

Συναρτησιακές Εξαρτήσεις & Κανονικές Μορφές

ΗΥ - 360 Αρχεία και Βάσεις Δεδομένων
6^ο Φροντιστήριο

18 Νοεμβρίου 2022

Συναρτησιακές Εξαρτήσεις

- ▶ Μια συναρτησιακή εξάρτηση μεταξύ X και Y συμβολίζεται με $X \rightarrow Y$ και ορίζει έναν περιορισμό ως προς τις τιμές των χαρακτηριστικών. Σε αυτή την περίπτωση το σύνολο χαρακτηριστικών Y λέγεται **συναρτησιακά εξαρτώμενο** (*functional dependent*) από το σύνολο X .
- ▶ Αν δύο πλειάδες έχουν τις ίδιες τιμές στα χαρακτηριστικά του συνόλου X , τότε θα έχουν ίδιες τιμές και στα χαρακτηριστικά του συνόλου Y .
- ▶ Σε οποιαδήποτε σχέση R , αν $S_1 \subseteq R$ και $S_2 \subseteq R$, η συναρτησιακή εξάρτηση $S_1 \rightarrow S_2$ καθορίζει ότι για οποιεσδήποτε δύο πλειάδες t_1 και t_2 της R ισχύει

$$\text{Αν } t_1[S_1] = t_2[S_1], \text{ τότε } t_1[S_2] = t_2[S_2]$$

,δηλαδή αν οποιεσδήποτε δυο πλειάδες μιας σχέσης της R συμφωνούν (έχουν την ίδια τιμή) σε κάποια γνωρίσματα $S_1 \subseteq R$ τότε συμφωνούν (έχουν την ίδια τιμή) και σε κάποια γνωρίσματα $S_2 \subseteq R$.

- ▶ Σε ένα σύνολο χαρακτηριστικών μπορεί να υπάρχουν **τετριμμένες** (trivial) συναρτησιακές εξαρτήσεις αλλά και άλλες που μπορούν να συναχθούν.

π.χ.

- 1) Έστω μια αλυσίδα καταστημάτων. Κάθε τμήμα έχει έναν διευθυντή και κάθε διευθυντής δεν μπορεί να ανήκει σε παραπάνω από ένα τμήμα. Ένα τμήμα μπορεί σε πολλές πόλεις.

- ▶ Ισχύει η συναρτησιακή εξάρτηση,

$$SupSSN \rightarrow DeptNo$$

- ▶ Δεν ισχύουν

$$SupSSN \rightarrow City$$

$$DeptNo \rightarrow City$$

#	DeptNo	SupSSN	City	#
	2	1234	N.York	
	3	5432	Paris	
	2	1234	Barcelona	
	3	5432	Paris	
	2	1234	Chicago	

- 2) Έστω ότι ένας διευθυντής διευθύνει μόνο ένα τμήμα σε κάθε πόλη.

- ▶ Ισχύει η συναρτησιακή εξάρτηση,

$$\{City, DeptNo\} \rightarrow SupSSN$$

- ▶ Δεν ισχύουν

$$SupSSN \rightarrow \{City, DeptNo\}$$

$$\{SupSSN, DeptNo\} \rightarrow City$$

#	DeptNo	SupSSN	City	#
	2	1234	N.York	
	3	5432	Paris	
	2	1234	Barcelona	
	3	5432	Paris	
	2	1234	Chicago	

Αξιώματα Armstrong

Είναι **βάσιμα** καθώς δεν δίνουν λανθασμένες εξαρτήσεις και **πλήρη** καθώς δίνουν όλες τις συναγόμενες συναρτησιακές εξαρτήσεις.

- ▶ Ανακλαστικότητα (Reflexivity):

$$\text{Αν } B \subseteq A, \text{ τότε } A \rightarrow B$$

- ▶ Προσαύξηση (Augmentation):

$$\text{Αν } A \rightarrow B, \text{ τότε } AC \rightarrow BC$$

- ▶ Μεταβατικότητα (Transitivity):

$$\text{Αν } A \rightarrow B \text{ και } B \rightarrow C, \text{ τότε } A \rightarrow C$$

Συνέπειες αξιωμάτων

- ▶ Ένωση (Union):

$$\text{Αν } A \rightarrow B \text{ και } A \rightarrow C, \text{ τότε } A \rightarrow BC$$

- ▶ Διάσπαση (Decomposition):

$$\text{Αν } A \rightarrow CB, \text{ τότε } A \rightarrow B \text{ και } A \rightarrow C$$

- ▶ Ψευδομεταβατικός κανόνας (Pseudo-transitivity):

$$\text{Αν } A \rightarrow B \text{ και } BC \rightarrow D, \text{ τότε } AC \rightarrow D$$

π.χ.

1) Έστω $R = (A, B, C, G, H, I)$ και οι συναρτησιακές εξαρτήσεις

$$A \rightarrow B, \quad A \rightarrow C, \quad CG \rightarrow H, \quad CG \rightarrow I, \quad B \rightarrow H$$

Μέσω των κανόνων Armstrong εξάγονται οι,

$$A \rightarrow H, \quad A \rightarrow BC, \quad AG \rightarrow I, \quad CG \rightarrow HI$$

□

2) Δίνεται πίνακας R με τα χαρακτηριστικά W, U, V, X, Y, Z και οι συναρτησιακές εξαρτήσεις:

$$W \rightarrow UV, \quad U \rightarrow Y, \quad VX \rightarrow YZ$$

Ζητείται να αποδειχθεί ότι ισχύει $WX \rightarrow Z$.

Λύση

- ▶ Με διάσπαση από τη $W \rightarrow UV$ προκύπτει $W \rightarrow V$.
- ▶ Με επαύξηση προκύπτει $WX \rightarrow VX$.
- ▶ Με μεταβατικότητα προκύπτει $WX \rightarrow YZ$.
- ▶ Με διάσπαση προκύπτει $WX \rightarrow Z$.

□

Ορισμός

Το K είναι **υποψήφιο κλειδί** της σχέσης R αν $K \rightarrow R$ για κανένα $\alpha \subset K$, $\alpha \rightarrow R$.

π.χ. Έστω σχεσιακό σχήμα με $R = (A, B, C, D, E)$ με ΣΕ:
 $\{A \rightarrow BC, CD \rightarrow E, B \rightarrow D, E \rightarrow A\}$ βρίσκουμε ότι

$$CD \rightarrow ABCDE, A \rightarrow ABCDE, E \rightarrow ABCDE$$

Τα υποψήφια κλειδιά για την σχέση R είναι

$$A, E, CD$$

Δηλαδή,

- ♦ Υποψήφιο κλειδί K : υπερκλειδί με την ιδιότητα ότι αν αφαιρεθεί ένα οποιοδήποτε γνώρισμα A από το K , το K' που προκύπτει δεν είναι υπερκλειδί.
- ▶ Κάθε σχέση έχει τουλάχιστον ένα υπερκλειδί.
- ▶ Από τον ορισμό, κάθε(σχήμα) σχέσης έχει τουλάχιστον ένα (πρωτεύον) κλειδί – δεν υπάρχουν ασθενείς σχέσεις.

Κλειστότητα

Ορισμός

Το σύνολο των συναρτησιακών εξαρτήσεων που μπορούν να παραχθούν από ένα σύνολο εξαρτήσεων F λέγεται **κλειστότητα** (closure) και συμβολίζεται με F^+ .

π.χ. Έστω $R = \{A, B, C, D\}$ και ισχύει

$$A \rightarrow B, \quad A \rightarrow C \quad B \rightarrow D$$

Αν έχουμε μια συναρτησιακή εξάρτηση όπου $A \rightarrow D$ όπου D είναι ένα υποσύνολο του A τότε η συγκεκριμένη συναρτησιακή εξάρτηση είναι τετριμμένη

$$A \rightarrow D, \quad A \rightarrow A(\text{trivial})$$

□

◇ Ο προσδιορισμός του συνόλου κλειστότητας F^+ είναι δαπανηρός αλγοριθμικά ακόμη και για μικρό σύνολο εξαρτήσεων F .

Ορισμός

Κλειστότητα του συνόλου χαρακτηριστικών X (attribute closure), ως προς το σύνολο των χαρακτηριστικών εξαρτήσεων F , είναι το σύνολο των χαρακτηριστικών που είναι συναρτησιακά εξαρτώμενα από το σύνολο των χαρακτηριστικών X και το συμβολίζουμε με X^+ .

π.χ. Στη σχέση $R = (A\Phi M, AM, \text{βαθμός}, \text{όνομα}, \text{διεύθυνση})$ με (ΣE)

$A\Phi M, AM \rightarrow \text{βαθμός}, \quad A\Phi M \rightarrow \text{όνομα}, \text{διεύθυνση}$

Παίρνουμε ότι $\{A\Phi M\}^+ = A\Phi M, \text{όνομα}, \text{διεύθυνση}$. □

Σημείωση

Αν το X^+ είναι το σύνολο όλων των γνωρισμάτων του πίνακα τότε το X είναι υποψήφιο κλειδί.

Υπολογισμός κλειστότητας A^+

Αλγόριθμος υπολογισμού A^+ της κλειστότητας του A κάτω από το σύνολο F

```
αποτέλεσμα :=  $A$ 
while (υπάρχουν αλλαγές στο αποτέλεσμα) do
  for each  $\beta \rightarrow \gamma$  στο  $F$  do
    begin
      if  $\beta \subseteq$  αποτέλεσμα then
        αποτέλεσμα := αποτέλεσμα  $\cup$   $\gamma$ 
    end
```

Παραδείγματα

1) Έστω $R = \{A, B, C, G, H, I\}$ και οι συναρτησιακές εξαρτήσεις του F :

$$I \rightarrow B, \quad A \rightarrow H, \quad B \rightarrow G, \quad C \rightarrow G, \quad CG \rightarrow I$$

Να υπολογιστεί το $\{A, C\}^+$.

1. $X = AC$
2. $X = ACH$ από $A \rightarrow H$.
3. $X = ACGH$ από $C \rightarrow G$.
4. $X = ACGHI$ από $CG \rightarrow I$.
5. $X = ABCGHI$ από $I \rightarrow B$.

2) Έστω $R = \{A, B, C, D, E, G\}$ και οι συναρτησιακές εξαρτήσεις του F :

$$A \rightarrow B, \quad A \rightarrow C, \quad BC \rightarrow D, \quad D \rightarrow E, \quad B \rightarrow G$$

Να υπολογιστεί την A^+

1. $X = A$
2. $X = AB$ από $A \rightarrow B$.
3. $X = ABC$ από $A \rightarrow C$.
4. $X = ABCD$ από $BC \rightarrow D$.
5. $X = ABCDE$ από $D \rightarrow E$.
6. $X = ABCDEG$ από $B \rightarrow G$.

3) Δίνεται ο πίνακας $R = \{V, Y, Z, W\}$ και το

$$F^+ = \{V \rightarrow Z, VZ \rightarrow W, W \rightarrow Y, VY \rightarrow W\}$$

Να βρεθεί η κλειστότητα του χαρακτηριστικού V .

1. $X = V$
2. $X = VZ$ από $V \rightarrow Z$.
3. $X = VZW$ από $VZ \rightarrow W$.
4. $X = VXWY$ από $W \rightarrow Y$.

Ελάχιστη Κάλυψη

Έστω δύο σύνολα (ΣΕ) F, G τότε,

1. Είναι ισοδύναμα αν κάθε (ΣΕ) στο F είναι δυνατόν να παραχθεί από το G και κάθε (ΣΕ) στο G είναι δυνατόν να παραχθεί από το F , δηλαδή τα F και G είναι ισοδύναμα αν ισχύει: $F^+ = G^+$.
2. Το F καλύπτει το G αν κάθε (ΣΕ) στο G είναι δυνατόν να παραχθεί από το F , δηλαδή ισχύει: $G^+ \subseteq F^+$.
3. Ένα σύνολο F από (ΣΕ) είναι **ελάχιστη κάλυψη** αν ικανοποιεί τα παρακάτω:
 - 3.1 Κάθε εξάρτηση στο F είναι της μορφής: $X \rightarrow A$, όπου A είναι ένα απλό γνώρισμα.
 - 3.2 Δεν μπορούμε να αποσύρουμε μια (ΣΕ) από το F και να εξακολουθούμε να έχουμε ένα σύνολο ισοδύναμο του F .
 - 3.3 Δεν μπορούμε να αντικαταστήσουμε μια (ΣΕ) $X \rightarrow A$ από το F με μια (ΣΕ) $Y \rightarrow A$, όπου $Y \subset X$ και να εξακολουθούμε να έχουμε ένα σύνολο ισοδύναμο του F .

Αλγόριθμος υπολογισμού της ελάχιστης κάλυψης ενός συνόλου ΣΕ

- I (**Διάσπαση**) Δημιουργούμε ένα νέο σύνολο G ισοδύναμο του F όπου φροντίζουμε να έχουμε (ΣΕ) με μόνο ένα γνώρισμα στο δεξιό μέλος της συνάρτησης.
- I Αντικαθιστούμε τις (ΣΕ) με άλλες, που έχουν λιγότερα γνωρίσματα στο αριστερό μέλος εφόσον δεν επηρεάζουν την κλειστότητα του G
- I (**Πλεονάζουσες**) Αφαιρούμε από το G όλες τις (ΣΕ) που δεν επηρεάζουν στην κλειστότητα του G αν αφαιρεθούν.
- I Συγχωνεύουμε τις (ΣΕ) που έχουν το ίδιο αριστερό μέλος, φροντίζοντας έτσι όλα τα αριστερά μέλη να είναι μοναδικά.

Παράδειγμα

Έστω $R = \{A, B, C\}$ και οι συναρτησιακές εξαρτήσεις του F :

$$A \rightarrow BC, \quad B \rightarrow C, \quad A \rightarrow B, \quad AB \rightarrow C$$

Να υπολογιστεί η ελάχιστη κάλυψη του F :

Έχουμε $G = \{A \rightarrow BC, B \rightarrow C, A \rightarrow B, AB \rightarrow C\}$.

1. Η $A \rightarrow BC$ έχει πάνω από ένα γνωρίσματα στο δεξί μέλος, οπότε τη διασπάμε σε $A \rightarrow B$ και $A \rightarrow C$. Άρα,
 $G = \{A \rightarrow B, A \rightarrow C, B \rightarrow C, A \rightarrow B, AB \rightarrow C\}$.
2. Η $AB \rightarrow C$ έχει πάνω από ένα γνωρίσματα στο αριστερό μέλος. Θα κοιτάξω τις κλειστός των A και B ως προς το $(G - (AB \rightarrow C))$.
 - 2.1 $A^+ = ABC$, στην οποία περιέχεται το B . Άρα μπορούμε να το βγάλουμε από την (ΣΕ).
 - 2.2 $G = \{A \rightarrow B, A \rightarrow C, B \rightarrow C, A \rightarrow B, A \rightarrow C\}$.
3.
 - 3.1 Οι δύο τελευταίες είναι επαναλαμβανόμενες, οπότε θα αφαιρεθούν.
 - 3.2 Λόγω μεταβατικότητας ($A \rightarrow B, B \rightarrow C$) η $A \rightarrow C$ καλύπτεται και θα αφαιρεθεί.
 - 3.3 $G = \{A \rightarrow B, B \rightarrow C\}$.
4. Δεν υπάρχουν (ΣΕ) με ίδιο αριστερό μέλος για να συγχωνεύσουμε, επομένως δεν μεταβάλλεται το G .

Συνεπώς, έχουμε την ελάχιστη κάλυψη $G = \{A \rightarrow B, B \rightarrow C\}$.

Υποψήφια κλειδιά

Εύρεση υποψήφινων κλειδιών δεδομένου ενός πίνακα R και των εξαρτήσεων F μεταξύ των πεδίων του.

Βήμα 1ο: Βρες τα χαρακτηριστικά που δεν εμφανίζονται ούτε στην αριστερή αλλά ούτε και στην δεξιά πλευρά της συναρτησιακής εξάρτησης.

Βήμα 2ο: Βρες τα χαρακτηριστικά που εμφανίζονται μόνο στη δεξιά πλευρά.

Βήμα 3ο: Βρες τα χαρακτηριστικά που εμφανίζονται μόνο στην αριστερή πλευρά.

Βήμα 4ο: Συνδύασε τα χαρακτηριστικά του βήματος 1 & 3.

Βήμα 5ο: Βρες την κλειστότητα των χαρακτηριστικών του βήματος 4. Αν σε αυτήν ανήκουν όλα τα υπόλοιπα χαρακτηριστικά τότε είναι υποψήφιο κλειδί. Διαφορετικά, επανέλαβε τα βήματα 4 & 5 προσαρτώντας κάθε φορά ένα χαρακτηριστικό που δεν συμπεριλαμβάνεται στο βήμα 2 & 4.

Παράδειγμα

Έστω $R = \{A, B, C, D, E, F, G\}$ μια σχέση και $F = \{AB \rightarrow F, AD \rightarrow E, F \rightarrow G\}$ ένα σύνολο εξαρτήσεων στη σχέση αυτή. Βρείτε ένα υποψήφιο κλειδί στην R .

Βήμα 1ο: C .

Βήμα 2ο: EG .

Βήμα 3ο: ABD .

Βήμα 4ο: $ABCD$.

Η κλειστότητα του συνόλου χαρακτηριστικών του βήματος 4 είναι το $ABCDFEG$. Άρα το $ABCD$ είναι υποψήφιο κλειδί.

Lossless-Join Decomposition

Ορισμός

Μια αποσύνθεση $\{R_1, R_2, \dots, R_k\}$ της σχέσης R με συναρτησιακές εξαρτήσεις F λέγεται αποσύνθεση **χωρίς απώλεια πληροφορίας** (Lossless - join decomposition), αν ανεξάρτητα από το περιεχόμενο της R , οι συναρτησιακές εξαρτήσεις εξασφαλίζουν ότι

$$R = R_1 \bowtie R_2 \bowtie \dots \bowtie R_k$$

Θεώρημα

Δεδομένης μιας σχέσης R και ενός συνόλου (ΣΕ) F οι οποίες πληρούνται στην R μια αποσύνθεση της R στις σχέσεις R_1 και R_2 δεν πάσχει από απώλεια πληροφορίας αν ισχύει τουλάχιστον μία από τις δύο σχεσιακές εξαρτήσεις:

$$Head(R_1) \cap Head(R_2) \rightarrow Head(R_1)$$

$$Head(R_1) \cap Head(R_2) \rightarrow Head(R_2)$$

Παράδειγμα

π.χ. Έστω $R = \{\text{Τίτλος}, \text{Έτος}, \text{Διάρκεια}, \text{Είδος}, \text{Όνομα Ηθοποιού}, \text{Διεύθυνση}, \text{Έτος Γέννησης}\}$.

Έχουμε τις (ΣΕ)

$\{\text{Τίτλος} \text{ Έτος} \rightarrow \text{Διάρκεια}, \text{Τίτλος} \text{ Έτος} \rightarrow \text{Είδος}\},$

$\{\text{Όνομα Ηθοποιού} \rightarrow \text{Διεύθυνση}, \text{Όνομα Ηθοποιού} \rightarrow \text{Έτος Γέννησης}\}$

και τις σχέσεις

$R_1 = \{\text{Τίτλος}, \text{Έτος}, \text{Διάρκεια}, \text{Είδος}\}$

$R_2 = \{\text{Τίτλος}, \text{Έτος}, \text{Όνομα Ηθοποιού}, \text{Διεύθυνση}, \text{Έτος Γέννησης}\}$

Άρα $R_1 \cap R_2 = \{\text{Τίτλος}, \text{Έτος}\}$. Γνωρίζουμε ότι ένα κλειδί K είναι υπερκλειδί μιας σχέσης R αν και μόνο αν ισχύει $K \rightarrow R$. Συνεπώς η $R_1 \cap R_2$ είναι υπερκλειδί για την R_1 , άρα και αποσύνθεση χωρίς απώλειες.

Κανονικές Μορφές

Ορισμός

Κανονικοποίηση είναι μια διαδικασία μετατροπής δεδομένων που έχουν παραπάνω από μια πιθανές αναπαραστάσεις σε μια 'τυπική' ή 'κανονική' μορφή.

Γιατί κανονικοποιούμε? Θέλουμε να μειώσουμε τον πλεονασμό των δεδομένων (data redundancy) καθώς προκύπτουν ανωμαλίες,

- ▶ Εγγραφής (*Insertion anomalies*).
- ▶ Ενημέρωσης (*Update anomalies*).
- ▶ Διαγραφής (*Deletion anomalies*).

Προσπαθούμε δηλαδή να μετασχηματίσουμε ένα σχεσιακό σχήμα σε μια νέα μορφή, η οποία είναι απαλλαγμένη από ανωμαλίες εισαγωγής, διαγραφής και τροποποίησης εγγράφων.

Τι είναι ο πλεονασμός δεδομένων;

	<i>att₁</i>	<i>att₂</i>	...	<i>att_m</i>
<i>row₁</i>				x
<i>row₂</i>				x
⋮				⋮
<i>row_n</i>				x

Προκαλεί,

- ▶ Αύξηση στο μέγεθος της βάσης.
- ▶ Προβλήματα εγγραφής, ενημέρωσης και διαγραφής.

π.χ.

<i>no</i>	<i>name</i>	<i>course</i>	<i>teacher</i>	<i>office</i>
1	Z	CS – 360	Dr.H	666
2	Y	CS – 360	Dr.H	666
3	X	CS – 360	Dr.H	666
4	W	CS – 360	Dr.H	666

Ανωμαλία Εγγραφής

<i>no</i>	<i>name</i>	<i>course</i>	<i>teacher</i>	<i>office</i>
1	Z	CS – 360	Dr.H	666
2	Y	CS – 360	Dr.H	666
3	X	CS – 360	Dr.H	666
4	W	CS – 360	Dr.H	666
5	V	CS – 360	Dr.H	666

Θα πρέπει δηλαδή να επαναλάβουμε τα ίδια δεδομένα για κάθε καινούργια εγγραφή.

Ανωμαλία Ενημέρωσης

<i>no</i>	<i>name</i>	<i>course</i>	<i>teacher</i>	<i>office</i>
1	Z	CS – 360	Dr.N	666
2	Y	CS – 360	Dr.N	666
3	X	CS – 360	Dr.N	666
4	W	CS – 360	Dr.N	666

Για μια αλλαγή, στον διδάσκοντα, θα πρέπει να αλλάξουμε **όλες** τις τιμές του χαρακτηριστικού *teacher*. Επίσης, οποιαδήποτε αστοχία στην ενημέρωση οδηγεί σε προβλήματα συνέπειας.

Ανωμαλία Διαγραφής

<i>no</i>	<i>name</i>	<i>course</i>	<i>teacher</i>	<i>office</i>
1	Z	CS – 360	Dr.H	666
2	Y	CS – 360	Dr.H	666
3	X	CS – 360	Dr.H	666
4	W	CS – 360	Dr.H	666

Αν επιθυμούμε να διαγράψουμε τους φοιτητές από τη βάση, τότε θα διαγράψουμε και όλες τις πληροφορίες που αφορούν το μάθημα. Άρα έχουμε απώλεια πληροφορίας όταν διαγράψουμε άλλη πληροφορία.

π.χ. Έχουμε τις παρακάτω καταγραφές

<u>Y_ID</u>	<u>E_ID</u>	ΥΕπώνυμο	ΕΌνομα	ΕΤοπος	Ωρες
AH 123456	123	Νικολάου	Γέφυρα	Πάτρα	34
AH 234567	56	Παπαδόπουλος	Γήπεδο	Πειραιάς	28
AZ 345678	123	Νικολάου	Γέφυρα	Πάτρα	5
AK 123456	34	Κανάκη	Π. Κέντρο	Ιωάννινα	76
AH 123458	56	Κουκος	Γήπεδο	Πειραιάς	56

Παρατηρούμε ότι όταν διαγραφεί ένα έργο θα έχει ως αποτέλεσμα να διαγραφούν οι εργαζόμενοι σε αυτό (ανωμαλία διαγραφής).

Ορισμός

Μια συναρτησιακή εξάρτηση (ΣΕ) $X \rightarrow Y$ καλείται **πλήρης** (ή ολική, *full*) όταν $\forall A \in X : X - \{A\} \not\rightarrow Y$, αλλιώς καλείται **μερική** (*partial*).

π.χ.

- 1) Στην σχέση $\{\text{Υπάλληλος_ID}, \text{Έργο_ID}\} \rightarrow \text{Ώρες}$ αν αφαιρέσουμε το Έργο_ID η (ΣΕ) $\{\text{Υπάλληλος_ID} \rightarrow \text{Ώρες}\}$ δεν ισχύει. Επίσης αν αφαιρέσουμε το γνώρισμα $\{\text{Υπάλληλος_ID}\}$.
- 2) Στην σχέση $\{\text{Υπάλληλος_ID}, \text{Έργο_ID}\} \rightarrow \text{Υπάλληλος_Επώνυμο}$ αν αφαιρέσουμε το Έργο_ID η (ΣΕ) $\{\text{Υπάλληλος_ID}\} \rightarrow \text{Υπάλληλος_Επώνυμο}$ εξακολουθεί να ισχύει.

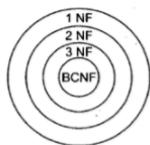
Ορισμός

Κανονική μορφή (*Normal form*) καλείται μία κλάση σχέσεων απαλλαγμένων από συγκεκριμένα προβλήματα τροποποιήσεων.

Πρώτη κανονική μορφή (1NF), Δεύτερη κανονική μορφή (2NF), Τρίτη κανονική μορφή (3NF), Boyce-Codd κανονική μορφή (BCNF)

1η Κανονική μορφή-(1NF)

Είναι κρίσιμη για το σχεσιακό μοντέλο, ενώ οι υπόλοιπες είναι προαιρετικές. Αν ένας πίνακας βρίσκεται σε μια κανονική μορφή τότε βρίσκεται και σε όλες τις προηγούμενες.



Hierarchy of normal forms

Η 2NF, η 3NF και η κανονική μορφή Boyce-Codd (BCNF) βασίζονται στις συναρτησιακές εξαρτήσεις μεταξύ των κλειδιών (πρωτεύοντα ή υποψήφια).
Συγκεκριμένα,

- ▶ Η 1NF θεωρείται τμήμα του τυπικού ορισμού μιάς σχέσης και ορίζει ότι τα χαρακτηριστικά περιλαμβάνουν μόνο **ατομικές** τιμές.
- ▶ Κάθε εγγραφή πρέπει να είναι μοναδική. Αυτό το διασφαλίζει το πρωτεύον κλειδί.
- ▶ Δεν υπάρχουν συνθετικά χαρακτηριστικά.
- ▶ Δεν υπάρχουν χαρακτηριστικά πολλαπλών τιμών.

Παράδειγμα

Δίνετε ο πίνακας

ID	Επώνυμο	Όνομα	Διεύθυνση	Hobby
AH 123456	Νικολάου	Κώστας	Αλαμάνας 20, 75400	Cart, Bowling
AH 234567	Παπαδόπουλος	Γιώργος	Ελ. Βενιζέλου 36Α, 13410	Sailing
AZ 345678	Νικολάου	Μιχάλης	Λ. Δημοκρατίας 12, 71409	BasketBall, Music
AK 123456	Κανάκη	Χαρούλα	Σουλίου 36, 35410	Trekking, Canyoning

Παρατηρούμε ότι δεν είναι σε 1NF.

Κάθε πλειάδα με πλειότιμα γνωρίσματα θα διασπαστεί σε περισσότερες (1 πλειάδα για κάθε διαφορετική τιμή του γνωρίσματος). Τα σύνθετα γνωρίσματα θα διασπαστούν σε περισσότερα, απλά γνωρίσματα.

ID	Επώνυμο	Όνομα	Δ_Οδός	Δ_Αριθμός	Δ_TK	Hobby
AH 123456	Νικολάου	Κώστας	Αλαμάνας	20	75400	Cart
AH 123456	Νικολάου	Κώστας	Αλαμάνας	20	75400	Bowling
AH 234567	Παπαδόπουλος	Γιώργος	Ελ. Βενιζέλου	36A	13410	Sailing
AZ 345678	Νικολάου	Μιχάλης	Λ. Δημοκρατίας	12	71409	BasketBall
AZ 345678	Νικολάου	Μιχάλης	Λ. Δημοκρατίας	12	71409	Music
AK 123456	Κανάκη	Χαρούλα	Σουλίου	36	35410	Trekking
AK 123456	Κανάκη	Χαρούλα	Σουλίου	36	35410	Canyoning

Ο παραγώμενος πίνακας είναι 1NF.

? Όμως έχουμε επανάληψη πληροφορίας...

2η Κανονική Μορφή-(2NF)

- ▶ 1NF.
- ▶ Απομακρύνουμε όλες τις μερικές συναρτησιακές εξαρτήσεις
- ▶ Δημιουργούμε τόσους πίνακες όσες είναι και οι πλήρεις συναρτησιακές εξαρτήσεις (ffd) των πεδίων του πίνακα, και τα τοποθετούμε σ'αυτούς.

Συνεπώς,

- Να είναι 1NF.
- Κάθε γνώρισμα που **δεν** είναι μέρος ενός υποψήφιου κλειδιού εξαρτάται συναρτησιακά από **ολόκληρο** το κλειδί και όχι από ένα μέρος του.
, οι εξαρτήσεις της μορφής

Πρωτεύον γνώρισμα (μη-κλειδί) → Μη πρωτεύον

παραβιάζουν το 2NF.

Παράδειγμα

Έχουμε τον πίνακα

<u>Y_ID</u>	<u>E_ID</u>	ΥΕπώνυμο	ΕΌνομα	ΕΤόπος	Ώρες
AH 123456	123	Νικολάου	Γέφυρα	Πάτρα	34
AH 234567	56	Παπαδόπουλος	Γήπεδο	Πειραιάς	28
AZ 345678	123	Νικολάου	Γέφυρα	Πάτρα	5
AK 123456	34	Κανάκη	Π. Κέντρο	Ιωάννινα	76
AH 123458	56	Κούκος	Γήπεδο	Πειραιάς	56

και τις πλήρεις συναρτησιακές εξαρτήσεις

$Y_ID, E_ID \rightarrow \text{Ώρες}$, $Y_ID \rightarrow \text{ΥΕπώνυμο}$, $E_ID \rightarrow \text{ΕΌνομα, ΕΤόπος}$

Παρατηρούμε ότι δεν αποτελεί 2NF καθώς οι σχέσεις $Y_ID \rightarrow \text{ΥΕπώνυμο}$, $E_ID \rightarrow \text{ΕΌνομα, ΕΤόπος}$ είναι της μορφής

Πρωτεύον γνώρισμα (μη-κλειδί) \rightarrow Μη πρωτεύον

Οι τρεις πλήρεις συναρτησιακές εξαρτήσεις οδηγούν στη δημιουργία τριών πινάκων.

<u>Υ ID</u>	ΥΕπώνυμο
AH 123456	Νικολάου
AH 234567	Παπαδόπουλος
AZ 345678	Νικολάου
AK 123456	Κανάκη
AH 123458	Κούκος

<u>Υ ID</u>	<u>E ID</u>	Ωρες
AH 123456	123	34
AH 234567	56	28
AZ 345678	123	5
AK 123456	34	76
AH 123458	56	56

<u>E ID</u>	ΕΌνομα	ΕΤόπος
123	Γέφυρα	Πάτρα
56	Γήπεδο	Πειραιάς
34	Π. Κέντρο	Ιωάννινα



3η Κανονική Μορφή-3NF

Ορισμός

Μία συναρτησιακή εξάρτηση $X \rightarrow Y$ λέγεται μεταβατική (transitive) αν υπάρχει σύνολο χαρακτηριστικών Z που είναι εκτός κλειδιού της R και ισχύουν $X \rightarrow Z$ και $Z \rightarrow Y$.

Το σύνολο των περιορισμών για την 3NF είναι,

- i) Να είναι 2NF.
- ii) **Δεν** υπάρχουν μεταβατικές εξαρτήσεις.

Αν $A \rightarrow B$, τότε δεν υπάρχει $B \rightarrow C$, όπου B, C μη πρωτεύοντα

Αλλιώς, Η $A \rightarrow B$ **δεν** είναι μεταβατική αν **δεν** υπάρχει σύνολο γνωρισμάτων C τ.ω. $A \rightarrow C$ **και** $C \rightarrow B$.

- iii) **Δεν** υπάρχουν εξαρτήσεις της μορφής,

Μη-πρωτεύον \rightarrow Μη-πρωτεύον

Παράδειγμα

Έχουμε τον πίνακα

ΑριθμόςΠαραγγελίας	ΚωδικόςΠελάτη	ΠόληΠελάτη
124/2010	C142	Ηράκλειο
138/2010	C139	Θεσσαλονίκη
221/2010	C127	Πάτρα
253/2010	C139	Θεσσαλονίκη
391/2010	C142	Ηράκλειο

με τις συναρτησιακές εξαρτήσεις

ΑριθμόςΠαραγγελίας \rightarrow ΚωδικόςΠελάτη, ΚωδικόςΠελάτη \rightarrow Πόλη Πελάτη

και υπάρχει η μεταβατικότητα

ΑριθμόςΠαραγγελίας \rightarrow ΠόληΠελάτη

Η μετατροπή σε 3NF οδηγεί στην διάσπαση

ΚωδικόςΠελάτη	ΠόληΠελάτη
C142	Ηράκλειο
C139	Θεσσαλονίκη
C127	Πάτρα

,

ΑριθμόςΠαραγγελίας	ΚωδικόςΠελάτη
124/2010	C142
138/2010	C139
221/2010	C127
253/2010	C139
391/2010	C142



Κανονική Μορφή Boyce-Codd (BCNF)

Για όλες τις (ΣΕ) στο F^+ της μορφής $X \rightarrow Y$ ισχύει τουλάχιστον ένα από τα παρακάτω:

- i) $X \rightarrow Y$ είναι μια τετριμμένη (ΣΕ) ή
- ii) X είναι υπερκλειδί (δηλαδή υποψήφιο κλειδί ή υπερσύνολο υποψήφιου κλειδιού) του σχήματος R

Δεν πρέπει ένα χαρακτηριστικό να έχει συναρτησιακή εξάρτηση από ένα άλλο χαρακτηριστικό που δεν είναι (ή περιέχει) το κλειδί. Επομένως το **αριστερό μέλος** κάθε μη τετριμμένης (ΣΕ) πρέπει να περιέχει ένα κλειδί, δηλαδή το X να είναι υπερκλειδί του σχήματος R .

π.χ. Έστω $R = \{A, B, C, D, E, F\}$ με κλειδί AB . Για να είναι BCNF επιτρέπονται μόνο οι εξαρτήσεις: $AB \rightarrow AB$, $AB \rightarrow A$, $AB \rightarrow B$, $ABC \rightarrow D$, $ABCD \rightarrow F$, ...
Απεναντίας δεν επιτρέπονται εξαρτήσεις, όπως: $A \rightarrow CD$, $DE \rightarrow F$, $BCD \rightarrow EF$, ...

□

Αλγόριθμος Διάσπασης σε BCNF

- ▶ Για κάθε σχήμα σχέσης R
 - ▶ Βρες μια μη τετριμμένη συνάρτησιακή εξάρτηση $X \rightarrow Y$, με $X \cap Y = \emptyset$, που παραβιάζει τον BCNF ορισμό.
 - ▶ Διάσπαση του σχήματος R σε δύο νέα, έτσι ώστε:
 - ▶ R_1 να έχει γνωρίσματα τα $X \cup Y$.
 - ▶ R_2 να έχει γνωρίσματα τα $R \setminus Y$.
- ▶ Τερματίζει όταν δεν υπάρχουν σχήματα σχέσεων που παραβιάζουν τον BCNF ορισμό.

Παραδείγματα

1) Έχουμε το σχήμα σχέσης που περιγράφεται από τον πίνακα

IDΈργου	<u>IDΣυμβουλος</u>	<u>IDΜηχανικος</u>	Αντικείμενο
E12	Σ12	M13	A111
E12	Σ34	M13	A111
E12	Σ65	M45	A111
E28	Σ28	M67	A678
E28	Σ71	M67	A678
E67	Σ43	M35	A904

με κλειδί: {IDΣύμβουλος, IDΜηχανικός} και συναρτησιακές εξαρτήσεις:

{IDΣύμβουλος, IDΜηχανικός} → IDΈργου, IDΈργου → Αντικείμενο

Παρατηρούμε ότι δεν είναι BCNF, μάλιστα δεν είναι καν 3NF καθώς η εξάρτηση

$ID'Εργου \rightarrow Αντικείμενο$

παραβιάζει τον ορισμό. Το διασπάμε σε δύο σχέσεις
{IDΣύμβουλος, IDΜηχανικός, ID'Εργου} και
{ID'Εργου, Αντικείμενο}

ID'Εργου	IDΣύμβουλος	IDΜηχανικός
E12	Σ12	M13
E12	Σ34	M13
E12	Σ65	M45
E28	Σ28	M67
E28	Σ71	M67
E67	Σ43	M35

ID'Εργου	Αντικείμενο
E12	A111
E28	A678
E67	A904



2)

Έχουμε το σχήμα σχέσης που περιγράφεται από τον πίνακα

Πελάτης	Συναλλαγή	ΠοσόΣ	ΥποΚατ	ΌνομαΥΚ	ΑποθεμαΥΚ
Π12	Σ1312	4060	ΥΚ13	Καβάλα	450
Π14	Σ1434	230	ΥΚ14	Καλαμάτα	720
Π78	Σ4565	4600	ΥΚ45	Ξάνθη	316
Ε26	Σ6728	2670	ΥΚ67	Ρέθυμνο	412
Ε67	Σ1371	13784	ΥΚ13	Καβάλα	450

με κλειδί: {Πελάτης, Συναλλαγή} και συναρτησιακές εξαρτήσεις

ΥποΚατ \rightarrow {ΌνομαΥΚ, ΑπόθεμαΥΚ}, Συναλλαγή \rightarrow {ΥποΚατ, ΠοσόΣ}

Παρατηρούμε ότι δεν είναι BCNF καθώς και οι δύο εξαρτήσεις παραβιάζουν τον BCNF ορισμό. Μετά από διάσπαση η εξάρτηση ΥποΚατ \rightarrow {ΌνομαΥΚ, ΑπόθεμαΥΚ} οδηγεί στην διάσπαση του R σε R_1 και R_2 ,

R_1

ΥποΚατ	ΌνομαΥΚ	ΑποθεμαΥΚ
ΥΚ13	Καβάλα	450
ΥΚ14	Καλαμάτα	720
ΥΚ45	Ξάνθη	316
ΥΚ67	Ρέθυμνο	412

R_2

Πελάτης	Συναλλαγή	ΠοσόΣ	ΥποΚατ
Π12	Σ1312	4060	ΥΚ13
Π14	Σ1434	230	ΥΚ14
Π78	Σ4565	4600	ΥΚ45
Ε26	Σ6728	2670	ΥΚ67
Ε67	Σ1371	13784	ΥΚ13

Όμως η εξάρτηση Συναλλαγή \rightarrow {ΥποΚατ, ΠοσόΣ} παραβιάζει τον BCNF ορισμό στην R_2 . Επομένως διασπάμε το R_2 , την εξάρτηση Συναλλαγή \rightarrow {ΥποΚατ, ΠοσόΣ} σε R_3 και R_4 και καταλήγουμε στις

R_1

ΥποΚατ	ΌνομαΥΚ	ΑποθεμαΥΚ
ΥΚ13	Καβάλα	450
ΥΚ14	Καλαμάτα	720
ΥΚ45	Ξάνθη	316
ΥΚ67	Ρέθυμνο	412

R_3

Συναλλαγή	ΠοσόΣ	ΥποΚατ
Σ1312	4060	ΥΚ13
Σ1434	230	ΥΚ14
Σ4565	4600	ΥΚ45
Σ6728	2670	ΥΚ67

R_4

Πελάτης	Συναλλαγή
Π12	Σ1312
Π14	Σ1434
Π78	Σ4565
Ε26	Σ6728
Ε67	Σ1371



Σύνοψη

Μια σχέση σε 3NF διασπάται σε ένα σύνολο από σχέσεις της BCNF όταν αφαιρεθούν από τον πίνακα όσα χαρακτηριστικά δεν εξαρτώνται πλήρως από κλειδιά.

Απλά

1NF: Όχι σύνθετα και πλειότιμα γνωρίσματα.

2NF: Μόνο πλήρεις συναρτησιακές εξαρτήσεις.

3NF: Όχι μεταβατικές συναρτησιακές εξαρτήσεις.

BCNF: Μόνο περιορισμοί κλειδιού.

Παρατηρήσεις

- ▶ Οδηγούμενοι σε πιο περιοριστική κανονική μορφή, αναγκαζόμαστε να διασπάσουμε την αρχική σχέση σε περισσότερες.
- ▶ Οι περισσότερες 3NF είναι και BCNF.
- ▶ Οι 1NF και 2NF δεν είναι ικανοποιητικές για να χαρακτηρίσουμε την βάση ως καλά σχεδιασμένη.
- ▶ Οι 3NF και BCNF εξαλείφουν προβλήματα που σχετίζονται με ενημερώσεις.
- ▶ Η BCNF ίσως να μην είναι πάντοτε εφικτή, αλλά η 3NF υπάρχει **πάντοτε**.

Τέλος