

Κανονικές Μορφές

Κανονικές Μορφές: Εισαγωγή

- Στόχος: Δοσμένου ενός σχήματος, αν είναι «καλό» ή χρειάζεται περαιτέρω διάσπαση.
Πως; Κανονικές μορφές.
- Ξέρουμε ότι αν ένα σχήμα είναι σε κάποια Κανονική Μορφή δεν υπάρχουν συγκεκριμένα προβλήματα
- Με φθίνουσα σειρά (από την πιο περιοριστική στη λιγότερο περιοριστική)
BCNF 3NF 2NF 1NF
- Βασίζεται σε Σ.Ε., οι Σ.Ε. έχουν σχέση με την επανάληψη πληροφορίας

Κανονικές Μορφές: Εισαγωγή

Πλεονασμός (επανάληψη πληροφορίας)

Ταινία

Τίτλος	Έτος	Διάρκεια	Είδος	Όνομα-Ηθοποιού
--------	------	----------	-------	----------------

Τι συμβαίνει με το (πρωτεύον) κλειδί και τις συναρτησιακές εξαρτήσεις;

Boyce-Codd Κανονική Μορφή

Ένα σχεσιακό σχήμα R είναι σε **Κανονική Μορφή Boyce-Codd (BCNF)** σε σχέση με ένα σύνολο F συναρτησιακών εξαρτήσεων αν για όλες τις ΣΕ στο F* της μορφής $X \rightarrow Y$ ισχύει τουλάχιστον ένα από τα παρακάτω:

- $X \rightarrow Y$ είναι μια τετριμμένη ΣΕ ή
- X είναι **υπερκλειδί** του σχήματος R

Δηλαδή το αριστερό μέρος κάθε μη τετριμμένης ΣΕ πρέπει να περιέχει ένα κλειδί

Το σχήμα μιας ΒΔ είναι σε BCNF αν το σχήμα **κάθε** σχέσης της είναι σε BCNF.

Boyce-Codd Κανονική Μορφή

Παράδειγμα 1

Ταινία (Τίτλος, Έτος, Διάρκεια, Είδος, Όνομα-Ηθοποιού)

Η σχέση Ταινία δεν είναι σε BCNF

(υποψήφιο) κλειδί: {Τίτλος, Έτος, Όνομα-Ηθοποιού}

Για παράδειγμα η ΣΕ Τίτλος Έτος \rightarrow Διάρκεια

Boyce-Codd Κανονική Μορφή

Παράδειγμα 2

Ταινία2 (Τίτλος, Έτος, Διάρκεια, Είδος)

Η σχέση Ταινία2 είναι σε BCNF

Παράδειγμα 3

Οποιαδήποτε σχέση με δύο γνωρίσματα είναι σε BCNF



Αλγόριθμος Αποσύνθεσης σε BCNF

- Βρες μια μη τετριμμένη ΣΕ που παραβιάζει τον BCNF ορισμό, έστω $X \rightarrow Y$ και $X \cap Y = \emptyset$
- Αποσύνθεση του αρχικού σχήματος R σε δύο σχήματα
 - R_1 με γνωρίσματα $X \cup Y$
 - R_2 με γνωρίσματα $R - Y$

Ευριστικός: στα δεξιά όσο το δυνατόν περισσότερα γνωρίσματα Αποσύνθεση χωρίς απώλειες:



Παράδειγμα 1

Ταινία (Τίτλος, Έτος, Διάρκεια, Είδος, Όνομα-Ηθοποιού)

Τίτλος Έτος \rightarrow Διάρκεια Είδος

Ταινία1(Τίτλος, Έτος, Διάρκεια, Είδος)

Ταινία2(Τίτλος, Έτος, Όνομα-Ηθοποιού)



Παράδειγμα 2

Ταινία-Εταιρεία (Τίτλος, Έτος, Διάρκεια, Είδος, Εταιρεία-Παραγωγής, Διεύθυνση-Εταιρείας) -- (Τίτλος, Έτος) (υποψήφιο) κλειδί

Πρόβλημα: υπάρχει μια **μεταβλητική** εξάρτηση

Τίτλος Έτος \rightarrow Εταιρεία-Παραγωγής

Εταιρεία-Παραγωγής \rightarrow Διεύθυνση-Εταιρείας

Τίτλος Έτος \rightarrow Διεύθυνση-Εταιρείας

Ταινία-Εταιρεία1 (Εταιρεία-Παραγωγής, Διεύθυνση-Εταιρείας)

Ταινία-Εταιρεία2 (Τίτλος, Έτος, Διάρκεια, Είδος, Εταιρεία-Παραγωγής)



- Μπορεί να χρειαστεί παραπάνω από μία αποσύνθεση

Αποσύνθεση του αρχικού σχήματος R σε δύο σχήματα - R_1 με γνωρίσματα $X \cup Y$ και R_2 με γνωρίσματα $R - Y$

η R_2 μπορεί να μην είναι σε BCNF



Δεν είναι πάντα δυνατή η αποσύνθεση σε μια BCNF που να διατηρεί τις εξαρτήσεις

Παράδειγμα

Έστω η σχέση Παιζει(Έργο, Κινηματογράφος, Πόλη) με τους περιορισμούς ότι

- (i) δεν υπάρχουν κινηματογράφοι με το ίδιο όνομα
- (ii) κάθε κινηματογράφος έχει πολλές αίθουσες (παίζει πολλά έργα) αλλά κάθε έργο παίζεται μόνο σε ένα κινηματογράφο σε κάθε πόλη

Κινηματογράφος \rightarrow Πόλη

Κλειδιά:

Έργο Πόλη \rightarrow Κινηματογράφος

{Έργο, Πόλη}
{Κινηματογράφος, Έργο}



Παιζει(Έργο, Κινηματογράφος, Πόλη)

Κινηματογράφος \rightarrow Πόλη

Κλειδιά

Έργο Πόλη \rightarrow Κινηματογράφος

{Έργο, Πόλη} {Κινηματογράφος, Έργο}

Αποσύνθεση σε: R_1 {Κινηματογράφος, Πόλη} και R_2 {Κινηματογράφος, Έργο}

Κινηματογράφος	Πόλη	Κινηματογράφος	Έργο
Odeon-ABANA	Αθήνα	Odeon-ABANA	Volver
Village Center Μαρούσι	Αθήνα	Village Center Μαρούσι	Volver

Δε μπορώ κοιτάζοντας μόνο την R2 να δω ότι η εισαγωγή της δεύτερης πλειάδας παραβιάζει μια ΣΕ



Ένα σχεσιακό σχήμα R είναι σε **τρίτη κανονική μορφή (3NF)** σε σχέση με ένα σύνολο F συναρτησιακών εξαρτήσεων αν για όλες τις ΣΕ στο F της μορφής $X \rightarrow Y$ ισχύει τουλάχιστον ένα από τα παρακάτω:

- $X \rightarrow Y$ είναι μια τετριμμένη ΣΕ ή
- X είναι υπερκλειδί του σχήματος R
- κάθε γνώρισμα A του Y - X περιέχεται σε κάποιο υποψήφιο κλειδί

BCNF πιο περιοριστική -- αν σε BCNF \Rightarrow 3NF

Πρωτεύον γνώρισμα (prime attribute): Γνώρισμα που ανήκει σε κάποιο υποψήφιο κλειδί



Παράδειγμα

Παίζει(Έργο, Κινηματογράφος, Πόλη)

Έργο Πόλη \rightarrow Κινηματογράφος

Κινηματογράφος \rightarrow Πόλη

Κλειδιά {Έργο, Πόλη}
{Κινηματογράφος, Έργο}

Υπάρχει μια μεταβατική εξάρτηση
Αλλά απαιτούμε να είναι σε πρωτεύον γνώρισμα

Η σχέση είναι σε 3NF



Αλγόριθμος Αποσύνθεσης σε 3NF

- Υπολόγισε το ελάχιστο κάλυμμα F_c του F
- Για κάθε α.μ. X μιας συναρτησιακής εξάρτησης του F_c έστω Y το σύνολο όλων των γνωρισμάτων A_i που εμφανίζονται στο δ.μ. μιας ΣΕ του F_c $X \rightarrow A_i$
νέα σχέση με γνωρίσματα $X \cup Y$
- Αν κανένα από τα σχήματα που δημιουργούνται δεν περιέχει κλειδί, δημιουργήσε ένα σχήμα που να περιέχει τα γνωρίσματα που σχηματίζουν κλειδί



Αλγόριθμος Αποσύνθεσης σε 3NF

- Απώλειες στη συνένωση;
- Διατήρηση εξαρτήσεων;



Παράδειγμα

Τραπεζίτης(Όνομα-Υποκαταστήματος, Όνομα-Πελάτη, Όνομα-Τραπεζίτη, Αριθμός Γραφείου)

Όνομα-Τραπεζίτη \rightarrow Όνομα-Υποκαταστήματος Αριθμός-Γραφείου

Όνομα-Πελάτη Όνομα-Υποκαταστήματος \rightarrow Όνομα-Τραπεζίτη

Κλειδιά {Όνομα-Πελάτη, Όνομα-Υποκαταστήματος}

3NF:

Τραπεζίτης1(Όνομα-Τραπεζίτη, Όνομα-Υποκαταστήματος Αριθμός-Γραφείου)

Τραπεζίτης2(Όνομα-Πελάτη, Όνομα-Υποκαταστήματος, Όνομα-Τραπεζίτη)

BCNF:



Κανονική Μορφή Boyce-Codd

Ένα σχεσιακό σχήμα R είναι σε BCNF σε σχέση με ένα σύνολο F συναρτησιακών εξαρτήσεων αν για όλες τις ΣΕ στο F της μορφής $X \rightarrow Y$ ισχύει τουλάχιστον ένα από τα παρακάτω:

- $X \rightarrow Y$ είναι μια τετριμμένη ΣΕ ή
- X είναι υπερκλειδί του σχήματος R

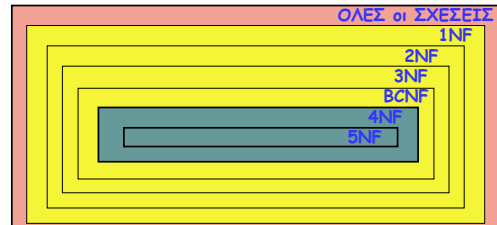
Τρίτη Κανονική Μορφή

-- κάθε γνώρισμα A του Y - X περιέχεται σε κάποιο υποψήφιο κλειδί (είναι πρωτεύον γνώρισμα)

Σχεδιασμός Σχεσιακών Σχημάτων (Επανάληψη)

	BCNF	3NF
• Αποφυγή επανάληψης πληροφορίας	ναι	όχι πάντα
• Αποσύνθεση χωρίς απώλειες στη συνένωση	ναι	ναι
• Διατήρηση εξαρτήσεων	όχι πάντα	ναι

Κανονικές Μορφές



Πρώτη Κανονική Μορφή

1NF (ιστορικοί λόγοι, κάθε γνώρισμα παίρνει ατομικές τιμές)

Δεύτερη Κανονική Μορφή

$X \rightarrow Y$

Y **πλήρης εξάρτηση** από το X αν δεν υπάρχουν περιττά γνώρισμα στο X (στο α.μ της εξάρτησης) (αν υπάρχουν, **μερική** εξάρτηση)

2NF κάθε μη πρωτεύον γνώρισμα (γνώρισμα που δεν ανήκει στο υποψήφιο κλειδί) είναι **πλήρως** εξαρτώμενο από το πρωτεύον κλειδί

Δηλαδή, αφορά κλειδιά με παραπάνω από ένα γνωρίσματα,

Όχι $X \rightarrow Y$, όπου X γνήσιο υποσύνολο του πρωτεύοντος κλειδιού και Y μη πρωτεύον γνώρισμα

Πλειότιμες Εξαρτήσεις

Υπάρχει επανάληψη πληροφορίας που δεν μπορεί να εκφραστεί με απλές ΣΕ

Πλειότιμες Εξαρτήσεις

Πλειότιμες Εξαρτήσεις

Προκύπτουν όταν δυο γνωρίσματα είναι ανεξάρτητα το ένα από το άλλο

Παράδειγμα

Ηθοποιοί(Όνομα, Οδός, Πόλη, Τίτλος, Έτος)

Υποθέτουμε ότι για κάθε ηθοποιό είναι πιθανόν να υπάρχουν πολλές διευθύνσεις

Κανένα από τα 5 γνωρίσματα δεν εξαρτάται συναρτησιακά από τα άλλα τέσσερα \Rightarrow δεν υπάρχουν μη μη τετριμμένες εξαρτήσεις \Rightarrow κλειδί ?

π.χ., Όνομα Οδός Τίτλος Έτος \rightarrow Πόλη δεν ισχύει

Παράδειγμα (συνέχεια)

Ηθοποιός(Όνομα, Οδός, Πόλη, Τίτλος, Έτος)

Όλες οι εξαρτήσεις είναι τετριμμένες

Το σχήμα είναι σε BCNF αλλά υπάρχει επανάληψη πληροφοριών που δεν οφείλεται όμως σε συναρτησιακές εξαρτήσεις

Παράδειγμα

Ηθοποιός(Όνομα, Οδός, Πόλη, Τίτλος, Έτος)

Όνομα →→ Οδός Πόλη

Όνομα	Οδός	Πόλη	Τίτλος	Έτος
C. Fisher	123 Mapple Str	Hollywood	Star Wars	1977
C. Fisher	5 Locust Ln	Malibu	Empire Strikes Back	1980
?				
?				

$X \rightarrow Y$

Για κάθε ζεύγος πλειάδων t_1 και t_2 της σχέσης R που συμφωνούν σε όλα τα γνωρίσματα του X μπορούμε να βρούμε στο R δυο πλειάδες t_3 και t_4 τέτοιες ώστε

• και οι δυο συμφωνούν με τις t_1 και t_2 στο X:

$$t_1[X] = t_2[X] = t_3[X] = t_4[X]$$

• η t_3 συμφωνεί με την t_1 στο Y: $t_3[Y] = t_1[Y]$

• η t_3 συμφωνεί με την t_2 στο R - X - Y: $t_3[R - X - Y] = t_2[R - X - Y]$

• η t_4 συμφωνεί με την t_2 στο Y: $t_4[Y] = t_2[Y]$

• η t_4 συμφωνεί με την t_1 στο R - X - Y: $t_4[R - X - Y] = t_1[R - X - Y]$

$A_1 A_2 \dots A_n \rightarrow B_1 B_2 \dots B_m$

Όνομα			Πόλη Οδός			Τίτλος Έτος					
X			Y			R - X - Y					
A_1	A_2	\dots	A_n	B_1	B_2	\dots	B_m	C_1	C_2	\dots	C_k
a_1	a_2	\dots	a_n	b_1	b_2	\dots	b_m	c_1	c_2	\dots	c_k
a_1	a_2	\dots	a_n	b'_1	b'_2	\dots	b'_m	c'_1	c'_2	\dots	c'_k

- Η διαδικασία Κανονικοποίησης έχει και μειονεκτήματα:
 - Δεν είναι δημιουργική
 - Συνήθως η κανονικοποίηση γίνεται αφού έχουμε κάποιο σχήμα (μας λέει αν είναι «καλό» ή «κακό»)
 - Δεν προσφέρει ένα εννοιολογικό σχήμα (ασχολείται μόνο με σχέσεις και γνωρίσματα)

Όμως, είναι μια ενδιαφέρουσα και πρακτικά χρήσιμη προσπάθεια να γίνουν με τυπικό και συστηματικό τρόπο πράγματα που τα κάνουμε συνήθως διασθητικά.

- Ένας μεγάλος αριθμός από εμπορικά εργαλεία, δοθέντων ενός συνόλου Σχημάτων Σχέσεων/Γνωρισμάτων και ενός συνόλου συναρτησιακών εξαρτήσεων δημιουργούν αυτόματα σχήματα σχέσεων σε μορφή 3NF (σπάνια πάνε σε BCNF, 4NF και 5NF)
- Μια άλλη χρήση τέτοιων εργαλείων είναι να ελέγχουν το επίπεδο κανονικοποίησης μιας σχέσης - γενικά, η χρήση ως ευριστικό εργαλείο επιλογής ενός σχεδιασμού έναντι κάποιου άλλου
- Υπάρχουν πρακτικά αποτελέσματα της θεωρίας που επιτρέπουν σε έναν σχεδιαστή να κάνει ανάλυση της μορφής:

Αν μια σχέση είναι σε 3NF και κάθε υποχρήσιο κλειδί αποτελείται ακριβώς από ένα γνώρισμα, τότε είναι και σε 5NF (Fagin, 1971)



Η Διαδικασία Σχεδιασμού

1. Συλλογή και ανάλυση απαιτήσεων
2. Εννοιολογικός σχεδιασμός
3. Επιλογή ΣΔΒΔ
4. Απεικόνιση στο μοντέλο δεδομένων (λογικός σχεδιασμός)
5. Φυσικός σχεδιασμός
6. Υλοποίηση



Εργαλείο για υπολογισμό κλειδιού κλπ:

http://dbtools.cs.cornell.edu/norm_index.html