

Εθνικό Μετσοβίο Πολυτέχνειο

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ

ATOMIKH

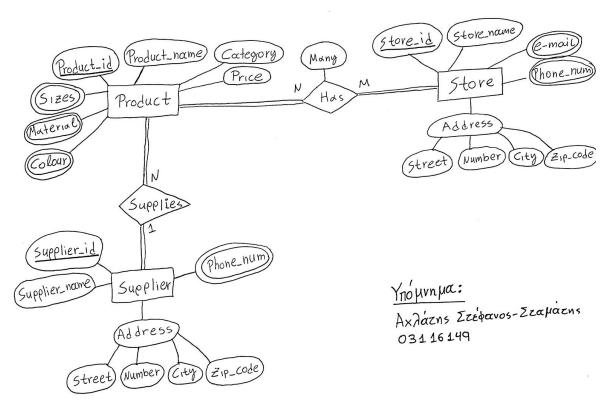
Αχλάτης Στέφανος-Σταμάτης (03116149) < el16149@central.ntua.gr>

Άσκηση 1η

Μια αλυσίδα καταστημάτων εμπορίας υποδημάτων αποφάσισε να δημιουργήσει μία ΒΔ ώστε να μπορεί να παρακολουθεί καλύτερα τα αποθέματα των προϊόντων της. Για να το πετύχει αυτό χρειάζεται να αποθηκεύει πληροφορίες για τα καταστήματα της, το όνομα, την διεύθυνση και τους πιθανούς τρόπους επικοινωνίας, καθώς και τα προϊόντα που είναι διαθέσιμα σε αυτό το κατάστημα. Πρέπει να αποθηκεύει πληροφορίες για τους προμηθευτές, την διεύθυνση τους, το όνομα της εταιρίας, τα προϊόντα τα οποία τους προμηθεύει και ένα ή περισσότερα τηλέφωνα επικοινωνίας. Τα προϊόντα μπορούν να ανήκουν σε δύο κατηγορίες, παπούτσια και τσάντες. Για όλα τα προϊόντα πρέπει να αποθηκεύονται, το όνομα τους, η τιμή τους και τα χρώματα στα οποία είναι διαθέσιμα. Ειδικά για τα παπούτσια, να αποθηκεύονται και τα μεγέθη που είναι διαθέσιμα, ενώ για τις τσάντες το υλικό κατασκευής τους.

Να σχεδιάσετε το Ε-R διάγραμμα που προκύπτει από την παραπάνω περιγραφή, να δικαιολογήσετε σύντομα την επιλογή των κλειδιών, των γνωρισμάτων και των σχέσεων.

Με βάση την προηγούμενη περιγραφή καταλήγουμε στο εξής διάγραμμα Ε-R:



Όπως έχουμε δει και από το σχετικό εργαστήριο ως οντότητες θεωρούμε τα ουσιαστικά που καθαρίζουν το κείμενο της περιγραφής της Βάσης Δεδομένων. Έτσι επιλέγουμε να χρησιμοποιήσουμε τρεις οντοτητες: Product, Store, Supplier. Με μια διαφορετική οπτική της άσκησης θα μπορούσε να δοθεί ένα ΕR στο οποίο υπήρχε και μια τρίτη οντότητα Category, αλλά θα έκανε αρκετά πιο σύνθετο το ΕR χωρίς να προσφέρει ουσιαστικά παραπάνω διευκόλυνση στην κατανόηση και στην μετέπειτα διαχείριση της Βάσης Δεδομένων από την SQL.

Ως σχέσεις επιλέγουμε τα ρήματα που συνδέουν αυτά τα ουσιαστικά. Πιο αναλυτικά στο συγκεκριμένο ΕR έχουμε ότι η σχέση που συνδέει το Product και το Supplier είναι η supplies που δηλώνει ότι ο προμηθευτείς προμηθεύει το προϊόν και η σχέση που συνδέει το Product με το Store είναι η σχέση has που δηλώνει ότι ένα Κατάστημα έχει ένα προϊόν.

Θα μπορούσαμε να δώσουμε και μια έξτρα σχέση μεταξύ του Supplier και του Store, που να δηλώνει ότι ο προμηθευτής συνεργάζεται με το Καταστημα, αλλά επειδή δεν δίνεται σαφω απο την εκφώνηση, ουτε θα προσέδιδε ουσιαστική διευκόλυνση στην συγγραφή της SQL παραλειπεται.

Στην συνέχεια αναλύουμε το ΕR:

Product:

- Product_id: Ένα μοναδικό αναγνωριστικό για κάθε προιόν, συχνά ονομάζεται και Barcode. Με αυτό ξεχωρίζουμε μοναδικά τα προϊόντα.
- Product name: Το όνομα του προϊόντος.
- Category: Η κατηγορία του προιόντος, δηλαδή αν είναι τσάντα ή παπούτσια. Αυτός ο διαχωρισμός θα μπορούσε να γίνει με πολλούς τρόπους από την sql π.χ. με ένα tinyInt ο οποίος θα ήταν π.χ. 0 για τσάντα και 1 για παπούτσια.
- Size: Multivalued attribute το οποίο αρχικά τιθεται ίσο με null και όταν το προϊόν είναι παπούτσια μπορεί να πάρει πολλές ακέραιες τιμές που αντιστοιχούν στα μεγέθη παπουτσιων.
- Material: Multivalued attribute το οποίο αρχικά τιθεται ίσο με null και όταν το προϊόν είναι παπούτσια μπορεί να πάρει πολλές τιμές που αντιστοιχούν στα υλικά που από τα οποία είναι κατασκευασμένα οι τσάντες.
- Colour: Multivalued attribute το οποίο δηλώνει τα χρώματα που διατίθενται για το εν λόγω προϊόν, μπορεί να πάρει πολλές τιμές γιατί μπορεί να υπάρχουν πολλά χρώματα.

Supplies:

- Supplier_id: Ένα μοναδικό αναγνωριστικό για κάθε προμηθευτή, θα μπορούσε να είναι το SSN του προσώπου που εργάζεται ως προμηθευτείς. Με αυτό ξεχωρίζουμε μοναδικά τους προμηθευτές.
- Supplier_name: Το όνομα του προμηθευτεί
- Phone_num: Multivalued attribute το οποίο αποθηκεύει τα τηλέφωνα επικοινωνίας του προμηθευτεί.
 Μπορεί να έχει πολλά τηλέφωνα π.χ. κινητό και σταθερό.
- Address: Composite Attribute το οποίο αποτελείται από τέσσερα γνωρίσματα Street, Number, City, Zip_code που αντιπροσωπεύουν την ακριβή τοποθεσία της αποθήκης του προμηθευτή.

Supplies:

Μέσω της σχέσης Supplies μπορεί να βρεθουν τα προϊόντα που προμηθεύει ο κάθε προμηθευτής

Store:

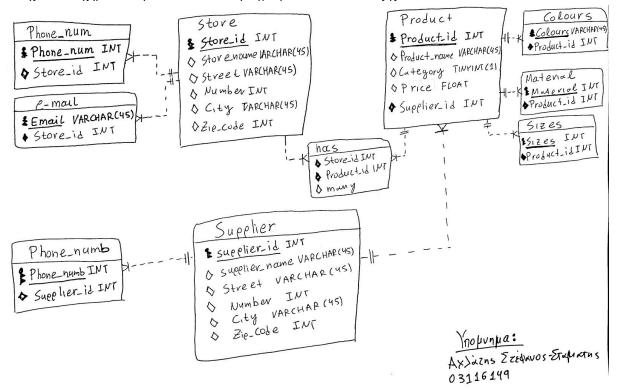
- Store_id: Ένα μοναδικό αναγνωριστικό για κάθε κατάστημα, θα μπορούσε να είναι ένας αυξον ακέραιος. Με αυτό ξεχωρίζουμε μοναδικά τα καταστήματα.
- Store name: Το όνομα του καταστήματος
- E-mail: Multivalued attribute το οποίο δηλώνει το οποίο αποθηκεύει τα email επικοινωνίας του μαγαζιού. Μπορεί να έχει πολλά τηλέφωνα π.χ. ένα για προωθητικές ενέργειες ένα για επικοινωνία με το κοινό κλπ.
- Phone_num: Multivalued attribute το οποίο αποθηκεύει τα τηλέφωνα επικοινωνίας του προμηθευτεί.
 Μπορεί να έχει πολλά τηλέφωνα π.χ. ένα για προωθητικές ενέργειες ένα για επικοινωνία με το κοινό κλπ.
- Address: Composite Attribute το οποίο αποτελείται από τέσσερα γνωρίσματα Street, Number, City, Zip_code που αντιπροσωπεύουν την ακριβή τοποθεσία του καταστήματος.

Has:

Μέσω της σχέσης has βλέπουμε ποια προϊόντα διατίθενται στο εκάστοτε κατάστημα και μέσω του attribute της σχέσης has, δηλαδή μέσω του many, βλέπουμε το πλήθος των διαθέσιμων προϊόντων.

Να σχεδιάσετε το σχεσιακό σχήμα που αντιστοιχεί στο Ε-R διάγραμμα του πρώτου ερωτήματος. Να δικαιολογήσετε σύντομα την επιλογή σας ως προς τα στοιχεία του διαγράμματος που έγιναν πίνακες.

Το σχεσιακό σχήμα που προκύπτει από το προηγούμενο ΕR είναι το εξής:



Ακολουθήθηκε το σχεδιαστικό πρότυπο που δίνει το mysql workbench και έτσι τα primary keys φαίνονται ότι έχουν ένα κλειδί σχηματισμένο από δίπλα τους και είναι υπογραμμισμένα, τα attributes έχουν έναν ρόμβο κενό ενώ τα foreign keys έχουν ένα bold ρομβο (το αντιστοιχο του κοκκινου για workbench).

Αξίζει να σημειωθεί ότι τα multivalued attributes γίνονται πάντα πίνακες και συνδέονται με την οντότητα στην οποία αναφέρονται, έχοντας ως foreign key το primary key του πίνακα στον οποίο αναφέρονται. Τα Composite Attribute σπάνε στα tables των οντοτήτων που ανήκουν στα επιμέρους attributes τους. Τέλος η σχέση has επειδή είναι NxM έγινε πίνακας που συνδέει του πινακες του Store και product, ενώ η σχέση Supplies επειδή είναι Nx1, βλέπουμε ότι απλά το table του Product παίρνει το κλειδί του Supplier ως foreign key. Τα υπόλοιπα είναι τα attributes που υπάρχουν και στο ER.

Προσέξτε ότι ι δυο πίνακες με τηλέφωνο διαχωρίζονται και ο δευτερος που αντιστοιχεί στον πινακα με τα τηλέφωνα του supplier έχει και ένα b.

Βασισμένοι στο σχεσιακό σχήμα του δεύτερου ερωτήματος να απαντήσετε με την χρήση σχεσιακής άλγεβρας τα παρακάτω ερωτήματα:

- 1. Δείξτε το όνομα των καταστημάτων που έχουν τουλάχιστον δύο διαφορετικά σχέδια τσάντας σε μαύρο χρώμα. Μην χρησιμοποιήσετε συναθροιστικές συναρτήσεις.
- 2. Δείξτε την τιμή όλων των παπουτσιών για το κατάστημα με όνομα 'Store 1'.
- 3. Δείξτε τα ονόματα των προμηθευτών που τους προμηθεύουν περισσότερα από 5 προϊόντα, χρησιμοποιώντας συναθροιστικές συναρτήσεις.

```
    iteam_1 <- (π Store_id, Product_id(has))</li>
    π iteam_2 <- (π Store_id, Product_id(has))</li>
    π iteam_1 .Name(iteam_1 ×((iteam_1 .Store_id=iteam_2.Store_id) and (iteam_1.Product_id<>iteam_2.Barcode) and (iteam_1 .Category=0) and (iteam_2.Category=0) and (iteam_1.Color="black") and (iteam_2.Color="black"))iteam_2)
    π Product.Price(σ Store.Name="Store_A" ^ Product.Category=1 ^ Has.Store_id=Store.Store_id ^ Has.Product_id=Product.Product_id)
    π Supplier name (π Supplier id, Supplier name(Supplier)×(Supplier id g(count(Product id)>5)(Supplies)))
```

Βασισμένοι στο σχεσιακό σχήμα του δεύτερου ερωτήματος να απαντήσετε με την χρήση SQL τα παρακάτω ερωτήματα:

- 1. Δείξτε την μέση τιμή όλων των παπουτσιών.
- 2. Δείξτε τα προϊόντα για τα οποία υπάρχουν 2 ή λιγότερες μονάδες και το όνομα του καταστήματος στο οποίο συμβαίνει αυτό.
- 3. Δείξτε την διεύθυνση και τα τηλέφωνα όλων των προμηθευτών.

1. SELECT AVG(Price) FROM Product WHERE Category = 1

2.
SELECT Product_id, Product_name
FROM has, Store
WHERE has.many<=2 AND has.Store_id=Store.Store_id

3. SELECT Supplier.Number, Supplier.Street, Supplier.City, Supplier.Zip_code, Phone_numb.Phone_Numb FROM Supplier, Phone_numb

Δίνεται η σχέση R(A,D,G,I,M,P,Q,V) για την οποία ισχύουν οι ακόλουθες συναρτησιακές εξαρτήσεις. A-> GI

AGP->Q

AI->DMP

DGQ->I

GP->AV

GV->P

I->D

V->IMQ

Χρησιμοποιώντας τους κατάλληλους αλγορίθμους και παρουσιάζοντας τα ενδιάμεσα βήματα να υπολογίσετε:

- 1. Τα υποψήφια κλειδιά.
- 2. Την κανονική κάλυψη (canonical cover) και την ελάχιστη κάλυψη (minimal cover).
- 3. Την καλύτερη (πιο αυστηρή) κανονική μορφή που ικανοποιεί η R.
- 4. Μια αποσύνθεση σε ένα σύνολο από 3NF σχέσεις. Σχολιάστε αν διατηρούν τις εξαρτήσεις.

Ερώτημα 1

Ξεκινάμε την εύρεση closure:

(A)+ = ADGIMPQV που είναι Minimum candidate key

(D)+=D

(G)+=G

(I)+=DI

(M)+=M

(P)+=P

(Q)+=Q

(MQ)+ = MQ(PQ)+ = PQ

(AV)+ = ADGIMPQV

(DV)+=DIMQV

(IV)+=DIMQV

(V)+=DIMQ(AD)+ = ADGIMPQVπου είναι Superkey (AG)+ = ADGIMPQVπου είναι Superkey (DG)+=DG(AI)+ = ADGIMPQVπου είναι Superkey (DI)+=DI(GI)+=DGI(AM)+ = ADGIMPQVπου είναι Superkey (DM)+=DM(GM)+=GM(IM)+=DIM(AP)+ = ADGIMPQVπου είναι Superkey (DP)+=DP(GP)+ = ADGIMPQVπου είναι Composite minimum candidate key (IP)+=DIP(MP)+=MP(AQ)+ = ADGIMPQVπου είναι Superkey (DQ)+=DQ(GQ)+=GQ(IQ)+=DIQ

που είναι Superkey

(GV)+ = ADGIMPQV που είναι Composite minimum candidate key

$$(MV)$$
+ = DIMQV
 (PV) + = DIMPQV
 (QV) + = DIMQV

Επομένως έχουμε τρία minimum candidate key εκ των οποίων τα δύο είναι Composite minimum candidate key.

Παρακάτω φαίνεται και η εφαρμογή του αλγορίθμου του closure για ορισμένες περιπτώσεις

AUNDUCIAN EGAPHOSA ZOU ASSOPICHOU SIA CLOSUVE:

initilize: (A)+ >	A	
Ano FD	Anozeseopa	
A -> GI	(A) [†] → A G I : } 1= leap	
AI > DMP	$(A)^{\dagger} \rightarrow A GIDMP$	
AGP > Q	(A) + > A GIPMPQ > 2= loop	
GP >V	(A) → AGIDMPQV	
	3=1000	,
, a a a a a a a a a a	KAMIA ANATH STU 3= LOOP	

· Tru zpizndopa nou Enzesoupe zou beogro repeat, for Neoni benan vées i Sioznzes no vesult mai o also reparga

Ano FD	Anozalwrz.	
VOIMQ	(V*1 -> VIMQ	} 1= loop
	KAMIA ANNATH	} 2= loop.

D TEPHOTESE ENERN 010 2= 600P der enaver-fin ellager.

Ερώτημα 2:

1η Διαδικασία Αλγορίθμου:

Όποτε έχουμε πάνω από δύο attributes στο δεξί μέλος σπάμε το FD στα επιμέρους με βάση τις γνωστές ιδιότητες. Έτσι έχουμε ότι:

- $A \rightarrow GI$
 - και σπάει στα:
 - $A \rightarrow G$
 - $A \rightarrow I$
- $V \rightarrow IMQ$

και σπάει στα:

- $V \rightarrow I$
- $V \rightarrow M$
- $V \rightarrow O$
- $AI \rightarrow DMP$

και σπάει στα:

- $AI \rightarrow D$
- $AI \rightarrow M$
- $AI \rightarrow P$
- $GP \rightarrow AV$

και σπάει στα:

- $GP \rightarrow A$
- $GP \rightarrow V$

2η Διαδικασία Αλγορίθμου:

Ελέγχουμε για κάθε FD που έχει περισσότερα του ενός attributes στο αριστερό μέλος αν είναι απαραίτητα. Έτσι, έχουμε ότι:

- \bullet AI \rightarrow D
 - Το Α δεν είναι απαραίτητο αφόυ το Ι μπορεί να ορίσει το D.

Το Ι είναι απαραίτητο για να κρατήσει το coverge του D.

- $\bullet \quad AI \to M$
 - Το A είναι απαραίτητο για να κρατήσει το coverage του M.

Το Ι δεν είναι απαραίτητο αφόυ το Α μπορεί να ορίσει το Μ.

- $\bullet \quad AI \to P$
 - Το A είναι απαραίτητο για να κρατήσει το coverage του P.

Το Ι δεν είναι απαραίτητο αφόυ το Α μπορεί να ορίσει το Ρ.

- $GP \rightarrow A$
 - Το G είναι απαραίτητο για να κρατήσει το coverage του A.

Το P είναι απαραίτητο για να κρατήσει το coverage του A.

- $GP \rightarrow V$
 - Το G είναι απαραίτητο για να κρατήσει το coverage του V.

Το P είναι απαραίτητο για να κρατήσει το coverage του V.

- $GP \rightarrow P$
 - Το G είναι απαραίτητο για να κρατήσει το coverage του P.

Το V είναι απαραίτητο για να κρατήσει το coverage του P.

- $\bullet \quad AGP \to Q$
 - Το Ι δεν είναι απαραίτητο αφού τα υπόλοιπα αριστερά attributes μπορούν να ορίσουν το Q.
 - Το G είναι απαραίτητο για να κρατήσει το coverage του Q.

Το P είναι απαραίτητο για να κρατήσει το coverage του Q.

- $DGQ \rightarrow I$
 - Το D είναι απαραίτητο για να κρατήσει το coverage του I.
 - Το G είναι απαραίτητο για να κρατήσει το coverage του I.
 - Το Q είναι απαραίτητο για να κρατήσει το coverage του I.

Άρα η κανονική κάλυψη είναι:

 $A \rightarrow GIMP$

 $\mathbf{I} \to \mathbf{D}$

 $V \rightarrow IMQ$

 $GP \rightarrow AVQ$

 $GV \rightarrow P$

 $DGQ \to I$

3η Διαδικασία Αλγορίθμου (για εύρεση της ελάχιστης):

Τώρα ελαχιστοποιούμε το σύνολο των FD. Για κάθε FD κάνουμε ένα προσωρινό συνολο απο FD χωρις το εν λόγω FD. Το εν λόγω FD είναι απαραίτητο αν στο νεο closure δεν περιλαμβάνει το δεξί μέλος του εν λόγω FD.

To FD A \rightarrow G είναι απαραίτητο

To FD A \rightarrow I den eίναι απαραίτητο

To FD I \rightarrow D είναι απαραίτητο

Το FD V→ Ι είναι απαραίτητο

To FD $V \rightarrow M$ είναι απαραίτητο

To FD $V \rightarrow Q$ είναι απαραίτητο

To FD A \rightarrow M den eínai aparaíthto

To FD A \rightarrow P είναι απαραίτητο

Το FD GP \rightarrow Α είναι απαραίτητο

To FD GP \rightarrow V είναι απαραίτητο

To FD GV \rightarrow P είναι απαραίτητο

To FD GP \rightarrow Q den eínai aparaíthto

To FD DGQ \rightarrow Ι είναι απαραίτητο

Αρα η ελάχιστη κάλυψη είναι:

 $A \rightarrow GP$

 $\boldsymbol{I} \to \boldsymbol{D}$

 $V \rightarrow IMQ$

 $GP \rightarrow AV$

 $GV \rightarrow P$

 $DGQ \rightarrow I$

Ερώτημα 3:

Η R είναι 1NF, καθώς κάθε attribute θεωρείτε ότι περιέχει μόνο μια τιμή για κάθε γραμμή

Η R δεν είναι 2NF. Υπάρχει τουλάχιστον μια partial dependency σε ενα composite minimum key. Για να ικανοποιείται το 2NF, δεν μπορει να υπάρχει κανένα non-prime attribute το οποίο μπορει να προσδιοριστει μέσω FD από ένα σύνολο από composite minimum key.

Η συνθήκη παραβιάζεται από τα εξής non-prime attributes:

- Το ελάχιστο πλήθος attribute που attribute το D είναι fd προσδιορίσιμα από το V, και θα έπρεπε να είναι μόνο fd προσδιορίσιμα από το συνολικό συνολο που συνθέτουν το compiste minimum key {G,V}.
- Το ελάχιστο πλήθος attribute που attribute το I είναι fd προσδιορίσιμα από το V, και θα έπρεπε να είναι μόνο fd προσδιορίσιμα από το συνολικό συνολο που συνθέτουν το compiste minimum key {G,V}.
- Το ελάχιστο πλήθος attribute που attribute το M είναι fd προσδιορίσιμα από το V, και θα έπρεπε να είναι μόνο fd προσδιορίσιμα από το συνολικό συνολο που συνθέτουν το compiste minimum key {G,V}.
- Το ελάχιστο πλήθος attribute που attribute το Q είναι fd προσδιορίσιμα από το V, και θα έπρεπε να είναι μόνο fd προσδιορίσιμα από το συνολικό συνολο που συνθέτουν το compiste minimum key $\{G,V\}$.

Η R δεν είναι 3NF. Δεν είναι σε 2NF και δεν ικανοποιούν όλα τα fd τις εξείς συνθήκες:

- Το δεξί μέλος να είναι να είναι υποσύνολο του αριστερού
- Το αριστερό μέλος να είναι superkey ή minimum key της σχέσης
- Το δεξί μέλος είναι ή είναι μέλος κάποιου minimum key της σχέσης.

Τα FD που αποτύχαν είναι τα:

- \bullet I \rightarrow D
- \bullet $V \rightarrow I$
- \bullet $V \rightarrow M$
- \bullet $V \rightarrow Q$
- $DGQ \rightarrow I$

Η R δεν είναι BCNF. Δεν είναι 3NF και δεν ικανοποιούν όλα τα fd τις εξείς συνθήκες:

- Το δεξί μέλος να είναι να είναι υποσύνολο του αριστερού
- Το αριστερό μέλος να είναι superkey ή minimum key της σχέσης

Τα FD που αποτύχαν είναι τα:

- \bullet I \rightarrow D
- $V \rightarrow I$
- \bullet $V \rightarrow M$
- $\bullet \quad V \to Q$
- $DGQ \rightarrow I$

Σχόλιο: Η R δεν είναι ούτε 4NF αφου δεν είναι BCNF κλπ.

Άρα η απάντηση είναι το 1NF.

Ερώτημα 4:

Για κάθε συναρτησιακή εξάρτηση του canonical cover set (συνενώνοντας τις συναρτησιακές εξαρτήσεις που έχουν το ίδιο attribute στο αριστερό μέλος) της αρχικής R με τα fd της, κατασκευάζουμε ένα relation schema με τα attributes σε αυτο το fd (και στα δύο μέλη).

Ελέγχουμε αν τουλάχιστον ένα κλειδί μπορεί να βρεθεί σε μια τουλάχιστον νεοσχηματισμενοι 3NF σχέση. Αφου το κλειδί Α είναι παρών σε τουλάχιστον μια από τις 3NF σχέσεις, καμία νέα σχέση δεν δημιουργήθηκε. Ελέγχουμε αν καποια σχέση περιλαμβάνει ολα τα attributes που βρεθηκαν σε άλλη σχέση (και διαγράφουμε τα διπλότυπα και τα μικρότερα). Διαγράφουμε τουλάχιστον μια νεα σχεση που ήταν διπλότυπο ή subset σε μια νέα σχέση.

Έχουμε:

$$\begin{split} &R2(I,M,Q,V) \ \text{na έχει FD: } V \rightarrow I; \ V \rightarrow M; \ V \rightarrow Q. \\ &R3(A,G,P,V) \ \text{na έχει FD: } A \rightarrow G; \ A \rightarrow P; \ GP \rightarrow A; \ GP \rightarrow V; \ GV \rightarrow P. \\ &R5(D,G,I,Q) \ \text{na έχει FD: } I \rightarrow D; \ DGQ \rightarrow I. \end{split}$$