



Βάσεις δεδομένων

(9^ο μάθημα)

Βασίλης Ευθυμίου
vefthym@hua.gr



Περιεχόμενα

- Βελτίωση σχεδιασμού
- Αποσύνθεση σχέσης
- Συναρτησιακές εξαρτήσεις
- Θεωρία κανονικών μορφών
 - 1^η NF
 - 2^η NF
 - 3^η NF

Καλά δομημένες σχέσεις και πρωτεύοντα κλειδιά

- Καλά δομημένη σχέση
 - μία σχέση που περιέχει ελάχιστο βαθμό **πλεονασμού**
 - επιτρέπει στους χρήστες να **εισάγουν**, να **τροποποιούν** και να **διαγράφουν** γραμμές χωρίς λάθη ή ασυνέπειες
- Πρωτεύον κλειδί
 - γνώρισμα/σύνολο γνωρισμάτων για το οποίο κάθε πλειάδα έχει μία **μοναδική** τιμή
 - κάθε σχέση έχει ένα πρωτεύον κλειδί
 - έτσι διασφαλίζεται η μοναδικότητα κάθε γραμμής

Παράδειγμα – Καλά δομημένη σχέση;

EMPLOYEE2

Emp_ID	Name	Dept	Salary	Course	Date_Completed
100	Margaret Simpson	Marketing	42,000	SPSS	6/19/2020
100	Margaret Simpson	Marketing	42,000	Surveys	10/7/2020
140	Alan Beeton	Accounting	39,000	Tax Acc	12/8/2020
110	Chris Lucero	Info Systems	41,500	SPSS	1/22/2020
110	Chris Lucero	Info Systems	41,500	C++	4/22/2020
190	Lorenzo Davis	Finance	38,000	Investments	5/7/2020
150	Susan Martin	Marketing	38,500	SPSS	6/19/2020
150	Susan Martin	Marketing	38,500	TQM	8/12/2020

- Πόσες πλειάδες θα έπρεπε να αλλάξουμε αν θέλαμε να αλλάξουμε το μισθό της Margaret Simpson;
- Αν είχε ολοκληρώσει 10 μαθήματα αντί για 2;

Παράδειγμα – Καλά δομημένες σχέσεις;

EMPLOYEE1

<u>Emp_ID</u>	Name	Dept	Salary
100	Margaret Simpson	Marketing	75,000
140	Allen Beeton	Accounting	95,000
110	Chris Lucero	Info Systems	90,000
190	Lorenzo Davis	Finance	90,000
150	Susan Martin	Marketing	62,000

EMP COURSE

<u>Emp_ID</u>	<u>Course</u>	<u>Date_Completed</u>
100	SPSS	6/19/2020
100	Surveys	10/7/2020
140	Tax Acc	12/8/2020
110	SPSS	1/22/2020
110	C++	4/22/2020
190	Investments	5/7/2020
150	SPSS	6/19/2020
150	TQM	8/12/2020

- Πόσες πλειάδες θα έπρεπε να αλλάξουμε αν θέλαμε να αλλάξουμε το μισθό της Margaret Simpson;
- Αν είχε ολοκληρώσει 10 μαθήματα αντί για 2 (και υπήρχαν ήδη στον πίνακα EMP COURSE);



Σχεδίαση ΒΔ

- Μέχρι τώρα: Εννοιολογικό → Σχεσιακό μοντέλο
- **Αποσύνθεση**: Γενικός τυπικός τρόπος κατασκευής του σχήματος
 - Ξεκινάμε από το καθολικό σχήμα σχέσης (περιέχει όλα τα γνωρίσματα)
 - Προσδιορίζουμε τις **συναρτησιακές εξαρτήσεις**
 - Συνεχείς διασπάσεις του καθολικού σχήματος έτσι ώστε τα σχήματα που προκύπτουν να ικανοποιούν κάποιες ιδιότητες (να είναι σε κάποια **κανονική μορφή**)

Συναρτησιακές εξαρτήσεις και πρωτεύοντα κλειδιά

- Συναρτησιακή εξάρτηση (ΣΕ):
 - Σχέση μεταξύ δύο γνωρισμάτων
 - Για μια δεδομένη σχέση, το γνώρισμα B είναι συναρτησιακά εξαρτώμενο από το γνώρισμα A, εάν **για κάθε έγκυρη τιμή του A, αυτή η τιμή του A καθορίζει μοναδικά την τιμή του B**
 - Αναπαράσταση: $A \rightarrow B$
- Τα δεδομένα δεν αποδεικνύουν συναρτησιακή εξάρτηση!
 - αυτό ορίζεται από τη γνώση του πεδίου/προβλήματος
- Ένα γνώρισμα μπορεί να εξαρτάται συναρτησιακά από **κανένα, ένα, ή περισσότερα** γνωρίσματα

EMPLOYEE

Emp_ID	Name
100	Mary
200	Allen
190	Susan

$Emp_ID \rightarrow Name$

Emp_ID value can have only one Name

EXAMPLE

A	B	C	D
X	U	X	Y
Ⓚ	X	Z	X
Z	Y	Y	Y
Ⓚ	Z	W	Z



Αποσύνθεση - Τυπικός ορισμός

- Αρχικά ένα καθολικό σχήμα

$$R = \{A1, A2, \dots, An\}$$

- αποσυντίθεται (decomposes) σε δύο σχήματα

$$R1 = \{B1, B2, \dots, Bm\} \text{ και } R2 = \{C1, C2, \dots, Ck\}$$

- τέτοια ώστε:

1. $\{A1, A2, \dots, An\} = \{B1, B2, \dots, Bm\} \cup \{C1, C2, \dots, Ck\}$
(διατήρηση γνωρισμάτων) γνωρίσματα
2. Οι πλειάδες της $r1(R1)$ είναι η προβολή των πλειάδων της $r(R)$ στα $\{B1, B2, \dots, Bm\}$ πλειάδες
3. Οι πλειάδες της $r2(R2)$ είναι η προβολή των πλειάδων της $r(R)$ στα $\{C1, C2, \dots, Ck\}$ πλειάδες



Παράδειγμα (α1)

- Καθολική σχέση

R = {EmployeeID, Name, Position, Title, BirthDate, HireDate, Address, City, Country, DeptNo, DeptName, Location}

- Συναρτησιακές εξαρτήσεις

- DeptNo $\rightarrow$ DeptName, Location
- EmployeeID $\rightarrow$ Name, Position, Title, BirthDate, HireDate, Address, City, Country, DeptNo

- Νέες σχέσεις

R1 = {DeptNo, DeptName, Location}

R2 = {EmployeeID, Name, Position, Title, BirthDate, HireDate, Address, City, Country, DeptNo}

Παράδειγμα (α2)

- Έλεγχοι

1. $R \stackrel{?}{=} R1 \cup R2$

Σε επίπεδο πλειάδων:

EmpID	Name	...	City	Country	DeptNo	DeptName	Location
101	John		London	UK	1	HR	London
102	Mary		Bristol	UK	1	HR	London
103	Bill		Bristol	USA	2	Sales	NY

EmpID	Name	...	City	Country	DeptNo
101	John		London	UK	1
102	Mary		Bristol	UK	1
103	Bill		Bristol	USA	2

DeptNo	DeptName	Location
1	HR	London
1	HR	London
2	Sales	NY

2. οι πλειάδες της $r1(R1)$ είναι προβολή των πλειάδων της $r(R)$ στα γνωρίσματα της $R1$;

3. Όμοια για την $r2(R2)$;

Μπορούμε να πάρουμε την αρχική σχέση με συνένωση των $R1$ και $R2$;

Λανθασμένη αποσύνθεση

EmpID	Name	...	City	Country	DeptNo	DeptName	Location
101	John		London	UK	1	HR	London
102	Mary		Bristol	UK	1	HR	London
103	Bill		Bristol	USA	2	Sales	NY



EmpID	Name		City	Country
101	John		London	UK
102	Mary		Bristol	UK
103	Bill		Bristol	USA

City	Country	DeptNo	DeptName	Location
London	UK	1	HR	London
Bristol	UK	1	HR	London
Bristol	USA	2	Sales	NY

Μπορούμε να πάρουμε την αρχική σχέση με συνένωση των R1 και R2;

Γενικότερα

- Για ένα σχεσιακό σχήμα R μπορούμε να παράγουμε μια αποσύνθεσή του σε σχήματα R_i που είναι υποσύνολα γνωρισμάτων του R και η ένωσή τους δίνει το R
- Επιπλέον πρέπει τα νέα στιγμιότυπα να δίνουν τα αρχικά μετά από φυσική συνένωση
- Δηλαδή για τα **γνωρίσματα**: $R \rightarrow \{R_1, R_2, \dots, R_n\}$
 - αρκεί $R = R_1 \cup R_2 \dots \cup R_n$
- Για τα **στιγμιότυπα**
 - Έστω $r(R)$ και $r_i = \pi_{R_i}(r)$, $\forall i = 1, \dots, n$
 - πρέπει $r \subseteq r_1 * r_2 * \dots * r_n$

Παράδειγμα (β1)

- Έστω το σχήμα $R(A, B, C)$.
- Παράγουμε μια αποσύνθεσή του:
 $R_1(A, B)$ και $R_2(B, C)$ όπου $(R_1 \cap R_2 = B)$

$r(R)$			$r_1(R_1)$		$r_2(R_2)$	
A	B	C	A	B	B	C
1	2	3	1	2	2	3
4	2	5	4	2	2	5

- Τι γίνεται με τα στιγμιότυπα (σχέσεις) που ανήκουν στο R , συμβολισμός $r(R)$;
- Μπορούμε να πάρουμε το αρχικό στιγμιότυπο με φυσική συνένωση; $(r_1 * r_2)$

Παράδειγμα (β2)

- Δεν είναι αυτό που θέλουμε
- Δεν μπορούμε να πάρουμε την αρχική σχέση r από τα r_1 και r_2

Αποσύνθεση
με απώλειες

$\mathbf{r}$	A	B	C		$\mathbf{r}_1$	A	B		$\mathbf{r}_2$	B	C		$\mathbf{r}_1 * \mathbf{r}_2$	A	B	C	$\mu\epsilon$
	1	2	3											1	2	3	
	4	2	5			1	2			2	3			1	2	5	
						4	2			2	5			4	2	3	
														4	2	5	

Παράδειγμα (β3)

- Έστω το σχήμα $R(A, B, C)$ αποσυντίθεται σε
 $R_1(A, C)$ και $R_2(B, C)$ όπου $(R_1 \cap R_2 = C)$
- Τι γίνεται με τα στιγμιότυπα (σχέσεις) που ανήκουν στο R ;
- Μπορούμε να πάρουμε το αρχικό στιγμιότυπο;

$r(R)$		
A	B	C
1	2	3
4	2	5

$r_1(R_1)$	
A	C
1	3
4	5

$r_2(R_2)$	
B	C
2	3
2	5

**Αποσύνθεση
χωρίς
απώλειες**

$r_1 * r_2$		
A	B	C
1	2	3
4	2	5

Επιθυμητές ιδιότητες

- Συνένωση χωρίς απώλειες
- Διατήρηση εξαρτήσεων
- Αποφυγή επανάληψης πληροφορίας

Επιθυμητή Ιδιότητα (1)

- Συνένωση χωρίς απώλειες
 - Έστω C το σύνολο περιορισμών.
 - Μια αποσύνθεση του R σε {R1, R2} είναι μια **αποσύνθεση άνευ απωλειών στη συνένωση** (lossless join decomposition) αν για όλες τις σχέσεις r(R) που είναι νόμιμες στο C ισχύει

$$r = \pi_{R1}(r) * \pi_{R2}(r)$$

- Αρκεί τα κοινά γνωρίσματα των δύο σχέσεων (R1, R2) να είναι κλειδί για τουλάχιστον μια από τις δύο σχέσεις

Παράδειγμα (α)

- Καθολική σχέση

$R = \{EmployeeID, Name, Position, Title, BirthDate, HireDate, AddressCity, Country, DeptNo, DeptName, Location\}$

- Συναρτησιακές εξαρτήσεις

- $DeptNo \rightarrow DeptName, Location$
- $EmployeeID \rightarrow Name, Position, Title, BirthDate, HireDate, AddressCity, Country, DeptNo$

- Νέες σχέσεις

$R1 = \{DeptNo, DeptName, Location\}$

$R2 = \{EmployeeID, Name, Position, Title, BirthDate, HireDate, AddressCity, Country, DeptNo\}$

$R_1 \cap R_2 = \{DeptNo\}$

Όμως το DeptNo είναι κλειδί στο R1

Παράδειγμα (β)

$r_1 * r_2$

r	r_1	r_2	
A B C	A B	B C	A B C
1 2 3	1 2	2 3	1 2 3
4 2 5	4 2	2 5	1 2 5
			4 2 3
			4 2 5

- $R1 \cap R2 = \{B\}$
- Συναρτησιακές εξαρτήσεις
 - $A \rightarrow B$
 - $A \rightarrow C$
 - $C \rightarrow A$
 - $B \rightarrow ??$
- Το B δε συμμετέχει σε καμία ΣΕ, άρα δεν εξαρτά τα R_1 ή R_2
- Άρα η διάσπαση είναι με απώλειες



Επιθυμητή Ιδιότητα (2)

- **Διατήρηση Εξαρτήσεων**

- Για να ελέγχουμε ότι διατηρούνται οι Σ.Ε. όταν γίνονται τροποποιήσεις σε μία από τις σχέσεις $r_i(R_i)$ να αρκεί να ελέγχουμε τη συγκεκριμένη σχέση
- να μη χρειάζεται να υπολογίσουμε τις αρχικές σχέσεις - αποφυγή των συνενώσεων



Παράδειγμα (α)

- Νέες σχέσεις

R1 = {DeptNo, DeptName, Location}

R2 = {EmployeeID, Name, Position, Title, BirthDate, HireDate, AddressCity, Country, DeptNo}

- Οι συναρτησιακές εξαρτήσεις

- DeptNo $\rightarrow$ DeptName, Location

- EmployeeID $\rightarrow$ Name, Position, Title, BirthDate, HireDate, AddressCity, Country, DeptNo

εξακολουθούν να ισχύουν

- Η αποσύνθεση είναι μια αποσύνθεση που διατηρεί τις εξαρτήσεις (dependency preserving)

Επιθυμητή Ιδιότητα (3)

- Αποφυγή επανάληψης πληροφορίας
 - Μας δίνεται ένα σχήμα και πρέπει να αποφασίσουμε αν είναι «καλό» ή χρειάζεται περαιτέρω διάσπαση
- Αν ένα σχήμα είναι σε κάποια Κανονική Μορφή δεν υπάρχουν συγκεκριμένα προβλήματα
- Κανονικές μορφές (normal forms): από την πιο περιοριστική στη λιγότερο περιοριστική
5NF 4NF BCNF 3NF 2NF 1NF
- Βασίζεται σε Σ.Ε., οι Σ.Ε. έχουν σχέση με την επανάληψη πληροφορίας



Κανονικοποίηση

- Μετατροπή μίας ΒΔ ώστε να ακολουθεί κάποια κανονική μορφή
- Πρώτη Κανονική Μορφή (1NF)
 - μοναδικές γραμμές, όχι γνωρίσματα με πολλές τιμές
 - όλες οι σχέσεις είναι σε 1NF
- Δεύτερη Κανονική Μορφή (2NF)
 - Κάθε μη-πρωτεύον γνώρισμα αναγνωρίζεται (συναρτησιακή εξάρτηση) από ολόκληρο το κλειδί
- Τρίτη Κανονική Μορφή (3NF)
 - Μη-πρωτεύοντα γνωρίσματα εξαρτώνται μόνο από κλειδιά (μη-μεταβατικές εξαρτήσεις)

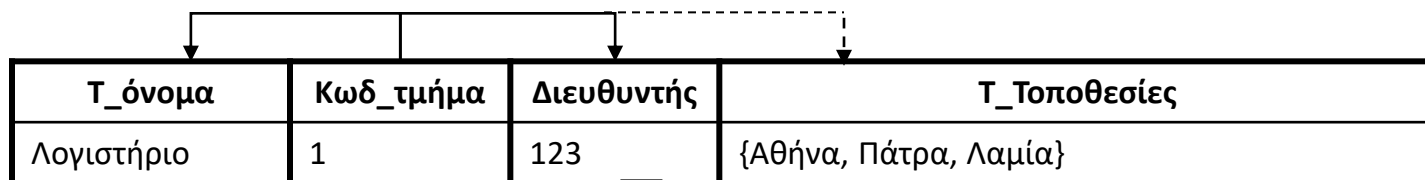
Αποτέλεσμα κανονικοποίησης: κάθε γνώρισμα που δεν ανήκει στο πρωτεύον κλειδί εξαρτάται από **ολόκληρο** το πρωτεύον κλειδί και **μονάχα** από αυτό



1^η κανονική μορφή (1NF)

- Τα γνωρίσματα μπορούν να περιλαμβάνουν μόνο ατομικές (απλές, αδιαίρετες) τιμές
- Η τιμή ενός γνωρίσματος σε μια πλειάδα πρέπει να είναι μία και μόνη τιμή από το πεδίο ορισμού του γνωρίσματος
- Ορίστηκε για να απαγορεύονται τα πλειότιμα γνωρίσματα, τα σύνθετα γνωρίσματα και οι συνδυασμοί τους (εμφωλευμένες σχέσεις)

Παράδειγμα



Τ_όνομα	Κωδ_τμήμα	Διευθυντής	Τ_Τοποθεσίες
Λογιστήριο	1	123	{Αθήνα, Πάτρα, Λαμία}

Τ_όνομα	Κωδ_τμήμα	Διευθυντής	Τ_Τοποθεσίες
Λογιστήριο	1	123	Αθήνα
Λογιστήριο	1	123	Πάτρα
Λογιστήριο	1	123	Λαμία

Τ_όνομα	Κωδ_τμήμα	Διευθυντής	Τ_Τοποθεσίες
Λογιστήριο	1	123	Αθήνα
Λογιστήριο	1	123	Πάτρα
Λογιστήριο	1	123	Λαμία

Τ_όνομα	Κωδ_τμήμα	Διευθυντής
Λογιστήριο	1	123

Κωδ_τμήμα	Τ_Τοποθεσίες
1	Αθήνα
1	Πάτρα
1	Λαμία

Μετατροπή σε 1NF

- Για τα πλειότιμα
 - Βρίσκουμε το πρωτεύον κλειδί της σχέσης $R (A_1, A_2, \dots, A_N)$ π.χ. A_1
 - Βρίσκουμε το πλειότιμο γνώρισμα π.χ. A_2
 - Αποσύνθεση της R σε $R_1(A_1, A_3, \dots, A_N)$ και $R_2(A_1, A_2)$
- Για τα σύνθετα, όμοια
 - Βρίσκουμε το πρωτεύον κλειδί της σχέσης $R (A_1, A_2, \dots, A_N)$ π.χ. A_1
 - Βρίσκουμε το σύνθετο γνώρισμα π.χ. $A_2 A_3 A_4$
 - Αποσύνθεση της R σε $R_1(A_1, A_5, \dots, A_N)$ και $R_2(A_1, A_2, A_3, A_4)$



2^η κανονική μορφή (2NF)

- Βασίζεται στην **πλήρη συναρτησιακή εξάρτηση**
 - $X \rightarrow Y \Rightarrow \forall A \in X, (X - \{A\}) \not\rightarrow Y$
- Το σχήμα R είναι σε 2NF αν
 - είναι σε 1NF και
 - κάθε μη πρωτεύον γνώρισμα Y είναι πλήρως συναρτησιακά εξαρτώμενο από το πρωτεύον κλειδί του R



Μετατροπή σε 2NF

ΕΝΤΟΠΙΣΜΟΣ

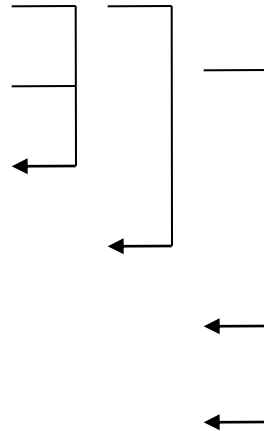
- Βρίσκουμε το πρωτεύον κλειδί της σχέσης $R (A_1, A_2, \dots, A_N)$
 - Αν είναι απλό (π.χ. A_1): οι ΣΕ είναι πλήρεις
 - Αν είναι σύνθετο (π.χ. A_1A_2): καταγράφουμε τις ΣΕ από το πρωτεύον κλειδί
- Εξετάζουμε αν οι ΣΕ είναι πλήρεις ή μερικές
 - Αν ισχύει η $A_1 \rightarrow A_3$ (μερική εξάρτηση), τότε η R δεν είναι σε 2NF

ΜΕΤΑΤΡΟΠΗ

- Αποσυνθέτω την R σε σχέσεις R_i που ομαδοποιούν τα μη πρωτεύοντα γνωρίσματα και το μέρος του κλειδιού που τα εξαρτά πλήρως
π.χ. $R_1(A_1, A_2, A_4, \dots, A_N), R_2(A_1, A_3)$

Παράδειγμα

ΕΡΓΑΖ_ΕΡΓΟ
<u>Αρ_ταυτ</u>
<u>Κωδ_έργου</u>
Ώρες
Εργαζ_Όνομα
Εργ_Όνομα
Τοπ_έργου



Είναι σε 1NF καθώς δεν έχει σύνθετα ή πλειότιμα γνωρίσματα
Ελέγχουμε τις εξαρτήσεις από το πρωτεύον κλειδί

1. Αρ_ταυτ, Κωδ_έργου $\rightarrow$ Ώρες
2. Αρ_ταυτ, Κωδ_έργου $\rightarrow$ Εργαζ_Όνομα
3. Αρ_ταυτ, Κωδ_έργου $\rightarrow$ Εργ_Όνομα
4. Αρ_ταυτ, Κωδ_έργου $\rightarrow$ Τοπ_έργου

Ελέγχουμε αν είναι πλήρεις:

1. Πλήρης
2. Μερική (λόγω της b)
3. Μερική (λόγω της c1)
4. Μερική (λόγω της c2)

Γνωρίζουμε τις ΣΕ:

- a) Αρ_ταυτ, Κωδ_έργου $\rightarrow$ Ώρες
- b) Αρ_ταυτ $\rightarrow$ Εργαζ_Όνομα
- c) Κωδ_έργου $\rightarrow$ Εργ_Όνομα, Τοπ_Εργου $\rightarrow$
 - c1) Κωδ_έργου $\rightarrow$ Εργ_Όνομα
 - c2) Κωδ_έργου $\rightarrow$ Τοπ_Εργου

Παράδειγμα



ΕΡΓΑΖ_ΕΡΓΟ
<u>Αρ ταυτ</u>
<u>Κωδ έργου</u>
Ώρες
Εργαζ_Ονομα
Εργ_Όνομα
Τοπ_έργου

ΕΡΓΑΖΟΜΕΝΟΣ
<u>Αρ ταυτ</u>
Εργαζ_Ονομα

ΕΡΓΟ
<u>Κωδ έργου</u>
Εργ_Όνομα
Τοπ_έργου

ΕΡΓΑΖ_ΕΡΓΟ_ΩΡΕΣ
<u>Αρ ταυτ</u>
<u>Κωδ έργου</u>
Ώρες

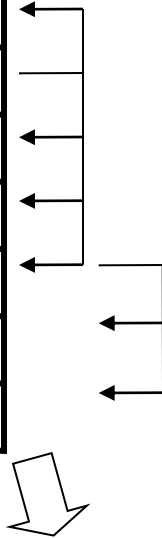


3^η κανονική μορφή (3NF)

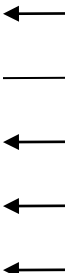
- Βασίζεται στη μεταβατική εξάρτηση
Η $X \rightarrow Y$ είναι μεταβατική αν $X \rightarrow Z$ και $Z \rightarrow Y$
- Το σχήμα R είναι σε 3NF αν
 - είναι σε 2NF και
 - κανένα μη πρωτεύον γνώρισμα Y δεν εξαρτάται μεταβατικά από το πρωτεύον κλειδί του R
- ΜΕΤΑΤΡΟΠΗ
 - Αποσυνθέτω την R σε σχέσεις R_i που ομαδοποιούν τα μη πρωτεύοντα ορίσματα και το μέρος του κλειδιού που τα εξαρτά πλήρως
π.χ. R (A_1 , $A_2, \dots, A_N$) και $A_1 \rightarrow A_2$ και $A_2 \rightarrow A_3$
 $R_1(\underline{A_1}, A_2, A_4, \dots, A_N)$ και $R_2(\underline{A_2}, A_3)$

Παράδειγμα

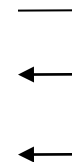
ΕΡΓΑΖ_ΤΜΗΜΑ
Εργαζ_Ονομα
<u>Αρ_ταυτ</u>
Ημ_Γεν
Διεύθυνση
Κωδ_τμήμα
Τμ_όνομα
Διευθυντής



ΕΡΓΑΖΟΜΕΝΟΣ
Εργαζ_Ονομα
<u>Αρ_ταυτ</u>
Ημ_Γεν
Διεύθυνση
Κωδ_τμήμα



ΤΜΗΜΑ
Κωδ_τμήμα
Τμ_όνομα
Διευθυντής



Είναι σε 1NF καθώς δεν έχει σύνθετα ή πλειότιμα γνωρίσματα
Είναι σε 2NF καθώς το πρωτεύον κλειδί είναι απλό και συνεπώς δεν υπάρχουν μερικές εξαρτήσεις
Έχει όμως δυο μεταβατικές εξαρτήσεις

1. ***Αρ_ταυτ* → Τμ_όνομα**
Αρ_ταυτ → Κωδ_τμήμα
Κωδ_τμήμα → Τμ_όνομα
2. ***Αρ_ταυτ* → Διευθυντής**
Αρ_ταυτ → Κωδ_τμήμα
Κωδ_τμήμα → Διευθυντής

Γενικός ορισμός 2NF

- 2NF: Κάθε μη πρωτεύον γνώρισμα A του R είναι πλήρως εξαρτώμενο από οποιοδήποτε κλειδί του R

Γνωρίζουμε τις:

a) Κωδικός $\rightarrow$ Έτος, Τίτλος, Διάρκεια, Είδος

b) Έτος, Τίτλος $\rightarrow$ Κωδικός, Διάρκεια, Είδος

c) Τίτλος $\rightarrow$ Είδος

d) Σκηνοθέτης $\rightarrow$ Χώρα

Πρωτεύον Κλειδί: Κωδικός

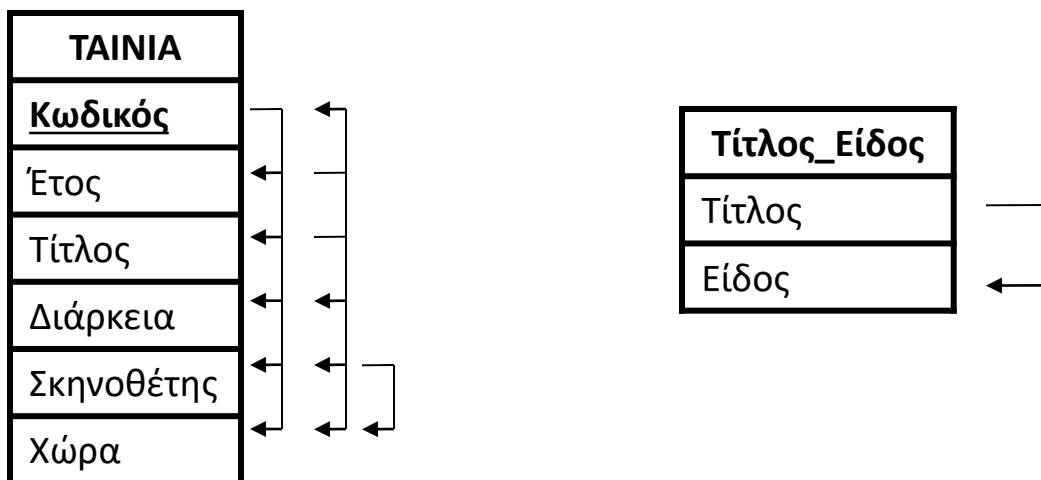
Κλειδί: Έτος, Τίτλος

Όλα τα γνωρίσματα είναι πλήρως εξαρτημένα από το πρωτεύον κλειδί

Το γνώρισμα *Είδος* δεν είναι πλήρως εξαρτημένο από το κλειδί (*Τίτλος*, *Έτος*)

ΤΑΙΝΙΑ	
<u>Κωδικός</u>	←
Έτος	←
Τίτλος	←
Διάρκεια	←
Είδος	←
Σκηνοθέτης	←
Χώρα	←

Παράδειγμα



Πρωτεύον Κλειδί: Κωδικός

Κλειδί: Έτος, Τίτλος

Όλα τα γνωρίσματα είναι πλήρως εξαρτημένα από το πρωτεύον κλειδί και από τα υπόλοιπα κλειδιά



Γενικός ορισμός 3NF

- 3NF: Όταν εμφανίζεται μια συναρτησιακή εξάρτηση $X \rightarrow A$ στο R τότε
 - το X είναι υπερκλειδί του R
 - ή*
 - το A είναι πρωτεύον γνώρισμα του R
- 3NF: Κάθε μη πρωτεύον γνώρισμά του είναι
 - πλήρως συναρτησιακά εξαρτώμενο από κάθε κλειδί του R
 - και*
 - μη μεταβατικά εξαρτώμενο από οποιοδήποτε κλειδί του R

Παράδειγμα

Γνωρίζουμε τις:

- a) Κωδικός → Έτος, Τίτλος, Διάρκεια, Είδος
- b) Έτος, Τίτλος → Κωδικός, Διάρκεια, Είδος
- c) Τίτλος → Είδος
- d) Σκηνοθέτης → Χώρα

Στη ΣΕ c) ο *Τίτλος* δεν είναι υπερκλειδί και το *Είδος* δεν είναι πρωτεύον γνώρισμα

Στη ΣΕ d) ο *Σκηνοθέτης* δεν είναι υπερκλειδί και η *Χώρα* δεν είναι πρωτεύον γνώρισμα



ΤΑΙΝΙΑ
<u>Κωδικός</u>
Έτος
Τίτλος
Διάρκεια
Είδος
Σκηνοθέτης
Χώρα

ΤΑΙΝΙΑ
<u>Κωδικός</u>
Έτος
Τίτλος
Διάρκεια
Σκηνοθέτης

Τίτλος_Είδος
Τίτλος
Είδος

Σκηνοθέτης_Χώρα
Σκηνοθέτης
Χώρα

1NF

STUD_NO	STUD_NAME	STUD_PHONE	STUD_STATE	STUD_COUNTRY
1	RAM	9716271721, 9871717178	HARYANA	INDIA
2	RAM	9898297281	PUNJAB	INDIA
3	SURESH		PUNJAB	INDIA

Table 1

Conversion to first normal form

STUD_NO	STUD_NAME	STUD_PHONE	STUD_STATE	STUD_COUNTRY
1	RAM	9716271721	HARYANA	INDIA
1	RAM	9871717178	HARYANA	INDIA
2	RAM	9898297281	PUNJAB	INDIA
3	SURESH		PUNJAB	INDIA

Table 2

μοναδικές γραμμές, όχι γνωρίσματα με πολλές τιμές
όλες οι σχέσεις είναι σε 1NF

2NF

EMPLOYEE2

<u>Emp_ID</u>	Name	Dept	Salary	<u>Course</u>	Date_Completed
100	Margaret Simpson	Marketing	42,000	SPSS	6/19/2020
100	Margaret Simpson	Marketing	42,000	Surveys	10/7/2020
140	Alan Beeton	Accounting	39,000	Tax Acc	12/8/2020
110	Chris Lucero	Info Systems	41,500	SPSS	1/22/2020
110	Chris Lucero	Info Systems	41,500	C++	4/22/2020
190	Lorenzo Davis	Finance	38,000	Investments	5/7/2020
150	Susan Martin	Marketing	38,500	SPSS	6/19/2020
150	Susan Martin	Marketing	38,500	TQM	8/12/2020

Συναρτησιακές Εξαρτήσεις:

Emp_ID → Name, Dept, Salary ❌

Emp_ID, Course → Date_Completed ✅

Κάθε μη-πρωτεύον γνώρισμα αναγνωρίζεται (συναρτησιακή εξάρτηση) από **ολόκληρο** το κλειδί (και όχι από αυστηρά υποσύνολά του)

2NF

EMPLOYEE1

<u>Emp_ID</u>	Name	Dept	Salary
100	Margaret Simpson	Marketing	75,000
140	Allen Beeton	Accounting	95,000
110	Chris Lucero	Info Systems	90,000
190	Lorenzo Davis	Finance	90,000
150	Susan Martin	Marketing	62,000

EMP COURSE

<u>Emp_ID</u>	Course	Date_ Completed
100	SPSS	6/19/2020
100	Surveys	10/7/2020
140	Tax Acc	12/8/2020
110	SPSS	1/22/2020
110	C++	4/22/2020
190	Investments	5/7/2020
150	SPSS	6/19/2020
150	TQM	8/12/2020

Συναρτησιακές Εξαρτήσεις:

Emp_ID → Name, Dept, Salary ✓

Emp_ID, Course → Date_Completed ✓

Κάθε μη-πρωτεύον γνώρισμα αναγνωρίζεται (συναρτησιακή εξάρτηση) από **ολόκληρο** το κλειδί (και όχι από αυστηρά υποσύνολά του)

3NF

SALES

<u>Customer_ID</u>	Customer_Name	Salesperson	Region
8023	Anderson	Smith	South
9167	Bancroft	Hicks	West
7924	Hobbs	Smith	South
6837	Tucker	Hernandez	East
8596	Eckersley	Hicks	West
7018	Arnold	Faulb	North

Συναρτησιακές Εξαρτήσεις:

Customer_ID → Customer_Name, Salesperson, Region ✓

Salesperson → Region ✗

Μη-μεταβατικές εξαρτήσεις

3NF

SALES1

<u>Customer_ID</u>	Customer_Name	<u>Salesperson</u>
8023	Anderson	Smith
9167	Bancroft	Hicks
7924	Hobbs	Smith
6837	Tucker	Hernandez
8596	Eckersley	Hicks
7018	Arnold	Faulb

SPERSON

<u>Salesperson</u>	Region
Smith	South
Hicks	West
Hernandez	East
Faulb	North

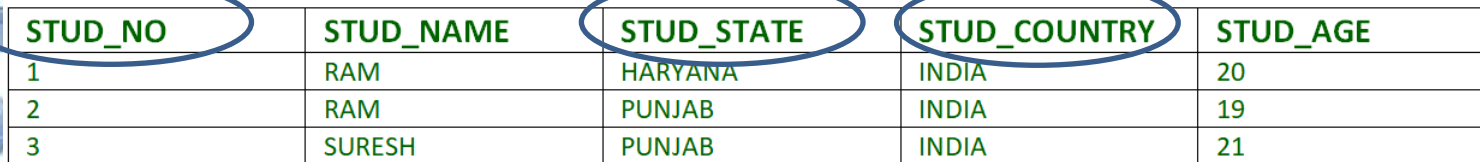
Συναρτησιακές Εξαρτήσεις:

Customer_ID → Customer_Name, Salesperson ✓

Salesperson → Region ✓

Μη-μεταβατικές εξαρτήσεις

3NF – Εξάσκηση



STUD_NO	STUD_NAME	STUD_STATE	STUD_COUNTRY	STUD_AGE
1	RAM	HARYANA	INDIA	20
2	RAM	PUNJAB	INDIA	19
3	SURESH	PUNJAB	INDIA	21

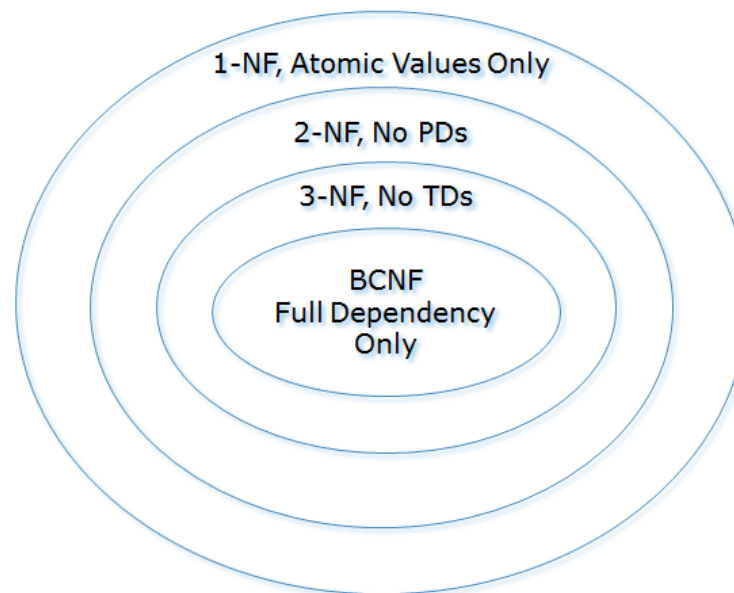
μετατροπή σε 3NF ↓

STUD_NO	STUD_NAME	STUD_STATE	STUD_AGE
1	RAM	HARYANA	20
2	RAM	PUNJAB	19
3	SURESH	PUNJAB	21

STUD_STATE	STUD_COUNTRY
HARYANA	INDIA
PUNJAB	INDIA

Μη-μεταβατικές εξαρτήσεις

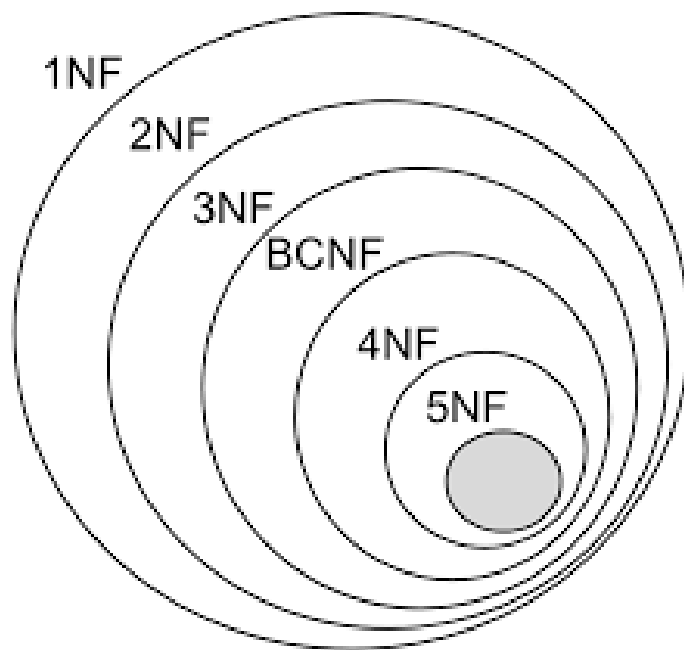
Κανονικές Μορφές - Σχηματικά



*PD: Partial Dependencies

*TD: Transitive Dependencies

Κανονικές Μορφές - Σχηματικά



Ron Fagin
IBM Fellow



Κανονικοποίηση

- Πλεονεκτήματα
 - Αποφυγή επαναλήψεων
- Μειονεκτήματα
 - Γίνεται αφού έχουμε κάποιο σχήμα και το αξιολογεί
 - Δεν προσφέρει ένα εννοιολογικό σχήμα (ασχολείται μόνο με σχέσεις και γνωρίσματα)
 - Ενδέχεται να καθυστερεί τις αναζητήσεις γιατί συχνά απαιτεί συνενώσεις σχέσεων
- Προσπάθεια να γίνουν με τυπικό και συστηματικό τρόπο πράγματα που τα κάνουμε συνήθως διαισθητικά
- Σε πολλά ΣΔΒΔ υπάρχουν εργαλεία ανάλυσης που φτάνουν με αυτόματο τρόπο σε σχήματα σχέσεων σε μορφή 3NF (σπάνια πάνε σε BCNF, 4NF και 5NF)
- Ταυτόχρονα μπορούν να ελέγξουν το βαθμό κανονικοποίησης μιας σχέσης και να βοηθήσουν στην επιλογή ενός καλού συνδυασμού