) as athens

WHERE NOT EXISTS (

(SELECT supply\_2.storeid FROM athens as supply\_2)

EXCEPT

(SELECT supply\_3.storeid FROM athens as supply\_3 WHERE supply\_3.gid = athens.gid)

);

**Q7:**

SELECT athens.gid

FROM

(

SELECT Supply.storeid, Supply.gid

FROM Supply, Store

WHERE Supply.storeid = Store.storeid AND Store.city = “Αθήνα”

) as athens

WHERE athens.gid NOT IN

(

SELECT Supply.gid

FROM Supply, Store

WHERE Supply.storeid = Store.storeid AND Store.city = “Πάτρα”

);

#### ♦ Άσκηση 4

Δίνονται, η σχέση  $R(ABCDE)$  με συναρτησιακές εξαρτήσεις:

$F = \{$

$B \rightarrow EA$

$EBC \rightarrow D$

$BED \rightarrow A$

$\}$

A)

Ακολουθούμε την μεθοδολογία και βρίσκουμε:

Set  $A = \{B, C\}$ : τα χαρακτηριστικά που ανήκουν στην LHS αποκλειστικά

Set  $B = \{A\}$ : τα χαρακτηριστικά που ανήκουν στην RHS αποκλειστικά

Set  $C = \{D, E\}$ : τα χαρακτηριστικά που βρίσκονται και στις 2 πλευρές

Βρίσκουμε τις κλειστότητες του set A και έχουμε:  $B^+ = ABE$ ,  $C^+ = C$ ,  $(BC)^+ = ABCED = R$

Επομένως το υποψήφιο κλειδί είναι το BC.

B)

Ζητείται η κανονική και η ελάχιστη κάλυψη. Επομένως:

Παρατηρώντας τις συναρτησιακές εξαρτήσεις, ψάχνουμε για extraneous χαρακτηριστικό. Στην τρίτη σχέση, το D είναι εξωτερικό\* οπότε τελικά έχουμε:

$$F = \{B \rightarrow EA, BE \rightarrow A, EBC \rightarrow D\}.$$

Επαναλαμβάνουμε την παραπάνω διαδικασία για τις νέες σχέσεις. Βλέπουμε πως το A είναι εξωτερικό στη σχέση  $B \rightarrow EA^{**}$  οπότε τελικά έχουμε:

$$F = \{B \rightarrow E, EBC \rightarrow D, BE \rightarrow A\}.$$

Επαναλαμβάνουμε την παραπάνω διαδικασία για τις νέες σχέσεις. Βλέπουμε πως το E είναι εξωτερικό χαρακτηριστικό στη σχέση  $BE \rightarrow A^{***}$  οπότε τελικά έχουμε:

$$F = \{B \rightarrow E, EBC \rightarrow D, B \rightarrow A\}. \text{ Ενώνουμε την πρώτη και τελευταία σχέση και καταλήγουμε σε: } F = \{B \rightarrow EA, EBC \rightarrow D\}.$$

Επαναλαμβάνοντας, βλέπουμε ότι το E είναι εξωτερικό στη σχέση  $EBC \rightarrow D$ . Εν τέλει, έχουμε:

$$F = \{B \rightarrow EA, BC \rightarrow D\}.$$

Στο σημείο αυτό δεν μπορούμε να βρούμε κάποιο νέο εξωτερικό χαρακτηριστικό και κάθε LHS είναι μοναδική. Επομένως, η κανονική κάλυψη είναι  $F = \{B \rightarrow EA, BC \rightarrow D\}$ .

Για να βρεθεί η ελάχιστη κάλυψη, απλώς χωρίζουμε τις σχέσεις που αποτελούνται από πολλαπλά χαρακτηριστικά στο δεξί μέρος.

Η ελάχιστη κάλυψη είναι  $F = \{B \rightarrow E, B \rightarrow A, BC \rightarrow D\}$ .

\*αφού από την  $B \rightarrow EA$  παράγεται το A και από την  $EBC \rightarrow D$  παράγεται το D

\*\*αφού με αντικατάσταση της σχέσης στο F με την  $B \rightarrow E$  παίρνουμε το  $B \rightarrow E$  και  $BE \rightarrow A$

\*\*\*αφού  $B \rightarrow E$  και  $BE \rightarrow A$  (το E παράγεται από το B)

C)

Ζητείται η πιο αυστηρή κανονική μορφή που ικανοποιεί η R. Θεωρούμε R της μορφής 1NF με υποψήφιο κλειδί το BC. Εξετάζουμε εάν ικανοποιείται η κανονική μορφή 2NF. Δηλαδή:



$F = \{B \rightarrow EA, EBC \rightarrow D, BED \rightarrow A\}$ , βλέπουμε στην πρώτη και τελευταία σχέση εξάρτηση από το B αλλά όχι σε συνδυασμό με το C. Όμως το υποψήφιο κλειδί είναι το BC. Επομένως, δεν ικανοποιείται η κανονική μορφή 2NF και άρα η αυστηρότερη κανονική μορφή της R είναι 1NF.

D)

Ζητείται αποσύνθεση σε σύνολο από 3NF σχέσεις. Άρα:

Παρατηρώντας την κανονική μορφή  $F = \{B \rightarrow AE, BC \rightarrow D\}$  βλέπουμε ότι η εξάρτηση  $B \rightarrow AE$  δεν γίνεται να ανήκει στο R, επομένως έχουμε το σχήμα  $R1 = (BEA)$  και  $R2 = (BCD)$ . Για την R1 το κλειδί είναι το B και για την R2 είναι το BC. Εφόσον το R2 περιέχει το υποψήφιο κλειδί και κανέναν από τα R1, R2 δεν περιέχει το ένα το άλλο τότε δεν χρειάζεται να προσθέσουμε κάποιο άλλο σχήμα και το ζητούμενο σύνολο από 3NF σχέσεις είναι το  $\{R1, R2\}$ .

### ♦ Άσκηση 5

Δίνονται, η σχέση R(ABCD) με συναρτησιακές εξαρτήσεις:

$F = \{$

$AB \rightarrow C$

$B \rightarrow D$

$C \rightarrow A$

$\}$

A)

Ακολουθούμε την μεθοδολογία και βρίσκουμε:

Set A = {B}: τα χαρακτηριστικά που ανήκουν στην LHS αποκλειστικά

Set B = {D}: τα χαρακτηριστικά που ανήκουν στην RHS αποκλειστικά

Set  $C = \{A, C\}$ : τα χαρακτηριστικά που βρίσκονται και στις 2 πλευρές

Βρίσκουμε τις κλειστότητες του set  $A$  και έχουμε:  $B^+ =$ . Επομένως, βρίσκουμε τις κλειστότητες των συνόλων των στοιχείων του  $A$  και των στοιχείων του  $R$  που δεν περιέχονται στο  $B$ . Παρατηρούμε ότι η κλειστότητα  $(BA)^+ = BCDA = R$  (δίνει όλα τα χαρακτηριστικά του  $R$ ). Αντίστοιχα, η κλειστότητα  $(BC)^+ = BCDA = R$  (δίνει όλα τα χαρακτηριστικά του  $R$ ). Άρα, τα υποψήφια κλειδιά είναι:  $BC$  και  $BA$ .

B)

Ζητείται αποσύνθεση σε ένα σύνολο από BCNF σχέσεις.

Γνωρίζουμε πως για να είναι μια σχέση σε BCNF μορφή πρέπει να είναι σε 3NF μορφή και για κάθε εξάρτηση  $A \rightarrow B$  το  $A$  πρέπει να είναι super key. Παρατηρούμε ότι η εξάρτηση  $B \rightarrow D$  παραβιάζει τις προϋποθέσεις του BCNF καθώς το  $B$  δεν είναι super key. Επομένως, χωρίζουμε το  $R$  σε δύο σχήματα  $R_1, R_2$  ως εξής:

$R_1 = (ABC)$  με  $F_1 = \{AB \rightarrow C, C \rightarrow A\}$

$R_2 = (BD)$  με  $F_2 = \{B \rightarrow D\}$

Προφανώς το  $R_2$  είναι στη ζητούμενη μορφή. Ωστόσο το  $R_1$  όχι καθώς, η σχέση  $C \rightarrow A$  παραβιάζει τις προϋποθέσεις (το  $C$  δεν είναι super key). Επαναλαμβάνοντας το παραπάνω, χωρίζουμε το  $R_1$  σε δύο σχήματα  $R_3, R_4$  ως εξής:

$R_3 = (BC)$  με  $F_3 = \{0\}$

$R_4 = (AC)$  με  $F_4 = \{C \rightarrow A\}$

Καταλήγουμε όλα τα  $R_2, R_3, R_4$  να είναι σε BCNF μορφή.

(Να σημειωθεί πως η εξάρτηση  $AB \rightarrow C$  μετά την 2η επανάληψη χάνεται).