

# Λογικός Σχεδιασμός

# Εισαγωγή

Θα εξετάσουμε πότε ένα σχεσιακό σχήμα για μια βάση δεδομένων είναι «καλό»

- Μη τυπικές γενικές κατευθύνσεις
- Θεωρία **κανονικών μορφών** η οποία βασίζεται στην έννοια των **συναρτησιακών εξαρτήσεων**

# Γενικές Κατευθύνσεις

1. Σηματολογία
2. Ελάττωση πλεονασμού
3. Ελάττωση τιμών null
4. Μη πλασματικές πλειάδες

# Σημασιολογία

- Εύκολη η εξήγηση της σημασίας του
- Αποφυγή συνδυασμού γνωρισμάτων από πολλές οντότητες και συσχετίσεις στην ίδια σχέση

**Ταινία**

<u>Τίτλος</u>	<u>Έτος</u>	Διάρκεια	Είδος
---------------	-------------	----------	-------

**Παίζει**

<u>Όνομα</u>	<u>Τίτλος</u>	<u>Έτος</u>
--------------	---------------	-------------

**Ηθοποιός**

<u>Όνομα</u>	Διεύθυνση	Έτος-Γέννησης
--------------	-----------	---------------

# Πλεονασμός (επανάληψη πληροφορίας)

Ταινία

<u>Τίτλος</u>	<u>Έτος</u>	Διάρκεια	Είδος	<u>Όνομα-Ηθοποιού</u>
---------------	-------------	----------	-------	-----------------------

εδώ βοηθούν οι συναρτησιακές εξαρτήσεις

## Εισαγωγή

- Για την εισαγωγή μιας νέας ταινίας πρέπει να εισάγουμε τουλάχιστον έναν ηθοποιό (τιμή null;)
- Για την εισαγωγή ενός ηθοποιού στην ταινία πρέπει να επαναλάβουμε τα γνωρίσματα (διάρκεια, είδος) της ταινίας

## Διαγραφή

- Τι γίνεται αν διαγράψουμε και τον τελευταίο ηθοποιό
- Διαγραφή μιας ταινίας;

# Πλεονασμός (επανάληψη πληροφορίας)

Ταινία

<u>Τίτλος</u>	<u>Έτος</u>	Διάρκεια	Είδος	<u>Όνομα-Ηθοποιού</u>
---------------	-------------	----------	-------	-----------------------

## Τροποποίηση

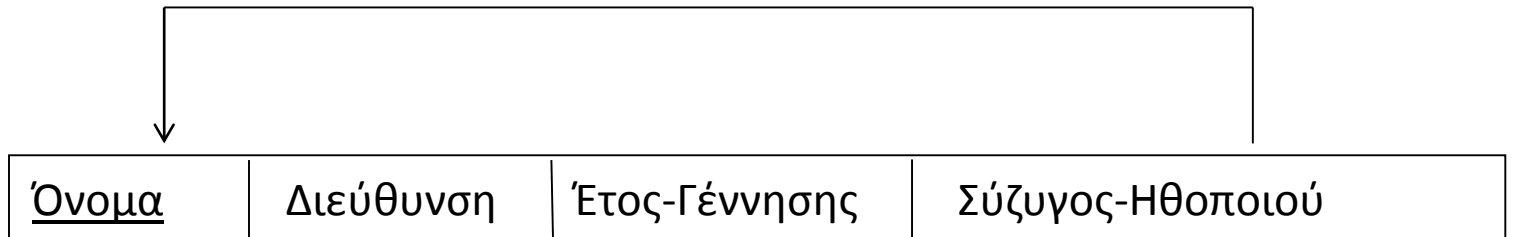
- Τι γίνεται αν θελήσουμε να τροποποιήσουμε τη διάρκεια μιας ταινίας;

## Σύνοψη Προβλημάτων Λόγω Πλεονασμού

- Πλεονασμός στην αποθήκευση
- Προβληματική ενημέρωση
- Προβληματική εισαγωγή
- Προβληματική διαγραφή

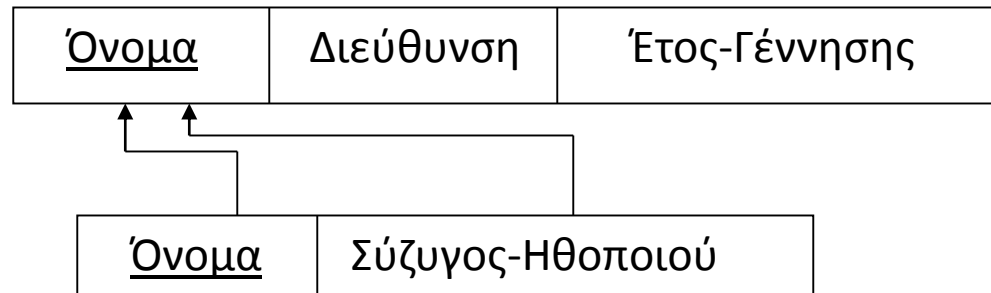
# Αποφυγή τιμών null

Ηθοποιός



Ηθοποιός

Ζευγάρι-Ηθοποιών



# Αποφυγή δημιουργίας πλασματικών πλειάδων

(αδυναμία αναπαράστασης συγκεκριμένης πληροφορίας)

<b>Ταινία</b>	<u>Τίτλος</u>	<u>Έτος</u>	Διάρκεια	Είδος
---------------	---------------	-------------	----------	-------

<b>Παίζει</b>	<u>Τίτλος</u>	<u>Όνομα-Ηθοποιού</u>
---------------	---------------	-----------------------

*Χάνουμε πληροφορία δεν μπορούμε να βρούμε ποιος ηθοποιός σε ποια ταινία*

<b>Ταινία</b>	<u>Τίτλος</u>	<u>Έτος</u>	Διάρκεια	Είδος	<u>Όνομα-Ηθοποιού</u>
---------------	---------------	-------------	----------	-------	-----------------------



# Γενικός Αλγόριθμος Σχεδιασμού

Ο τρόπος που σχεδιάζαμε ένα σχήμα ΒΔ μέχρι τώρα:

- ✓ από το εννοιολογικό στο σχεσιακό μοντέλο

Θα δούμε ένα **γενικό τυπικό τρόπο** κατασκευής του σχήματος

Γενικά:

- Ξεκινάμε από το **καθολικό σχήμα** (που περιέχει όλα τα γνωρίσματα)
- **Διαδοχικές διασπάσεις** έτσι ώστε τα σχήματα που προκύπτουν να ικανοποιούν κάποιες ιδιότητες (με βάση συναρτησιακές εξαρτήσεις)

# Συναρτησιακές Εξαρτήσεις

# Εισαγωγή

- *Τι είναι;*

Εξαρτήσεις ανάμεσα σε σύνολα από γνωρίσματα

- *Συμβολισμός*

**S1** → **S2** (όπου S1, S2 σύνολα γνωρισμάτων)

- *Τι σημαίνει:*

Αν ίδιες τιμές στα γνωρίσματα του S1  $\Rightarrow$  ίδιες τιμές στα γνωρίσματα του S2

# Παράδειγμα

Παράδειγμα: Σχήμα σχέσης R(A, B, C, D) (Υπενθύμιση συμβολισμού)

Στιγμιότυπο,  $r(R)$

	A	B	C	D
<b>r1</b>	$a_1$	$b_1$	$c_1$	$d_1$
<b>r2</b>	$a_1$	$b_2$	$c_1$	$d_2$
<b>r3</b>	$a_2$	$b_3$	$c_2$	$d_3$
<b>r4</b>	$a_3$	$b_3$	$c_2$	$d_4$

Συμβολισμός

$$r1[A] = a_1$$

$$r2[BC] = b_2 c_1$$

# Ορισμός

Έστω ένα σχήμα σχέσης  $R(A_1, A_2, \dots, A_n)$ .

Θα συμβολίζουμε με  $R = \{A_1, A_2, \dots, A_n\}$  το σύνολο των γνωρισμάτων της  $R$ .

Έστω  $X \subseteq R$  και  $Y \subseteq R$ ,

μια **συναρτησιακή εξάρτηση (functional dependency)**

$$X \rightarrow Y$$

ισχύει στο σχήμα  $R$

αν για κάθε σχέση  $r(R)$ , για κάθε ζεύγος πλειάδων  $t_1$  και  $t_2$  της  $r$ , ισχύει  $t_1[X] = t_2[X] \Rightarrow t_1[Y] = t_2[Y]$  (If  $t_1[X] = t_2[X]$  then  $t_1[Y] = t_2[Y]$ )

Με απλά λόγια, μια συναρτησιακή εξάρτηση  $X \rightarrow Y$  μας λέει ότι αν οποιεσδήποτε δυο πλειάδες μιας σχέσης της  $R$  συμφωνούν (έχουν την ίδια τιμή) στα γνωρίσματα  $X \subseteq R$  τότε συμφωνούν (έχουν την ίδια τιμή) και στα γνωρίσματα  $Y \subseteq R$ .

# Συναρτησιακές Εξαρτήσεις

Αντί  $\{A_1, A_2, \dots, A_n\} \rightarrow \{B_1, B_2, \dots, B_m\}$  γράφουμε

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

✓ **Ισχύουν στο σχήμα** - δηλαδή για όλες τις πιθανές σχέσεις (πλειάδες)

Παράδειγμα: Ποιες (μη τετριμμένες) συναρτησιακές εξαρτήσεις δεν παραβιάζει η παρακάτω σχέση – δεν ξέρουμε αν ισχύουν στο σχήμα

Μπορούμε όμως να πούμε ποιες **δεν ισχύουν**

A	B	C	D
a <sub>1</sub>	b <sub>1</sub>	c <sub>1</sub>	d <sub>1</sub>
a <sub>1</sub>	b <sub>2</sub>	c <sub>1</sub>	d <sub>2</sub>
a <sub>2</sub>	b <sub>3</sub>	c <sub>2</sub>	d <sub>3</sub>
a <sub>3</sub>	b <sub>3</sub>	c <sub>2</sub>	d <sub>4</sub>

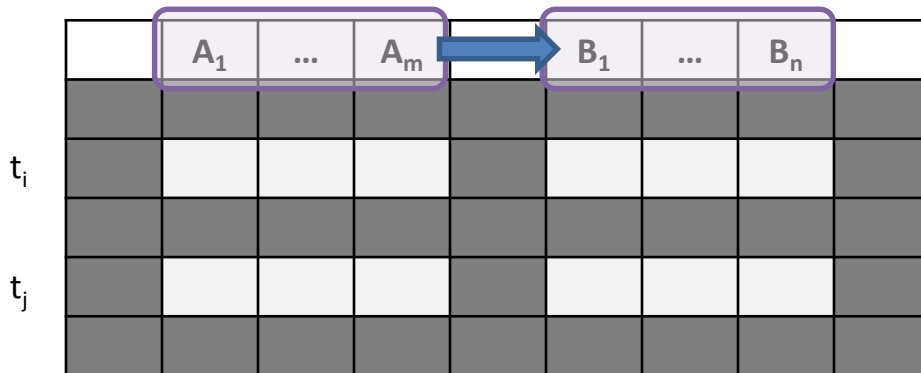
# Συναρτησιακές Εξαρτήσεις (επανάληψη)

	$A_1$	...	$A_m$		$B_1$	...	$B_n$	

Ορισμός:

Έστω δύο σύνολα γνωρισμάτων  
 $\mathbf{A} = \{A_1, \dots, A_m\}$  και  $\mathbf{B} = \{B_1, \dots, B_n\}$  ενός  
πίνακα  $\mathbf{R}$ ,

# Συναρτησιακές Εξαρτήσεις (εικόνα)



Ορισμός:

Έστω δύο σύνολα γνωρισμάτων  $A = \{A_1, \dots, A_m\}$  και  $B = \{B_1, \dots, B_n\}$  ενός πίνακα  $R$ ,

Η *συναρτησιακή εξάρτηση*  $A \rightarrow B$  στο  $R$



# Συναρτησιακές Εξαρτήσεις (εικόνα)

	$A_1$	...	$A_m$		$B_1$	...	$B_n$	
$t_i$								
$t_j$								

Αν  $t_1, t_2$  συμφωνούν εδώ..

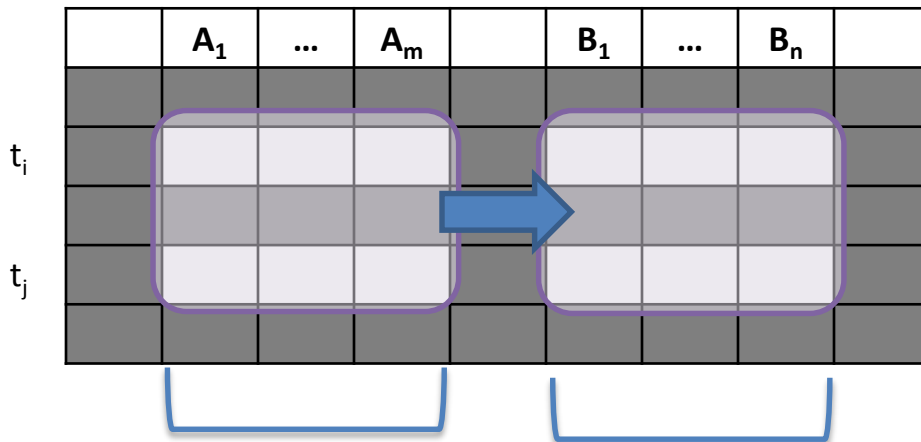
Ορισμός:

Έστω δύο σύνολα γνωρισμάτων  $A = \{A_1, \dots, A_m\}$  και  $B = \{B_1, \dots, B_n\}$  ενός πίνακα  $R$ ,

Η **συναρτησιακή εξάρτηση**  $A \rightarrow B$  στο  $R$  ισχύει, αν για κάθε  $t_i, t_j$  στο  $R$ :

Αν  $t_i[A_1] = t_j[A_1]$  AND  $t_i[A_2] = t_j[A_2]$   
AND ... AND  $t_i[A_m] = t_j[A_m]$

# Συναρτησιακές Εξαρτήσεις (εικόνα)



Αν t<sub>1</sub>, t<sub>2</sub> συμφωνούν εδώ.. ... συμφωνούν και εδώ!

Ορισμός:

Έστω δύο σύνολα γνωρισμάτων  $A = \{A_1, \dots, A_m\}$  και  $B = \{B_1, \dots, B_n\}$  ενός πίνακα  $R$ ,

Η **συναρτησιακή εξάρτηση**  $A \rightarrow B$  στο  $R$  ισχύει, αν για κάθε  $t_i, t_j$  στο  $R$ :

Αν  $t_i[A_1] = t_j[A_1]$  AND  $t_i[A_2] = t_j[A_2]$   
AND ... AND  $t_i[A_m] = t_j[A_m]$

**Τότε**  $t_i[B_1] = t_j[B_1]$  AND  $t_i[B_2] = t_j[B_2]$   
AND ... AND  $t_i[B_n] = t_j[B_n]$

# Συναρτησιακές Εξαρτήσεις

- Το  $Y$  εξαρτάται συναρτησιακά από το  $X$
- Γιατί καλούνται συναρτησιακές;
- $K \subseteq R$  κλειδί της  $R$  αν  $K \rightarrow ?$ 
  - Υπενθύμιση:  $R$  είναι το σύνολο των γνωρισμάτων του σχήματος
- Μια γενίκευση της έννοιας του κλειδιού

## Παρατήρηση

$$A_1A_2 \dots A_n \rightarrow B_1 \text{ και } A_1A_2 \dots A_n \rightarrow B_2 \Leftrightarrow A_1A_2 \dots A_n \rightarrow B_1B_2$$

# Παράδειγμα I (φυσική σημασία)

Όπως και τα κλειδιά, οι συναρτησιακές εξαρτήσεις προκύπτουν από τη φυσική περιγραφή του προβλήματος, δηλαδή από τον πραγματικό κόσμο

Έστω το παρακάτω σχεσιακό σχήμα:

Εγγραφή(Μάθημα, Φοιτητής, Ώρα&Μέρα, Αίθουσα, Βαθμός)

(συντομογραφία) **E(M, Φ, Ω, Α, Β)**

Ποιες συναρτησιακές εξαρτήσεις εκφράζουν τα 1 έως 4

1. Τα μαθήματα προσφέρονται μόνο μια φορά σε μια συγκεκριμένη ώρα&μέρα και αίθουσα.
2. Οι φοιτητές δεν μπορούν να είναι ταυτόχρονα (δηλαδή, την ίδια ώρα&μέρα) σε δυο διαφορετικές αίθουσες
3. Δε γίνεται να έχουμε δυο μαθήματα ταυτόχρονα (την ίδια ώρα&μέρα) στην ίδια αίθουσα
4. Ένας φοιτητής παίρνει μόνο ένα βαθμό σε κάθε μάθημα

Ποιο (ποια) είναι το κλειδί αν ισχύουν τα 1 έως 4

Τι σημαίνει

1.  $\Phi \rightarrow M$
2.  $MB \rightarrow \Phi$

# Παράδειγμα II (φυσική σημασία)

Παράδειγμα: Στο παρακάτω σχήμα **Λογαριασμός** θεωρούμε ότι ένας λογαριασμός μπορεί να ανήκει σε παραπάνω από έναν πελάτη και ένας πελάτης μπορεί να έχει πολλούς λογαριασμούς. Ποιες άλλες (εκτός του κλειδιού) συναρτησιακές εξαρτήσεις μπορεί να ισχύουν αλλά δε φαίνονται στο παρακάτω σχήμα;

## Λογαριασμός

Όνομα-Υποκαταστήματος	<u>Αριθμός-Λογαριασμού</u>	Ποσό	<u>Όνομα-Πελάτη</u>
-----------------------	----------------------------	------	---------------------

Παράδειγμα: Παρόμοια, στο παρακάτω σχήμα ένας **Πελάτης** και επίσης ένας πελάτης δίνει μόνο μια διεύθυνση

## Πελάτης

<u>Όνομα-Πελάτη</u>	Οδός	Πόλη	<u>Αριθμός-Λογαριασμού</u>
---------------------	------	------	----------------------------

{ Διεύθυνση πελάτη

- ✓ Στα παραπάνω σχεσιακά μοντέλα, με τα κλειδιά εκφράζεται μόνο ένα υποσύνολο των περιορισμών
- ✓ Διαισθητικά, οι δύο παραπάνω σχεδιασμοί δεν είναι «καλοί», γιατί;

# Τετριμμένη Συναρτησιακή Εξάρτηση

Τετριμμένες (trivial) εξαρτήσεις: ισχύουν για όλα τα σχήματα

Παράδειγμα:  $A \rightarrow A$  ή  $AB \rightarrow B$

Γενικά,

$X \rightarrow Y$  **τετριμμένη**, όταν  $Y \subseteq X$

# Περιορισμοί Σχήματος

- Οι συναρτησιακές εξαρτήσεις ορίζονται στο **σχήμα** μιας σχέσης, εκφράζουν περιορισμούς ορθότητας (integrity constraints)
- Ένα σύνολο από συναρτησιακές εξαρτήσεις  $F$  ισχύει σε ένα σχήμα, όλα τα έγκυρα (valid) στιγμιότυπα πρέπει να ικανοποιούν το σύνολο των εξαρτήσεων
- Έλεγχος αν μια σχέση (στιγμιότυπο) ικανοποιεί το σύνολο  $F$

# Κανόνες Συμπερασμού (Inference Rules)

Πως μπορούμε να συνάγουμε νέες εξαρτήσεις από ένα δεδομένο σύνολο εξαρτήσεων

$F \models X \rightarrow Y$  : η συναρτησιακή εξάρτηση  $X \rightarrow Y$  **συνάγεται** (inferred/implies) από το σύνολο εξαρτήσεων  $F$

Η  $X \rightarrow Y$  ισχύει σε κάθε στιγμιότυπο που ικανοποιεί το σύνολο των εξαρτήσεων στο  $F$

$F^+$ : **κλειστότητα (εγκλεισμός)** του  $F$  (closure): σύνολο όλων των συναρτησιακών εξαρτήσεων που συνάγονται από το  $F$

**Κανόνες Συμπερασμού** - για τη δημιουργία εξαρτήσεων



# Κανόνες Συμπερασμού (Inference Rules)

## 1. Ανακλαστικός Κανόνας (reflexivity)

Αν  $X \supseteq Y$ , τότε  $X \rightarrow Y$

## 2. Επαυξητικός Κανόνας (augmentation)

$\{X \rightarrow Y\} \quad | = \quad XZ \rightarrow YZ$

## 3. Μεταβατικός Κανόνας (transitivity)

$\{X \rightarrow Y, Y \rightarrow Z\} \quad | = \quad X \rightarrow Z$

**Κανόνες του Armstrong:** *βάσιμοι* (sound) δε δίνουν λανθασμένες εξαρτήσεις και *πλήρεις* (complete) μας δίνουν όλο το  $F^+$

# Κανόνες Συμπερασμού

$$\{X \rightarrow Y\} \quad | = \quad XZ \rightarrow YZ$$

## Επαυξητικός Κανόνας

Απόδειξη των 3 κανόνων με βάση τον ορισμό

Απόδειξη

(με επαγωγή σε άτοπο:) έστω ότι σε κάποιο στιγμιότυπο της  $r$  ισχύει

$$X \rightarrow Y \text{ (1) αλλά όχι } XZ \rightarrow YZ \text{ (2)}$$

Από (2 & ορισμό), υπάρχουν δυο πλειάδες,  $t_1$  και  $t_2$ , τέτοιες ώστε  $t_1[XZ] = t_2[XZ]$  (3)

$$\text{και } t_1[YZ] \neq t_2[YZ]$$

Από (3),  $t_1[X] = t_2[X]$  (4) και  $t_1[Z] = t_2[Z]$  (5)

Από (1) και (4),  $t_1[Y] = t_2[Y]$  (6)

Από (5) και (6),  $t_1[YZ] = t_2[YZ]$  Άτοπο!

# Κανόνες Συμπερασμού

## Επιπρόσθετοι κανόνες

### 4. Ενωτικός Κανόνας (union)

$$\{X \rightarrow Y, X \rightarrow Z\} \quad | = \quad X \rightarrow YZ$$

### 5. Διασπαστικός Κανόνας (decomposition)

$$\{X \rightarrow YZ\} \quad | = \quad X \rightarrow Y$$

### 6. Ψευδομεταβατικός Κανόνας

$$\{X \rightarrow Y, YZ \rightarrow W\} \quad | = \quad XZ \rightarrow W$$

# Κανόνες Συμπερασμού

## Ενωτικός Κανόνας

$$\{X \rightarrow Y (1), X \rightarrow Z (2)\} \quad | \quad = X \rightarrow YZ$$

Απόδειξη των επιπλέον κανόνων με βάση τον ορισμό ή/και των κανόνων του Amstrong

Απόδειξη (με χρήση των κανόνων του Amstrong)

$$(2) + \text{Επαυξ. } XY \rightarrow YZ \quad (3)$$

$$(1) + \text{Επαυξ. } X \rightarrow XY \quad (4)$$

$$(3) (4) \text{ Μεταβ. } X \rightarrow YZ$$

Ανακλαστικός Κανόνας

Αν  $X \supseteq Y$ , τότε  $X \rightarrow Y$

Επαυξητικός Κανόνας

$$\{X \rightarrow Y\} \quad | \quad = XZ \rightarrow YZ$$

Μεταβατικός Κανόνας

$$\{X \rightarrow Y, Y \rightarrow Z\} \quad | \quad = X \rightarrow Z$$

# Κανόνες Συμπερασμού (σύνοψη)

1. Ανακλαστικός Κανόνας  $\text{An } X \supseteq Y, \text{ τότε } X \rightarrow Y$
2. Επαυξητικός Κανόνας  $\{X \rightarrow Y\}$  συνάγει  $XZ \rightarrow YZ$
3. Μεταβατικός Κανόνας  $\{X \rightarrow Y, Y \rightarrow Z\}$  συνάγει  $X \rightarrow Z$
4. Ενωτικός Κανόνας  $\{X \rightarrow Y, X \rightarrow Z\}$  συνάγει  $X \rightarrow YZ$
5. Διασπαστικός Κανόνας  $\{X \rightarrow YZ\}$  συνάγει  $X \rightarrow Y$
6. Ψευδομεταβατικός Κανόνας  $\{X \rightarrow Y, YZ \rightarrow W\}$  συνάγει  $XZ \rightarrow W$

# Παράδειγμα

Έστω  $R = \{A, B, C, G, H, I\}$  και  $F = \{A \rightarrow B, A \rightarrow C, CG \rightarrow H, CG \rightarrow I, B \rightarrow H\}$

Παραδείγματα συναρτησιακών εξαρτήσεων που συνάγονται από το  $F$

- $A \rightarrow H$
- $CG \rightarrow HI$
- $AG \rightarrow I$

(α) Υπάρχει τρόπος/αλγόριθμος να τις υπολογίσουμε όλες;

(β) Πως μπορούμε να υπολογίσουμε το κλειδί;

# Εγκλεισμός Γνωρισμάτων

$X^+$  : κλειστότητα (εγκλεισμός) (closure) ενός συνόλου  $X$  από γνωρίσματα από το  $F$  : σύνολο όλων των γνωρισμάτων που εξαρτώνται συναρτησιακά από το  $X$  μέσω του  $F$

## Υπολογισμός του $X^+$

Result :=  $X$

while (αλλαγή στο Result)

    Για κάθε συναρτησιακή εξάρτηση:  $Y \rightarrow Z \in F$

        Αν  $Y \subseteq \text{Result}$ , Result := Result  $\cup$   $Z$

# Παράδειγμα

Έστω  $R = \{A, B, C, G, H, I\}$  και  $F = \{A \rightarrow B, A \rightarrow C, CG \rightarrow H, CG \rightarrow I, B \rightarrow H\}$

Υπολογισμός του  $\{A\}^+$ ,  $\{B\}^+$ ,  $\{A, G\}^+$



# Εγκλεισμός Γνωρισμάτων

- Είναι ο αλγόριθμος σωστός
  - (α) Για κάθε  $Y \in \text{Result}$ , ισχύει  $Y \in X^+$
  - (β) Για κάθε  $Y \in X^+$ , ισχύει  $Y \in \text{Result}$
- Πολυπλοκότητα χειρότερης περίπτωσης

# Παράδειγμα I

$R(A, B, C, D) \quad F = \{AB \rightarrow C, C \rightarrow D, D \rightarrow A\}$

1. Δείξουμε αν μια συναρτησιακή εξάρτηση ισχύει (συνάγεται από την  $F$ )

$C \rightarrow A ?$

$A \rightarrow D ?$

$AB \rightarrow D ?$

2. Υπολογισμός κλειδιών

3. Υπολογίσουμε το  $F^+$

# Άσκηση

$R(A, B, C, D, E) \quad F = \{A \rightarrow B, C \rightarrow E, AD \rightarrow E\}$

1. Ισχύει  $DC \rightarrow E$  ?
2. Υπολογίστε τα  $A^+, B^+, C^+, D^+, E^+$
3. Υποψήφια κλειδιά;
4. Δώστε ένα στιγμιότυπο που να παραβιάζει **μόνο** την  $AD \rightarrow E$

# Παράδειγμα II

$R(A, B, C, D, E) \quad F = \{A \rightarrow BC, C \rightarrow AD, B \rightarrow ED, AD \rightarrow E\}$

1. Υπολογίστε το

$A^+, B^+, C^+, D^+, E^+$

2. Υποψήφια κλειδιά;

# Παράδειγμα II

$R(A, B, C, D, E) \quad F = \{A \rightarrow BC, C \rightarrow AD, B \rightarrow ED, AD \rightarrow E\}$

1. Υπολογίστε το

$A^+, B^+, C^+, D^+, E^+$

2. Υποψήφια κλειδιά;

# Κάλυμμα

**Στόχος η απλοποίηση** ενός συνόλου συναρτησιακών εξαρτήσεων χωρίς να μεταβάλλουμε την κλειστότητά του

Έστω δυο σύνολα συναρτησιακών εξαρτήσεων  $E$  και  $F$

Λέμε ότι το  $F$  **καλύπτει** το  $E$  (ή το  $E$  καλύπτεται από το  $F$ ), **αν κάθε  $\Sigma E$  στο  $E$  ανήκει στο  $F^+$**  (δηλαδή, συνάγεται από το  $F$ ) (ισοδύναμα, αν  $E \subseteq F^+$ )

Δυο σύνολα συναρτησιακών εξαρτήσεων  $E$  και  $F$  είναι **ισοδύναμα**

ανν  $E^+ = F^+$ .

(δηλαδή, αν το  $E$  καλύπτει το  $F$  και το  $F$  καλύπτει το  $E$ )

# Κάλυμμα

- Πως μπορούμε να υπολογίσουμε αν ένα σύνολο  $F$  καλύπτει ένα σύνολο  $E$ ;
- Πως μπορούμε να υπολογίσουμε αν ένα σύνολο  $F$  είναι ισοδύναμο με ένα σύνολο  $E$ ;

# Παράδειγμα

$$F1 = \{A \rightarrow C, B \rightarrow C\}$$

$$F2 = \{A \rightarrow B, A \rightarrow C\}$$

$$F3 = \{A \rightarrow B, AB \rightarrow C\}$$

F1 καλύπτει το F3;

F3 καλύπτει το F1;

F1 ισοδύναμο του F3;

F2 καλύπτει το F1;

F2 καλύπτει το F3;

F2 ισοδύναμο με το F3;



# Ελάχιστο Κάλυμμα

Ένα σύνολο  $F$  συναρτησιακών εξαρτήσεων είναι **ελάχιστο** αν:

1. κάθε ΣΕ στο  $F$  έχει *ένα μόνο γνώρισμα* στο δεξιό της μέρος
2. δε μπορούμε να αντικαταστήσουμε μια ΣΕ  $X \rightarrow Z$  στο  $F$  με μια ΣΕ  $Y \rightarrow Z$  τέτοια ώστε  $Y \subset X$  και να πάρουμε ένα σύνολο ισοδύναμο του  $F$  (δηλαδή, δεν υπάρχει *περιττό γνώρισμα* στο α.μ της συναρτησιακής εξάρτησης)
3. δε *μπορούμε να αφαιρέσουμε* μια ΣΕ από το  $F$  και να πάρουμε ένα σύνολο ισοδύναμο του  $F$  (δηλαδή, δεν υπάρχει περιττή ΣΕ)

**Ελάχιστο κάλυμμα**  $F_{\min}$  της  $F$ : ελάχιστο σύνολο από ΣΕ που είναι ισοδύναμο με την  $F$

# Αλγόριθμος Υπολογισμού Ελάχιστου Καλύμματος

1. Αντικατέστησε τις συναρτησιακές εξαρτήσεις

$$X_1 \rightarrow Y_1 Y_2 \text{ με } X_1 \rightarrow Y_1 \text{ και } X_1 \rightarrow Y_2.$$

2. Για κάθε ΣΕ

(i) Βρες τα περιττά γνωρίσματα στο α.μ., αφάιρεσε τα

(ii) Έλεγξε αν είναι περιττή, αν ναι αφάιρεσέ τη

# Περισσότερα Γνωρίσματα

**Περισσότερα γνωρίσματα:** γνωρίσματα που αν αφαιρεθούν δεν επηρεάζουν τη κλειστότητα (δηλαδή προκύπτει ισοδύναμο σύνολο)

Για παράδειγμα: το γνώρισμα  $AB \rightarrow C$  το **A είναι περιττό** στην εξάρτηση (στο αριστερό μέρος) αν

$$F \text{ ισοδύναμο } \underbrace{(F - \{AB \rightarrow C\}) \cup \{B \rightarrow C\}}_{F'}$$

*Δηλαδή, αν αφαιρέσουμε το A από την ΣΕ, το σύνολο  $F'$  που προκύπτει είναι ισοδύναμο με το αρχικό σύνολο  $F$*

- ✓ Προφανώς το  $F'$  καλύπτει το  $F$ , άρα **αρκεί να ελέγξουμε αν το  $F$  καλύπτει το  $F'$**

# Περισσότερο Γνώρισμα

Έστω ένα σύνολο  $F$  συναρτησιακών εξαρτήσεων και η ΣΕ  $X \rightarrow Y \in F$

Το γνώρισμα  $A \in X$  είναι **περισσότερο στο  $X$**  αν

$F$  καλύπτει το  $(F - \{X \rightarrow Y\}) \cup \{(X - A) \rightarrow Y\}$

- Πως θα υπολογίσουμε αν ένα γνώρισμα στο **α.μ.** μιας ΣΕ είναι περισσότερο; Θα πρέπει να δείξουμε ότι οι ΣΕ του  $F'$  ανήκουν στο  $F^+$ , δηλαδή:

Υπολόγισε το  $(X - \{A\})^+$  με βάση τις ΣΕ του συνόλου  $F$ :

Το  $A$  είναι περισσότερο αν το  $Y$  ανήκει στο  $(X - \{A\})^+$  στο  $F$

# Παράδειγμα (περιττού γνωρίσματος στο αριστερό μέρος)

$R(A, B, C, D, E) \ F = \{A \rightarrow BC, C \rightarrow AD, B \rightarrow ED, AD \rightarrow E\}$

Είναι κάποιο γνώρισμα στο α.μ. της  $AD \rightarrow E$  περιττό;

# Περιττή Εξάρτηση

- Πως θα υπολογίσουμε αν μια ΣΕ  $X \rightarrow B$  (με ένα γνώρισμα στο δ.μ.) είναι περιττή;

Υπολογίζουμε το  $(X)^+$  χρησιμοποιώντας το  $F - \{X \rightarrow B\}$

Περιττό αν το  $B$  ανήκει στο  $(X)^+$

# Αλγόριθμος Υπολογισμού Ελάχιστου Καλύμματος

1. Αντικατέστησε τις συναρτησιακές εξαρτήσεις

$$X_1 \rightarrow Y_1 Y_2 \text{ με } X_1 \rightarrow Y_1 \text{ και } X_1 \rightarrow Y_2.$$

2. Για κάθε ΣΕ

(i) Βρες τα περιττά γνωρίσματα στο α.μ.

Α περιττό στο  $X (X \rightarrow Y)$ : υπολόγισε το  $(X - \{A\})^+$

(ii) Έλεγξε αν είναι περιττή, αν ναι αφάιρεσε τη

Εξάρτηση  $X \rightarrow B$  περιττή: υπολόγισε το  $X^+$

# Παράδειγμα

Έστω  $R(A, B, C)$  και  $F = \{A \rightarrow BC, B \rightarrow C, A \rightarrow B, AB \rightarrow C\}$ .

Βρείτε το  $F_{\min}$ .



# Παράδειγμα

Έστω  $R(A, B, C)$  και  $F = \{A \rightarrow BC, B \rightarrow C, A \rightarrow B, AB \rightarrow C\}$ . Βρείτε το  $F_{\min}$ .

Μετά το βήμα 1:  $\{A \rightarrow B, A \rightarrow C, B \rightarrow C, A \rightarrow B, AB \rightarrow C\}$

Βήμα 2: Εξέταση αν το  $A$  είναι περιττό στο  $AB \rightarrow C$ , υπολογίζοντας το  $(B)^+$

*είναι περιττό*

*Νέο σύνολο:  $\{A \rightarrow B, A \rightarrow C, B \rightarrow C, B \rightarrow C\}$*

Βήμα 3: Εξέταση αν η ΣΕ  $A \rightarrow B$  είναι περιττή όχι

Εξέταση αν η ΣΕ  $A \rightarrow C$  είναι περιττή ναι

*Νέο σύνολο:  $\{A \rightarrow B, B \rightarrow C\}$*

Εξέταση αν η ΣΕ  $B \rightarrow C$  είναι περιττή όχι

Αποτέλεσμα:  $\{A \rightarrow B, B \rightarrow C\}$

# Ελάχιστο Κάλυμμα

## Παρατηρήσεις

- Το ελάχιστο κάλυμμα *δεν* είναι μοναδικό
- Το βήμα (i) πρέπει να προηγηθεί του βήματος (ii), δηλαδή πρέπει πρώτα να βρούμε τα περιττά γνωρίσματα στο α.μ. και μετά τις περιττές εξαρτήσεις

# Άσκηση

$R(A, B, C, D, E) \quad F = \{A \rightarrow B, C \rightarrow E, AD \rightarrow E\}$

1. Ισχύει  $DC \rightarrow E$  ?
2. Υποψήφια κλειδιά;
3. Δώστε ένα στιγμιότυπο που να παραβιάζει **μόνο** την  $AD \rightarrow E$

# Άσκηση

$R(A, B, C, D, E) \quad F = \{A \rightarrow BC, C \rightarrow AD, B \rightarrow ED, AD \rightarrow E\}$

Ποιο είναι το ελάχιστο κάλυμμα της  $F$ ;

# Σύνοψη

## Τι είδαμε:

- Ορισμό συναρτησιακής εξάρτησης
- Κανόνες συμπερασμού συναρτησιακών εξαρτήσεων
- Κλειστότητα γνωρίσματος
- Ισοδυναμία συνόλου εξαρτήσεων
- Ελάχιστο κάλυμμα

# Ερωτήσεις;

# Διάσπαση

# Αλγόριθμος Σχεδιασμού

Ο τρόπος που σχεδιάζαμε ένα σχήμα ΒΔ μέχρι τώρα:

από το εννοιολογικό στο σχεσιακό μοντέλο

Θα δούμε ένα **γενικό θεωρητικό (formal) τρόπο** κατασκευής του σχήματος

Γενικά:

- Ξεκινάμε από το **καθολικό σχήμα** (που περιέχει όλα τα γνωρίσματα)
- Διαδοχικές διασπάσεις έτσι ώστε τα σχήματα που προκύπτουν να ικανοποιούν κάποιες ιδιότητες (να είναι σε κάποια κανονική μορφή)

-- top-down τεχνική



# Αλγόριθμος Σχεδιασμού

Μπορούμε να εφαρμόσουμε αυτόν τον τρόπο και για να «διασπάσουμε» μια «προβληματική» σχέση που προέκυψε από την μετατροπή του εννοιολογικού σχεδιασμού

Μειονέκτημα των διασπάσεων:

μπορεί να απαιτεί συνενώσεις (join) για να απαντηθούν ερωτήματα ή για να ελεγχθούν εξαρτήσεις -> αποδοτικότητα του συστήματος

# Αλγόριθμος Σχεδιασμού

Αποσύνθεση/Διάσπαση (decomposition)

## Αλγόριθμος σχεδιασμού

1. Αρχικά ένα καθολικό (universal) σχήμα σχέσης που περιέχει όλα τα γνωρίσματα
2. Προσδιορισμός των συναρτησιακών εξαρτήσεων
3. Διάσπαση σε ένα σύνολο από σχήματα σχέσεων που ικανοποιούν κάποιες ιδιότητες

# Παράδειγμα Διάσπασης Σχήματος

*Καθολικό Σχήμα:*  $R = \{\text{Τίτλος, Έτος, Διάρκεια, Είδος, Όνομα-Ηθοποιού, Διεύθυνση, Έτος-Γέννησης}\}$

Σύνολο ΣΕ που ισχύουν στο πρόβλημα:

$F = \{\text{Τίτλος Έτος} \rightarrow \text{Είδος Διάρκεια, Όνομα-Ηθοποιού} \rightarrow \text{Διεύθυνση Έτος-Γέννησης}\}$

Πιθανή διάσπαση:

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

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

- ✓ Ποια είναι μια καλή διάσπαση; Πως μπορούμε να πάρουμε την αρχική σχέση;
- ✓ Μπορούμε να διασπάσουμε την  $R_2$  με τον ίδιο τρόπο.

# Επιθυμητές Ιδιότητες Διάσπασης

1. Διασπάσεις χωρίς απώλειες στη συνένωση
2. Διατήρηση των συναρτησιακών εξαρτήσεων
3. Οι σχέσεις που προκύπτουν σε κάποια κανονική μορφή

# Τυπικός Ορισμός Διάσπασης

Αρχικά ένα καθολικό σχήμα  $R = \{A_1, A_2, \dots, A_n\}$ .

Διάσπαση/αποσύνθεση (decomposition) του σε δύο σχήματα

$$R_1 = \{B_1, B_2, \dots, B_m\} \text{ και } R_2 = \{C_1, C_2, \dots, C_k\}$$

τέτοια ώστε:

1.  $\{A_1, A_2, \dots, A_n\} = \{B_1, B_2, \dots, B_m\} \cup \{C_1, C_2, \dots, C_k\}$  (διατήρηση γνωρισμάτων)  
*[γνωρίσματα]*
2. Οι πλειάδες της  $r_1(R_1)$  είναι η *προβολή των πλειάδων της  $r(R)$  στα  $\{B_1, B_2, \dots, B_m\}$*   
*[πλειάδες]*
3. Οι πλειάδες της  $r_2(R_2)$  είναι η *προβολή των πλειάδων της  $r(R)$  στα  $\{C_1, C_2, \dots, C_k\}$*   
*[πλειάδες]*

# Παράδειγμα

Έστω το (καθολικό) σχήμα  $R(A, B, C)$  και διάσπαση σε  $R_1(A, B)$  και  $R_2(B, C)$

Τα αντίστοιχα στιγμιότυπα (σχέσεις) (συμβολισμός  $r(R)$  ή  $r$ )

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

*Μπορούμε να πάρουμε το αρχικό στιγμιότυπο;*

Φυσική συνένωση  $r_1 * r_2$ ;

# Διάσπαση

γνωρίσματα

Έστω ένα σχεσιακό σχήμα  $R$ . Ένα σύνολο από σχεσιακά σχήματα  $\{R_1, R_2, \dots, R_n\}$  είναι μια **διάσπαση** του  $R$  αν

$$R = R_1 \cup R_2 \dots \cup R_n$$

Δηλαδή,  $\forall i = 1, \dots, n \quad R_i \subseteq R$

στιγμιότυπα

Έστω  $r(R)$  και  $r_i = \pi_{R_i}(r)$ ,  $\forall i = 1, \dots, n$

$$r \subseteq r_1 * r_2 * \dots * r_n$$

# Διάσπαση Άνευ Απωλειών στη Συνένωση

Έστω  $C$  το σύνολο περιορισμών στην  $R$ . Μια διάσπαση του  $R$  σε  $\{R_1, R_2, \dots, R_n\}$  είναι μια **διάσπαση άνευ απωλειών στη συνένωση (lossless join decomposition)** αν για όλες τις σχέσεις  $r(R)$  που είναι νόμιμες στο  $C$  ισχύει

$$r = \pi_{R_1}(r) * \pi_{R_2}(r) * \dots * \pi_{R_n}(r)$$

Ονομάζεται και **μη προσθετική συνένωση (non-additive join)**



# Παράδειγμα

<b>r</b>	<table style="border-collapse: collapse; text-align: center;"> <tr><td>A</td><td>B</td><td>C</td></tr> <tr><td>1</td><td>2</td><td>3</td></tr> <tr><td>4</td><td>2</td><td>5</td></tr> </table>	A	B	C	1	2	3	4	2	5	<b>r<sub>1</sub></b>	<table style="border-collapse: collapse; text-align: center;"> <tr><td>A</td><td>B</td></tr> <tr><td>1</td><td>2</td></tr> <tr><td>4</td><td>2</td></tr> </table>	A	B	1	2	4	2	<b>r<sub>1</sub> * r<sub>2</sub></b>	<table style="border-collapse: collapse; text-align: center;"> <tr><td>A</td><td>B</td><td>C</td></tr> <tr><td>1</td><td>2</td><td>3</td></tr> <tr><td>1</td><td>2</td><td>5</td></tr> <tr><td>4</td><td>2</td><td>3</td></tr> <tr><td>4</td><td>2</td><td>5</td></tr> </table>	A	B	C	1	2	3	1	2	5	4	2	3	4	2	5	$R_1 \cap R_2 = B$
A	B	C																																		
1	2	3																																		
4	2	5																																		
A	B																																			
1	2																																			
4	2																																			
A	B	C																																		
1	2	3																																		
1	2	5																																		
4	2	3																																		
4	2	5																																		
		<b>r<sub>2</sub></b>	<table style="border-collapse: collapse; text-align: center;"> <tr><td>B</td><td>C</td></tr> <tr><td>2</td><td>3</td></tr> <tr><td>2</td><td>5</td></tr> </table>	B	C	2	3	2	5																											
B	C																																			
2	3																																			
2	5																																			

✓ Δεν μπορούμε να πάρουμε την αρχική σχέση  $r$  από τα  $r_1$  και  $r_2$

<b>r<sub>1</sub></b>	<table style="border-collapse: collapse; text-align: center;"> <tr><td>A</td><td>C</td></tr> <tr><td>1</td><td>3</td></tr> <tr><td>4</td><td>5</td></tr> </table>	A	C	1	3	4	5	<b>r<sub>2</sub></b>	<table style="border-collapse: collapse; text-align: center;"> <tr><td>B</td><td>C</td></tr> <tr><td>2</td><td>3</td></tr> <tr><td>2</td><td>5</td></tr> </table>	B	C	2	3	2	5	$R_1 \cap R_2 = C$
A	C															
1	3															
4	5															
B	C															
2	3															
2	5															

# Διάσπαση Άνευ Απωλειών στη Συνένωση

## Θεώρημα

Έστω  $R$  ένα σχεσιακό σχήμα και  $F$  ένα σύνολο από συναρτησιακές εξαρτήσεις στο  $R$ . Έστω  $R_1$  και  $R_2$  μια διάσπαση του  $R$ . Αν μια τουλάχιστον από τις ΣΕ

$$R_1 \cap R_2 \rightarrow R_1 \text{ ή } R_1 \cap R_2 \rightarrow R_2 \text{ ανήκει στο } F^+$$

τότε η διάσπαση είναι χωρίς απώλειες στη συνένωση.

Δηλαδή τα κοινά γνωρίσματα των δύο σχημάτων είναι κλειδί για τουλάχιστον ένα από τα δύο σχήματα

# Παράδειγμα

Παράδειγμα:  $R = \{\text{Τίτλος, Έτος, Διάρκεια, Είδος, Όνομα-Ηθοποιού, Διεύθυνση, Έτος-Γέννησης}\}$

Τίτλος Έτος  $\rightarrow$  Διάρκεια

Τίτλος Έτος  $\rightarrow$  Είδος

Όνομα Ηθοποιού  $\rightarrow$  Διεύθυνση

Όνομα-Ηθοποιού  $\rightarrow$  Έτος Γέννησης

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

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

$$R_1 \cap R_2 = \{\text{Τίτλος, Έτος}\}$$

# Διατήρηση Εξαρτήσεων

## Στόχος

Για να ελέγξουμε ότι διατηρούνται οι Σ.Ε. στο αρχικό σχήμα, όταν γίνονται τροποποιήσεις σε μία από τις σχέσεις  $r_i(R_i)$ , να αρκεί να ελέγξουμε μόνο τη συγκεκριμένη σχέση (δηλαδή, να μη χρειάζεται να υπολογίσουμε την αρχική σχέση - αποφυγή των συνενώσεων)

Έστω  $F$  ένα σύνολο από ΣΕ στο σχήμα  $R$  και  $\{R_1, R_2, \dots, R_n\}$  μια διάσπαση του  $R$ .

$F_i$  περιορισμός (ή προβολή) του  $F$  στο  $R_i$  είναι το σύνολο όλων των συναρτησιακών εξαρτήσεων του  $F^+$  που περιέχουν μόνο γνωρίσματα του  $R_i$ .

Προσοχή:  $F^+$  όχι  $F$

# Παραδείγματα

**Παράδειγμα 1:** Έστω  $R(A, B, C, D)$ ,  $F = \{A \rightarrow B, B \rightarrow C\}$ .

Περιορισμός του  $F$  στο  $S(A, C)$  (δηλαδή ποιες  $\Sigma E$  του  $F^+$  ισχύουν στο  $S$ )

**Παράδειγμα 2:** Έστω  $R(A, B, C, D, E)$ ,  $F = \{A \rightarrow D, B \rightarrow E, DE \rightarrow C, C \rightarrow D\}$ .

Περιορισμός του  $F$  στο  $S(A, B, C)$

# Διατήρηση Εξαρτήσεων

Έστω  $F$  ένα σύνολο από ΣΕ στο σχήμα  $R$  και  $\{R_1, R_2, \dots, R_n\}$  μια διάσπαση του  $R$  και  $F_i$  η προβολή (περιορισμός της  $F$  στο  $R_i$ ).

$$\text{Έστω } F' = F_1 \cup F_2 \dots \cup F_n$$

Μια διάσπαση είναι μια **διάσπαση που διατηρεί τις εξαρτήσεις** (dependency preserving) αν  $F'^+ = F^+$

# Παραδείγματα

Έστω  $R(A, B, C, D)$ ,  $F = \{A \rightarrow C, B \rightarrow C, BD \rightarrow A\}$  και η διάσπαση του  $R$  σε  $R_1(A, C)$  και  $R_2(A, B, D)$ .

(α) Είναι χωρίς απώλειες (lossless join);

(β) Διατηρεί τις εξαρτήσεις;

Έστω  $R(A, B, C, D, E)$ ,  $F = \{A \rightarrow D, B \rightarrow E, DE \rightarrow C, B \rightarrow C\}$ .

(α) Η διάσπαση του  $R$  σε  $S(A, B, C)$  και  $T(A, B, D, E)$  διατηρεί τις εξαρτήσεις;

(β) Είναι χωρίς απώλειες (lossless join);

# Σχεδιασμός

Διάσπαση καθολικού σχήματος

Επιθυμητές ιδιότητες

1. όχι απώλειες στη συνένωση
2. διατήρηση εξαρτήσεων
3. όχι επανάληψη πληροφορίας λόγω ΣΕ



Κανονικές μορφές



# Ερωτήσεις;

# Κανονικοποίηση Σχήματος

# Επιθυμητές Ιδιότητες (ανασκόπηση)

## 1. Συνενώσεις Άνευ Απωλειών

Η φυσική συνένωση των σχέσεων που προκύπτουν μας δίνει ακριβώς την αρχική σχέση (χωρίς επιπρόσθετες πλειάδες):  $r = \pi_{R_1}(r) * \pi_{R_2}(r) * \dots * \pi_{R_n}(r)$

$R_1 \cap R_2 \rightarrow R_1$  ή  $R_1 \cap R_2 \rightarrow R_2$  ανήκει στο  $F^+$ , δηλαδή τα κοινά γνωρίσματα των δύο σχημάτων είναι κλειδί για τουλάχιστον ένα από τα δύο

## 2. Διατήρηση Εξαρτήσεων

Στόχος: Έλεγχος διατήρησης εξαρτήσεων όταν γίνονται τροποποιήσεις χωρίς να υπολογίζουμε τις αρχικές σχέσεις (αποφυγή των συνενώσεων)

$F' = F_1 \cup F_2 \dots \cup F_n$ , πρέπει  $F'^+ = F^+$

## 3. Αποφυγή Επανάληψης Πληροφορίας

πως; Οι σχέσεις που προκύπτουν σε κανονική μορφή

# Κανονικές Μορφές

**Στόχος:** Δοθέντος 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**

- Οι ορισμοί των κανονικών μορφών βασίζονται σε Σ.Ε., οι Σ.Ε. έχουν σχέση με την επανάληψη πληροφορίας

# BCNF

Ένα σχεσιακό σχήμα R είναι σε **Κανονική Μορφή Boyce-Codd (BCNF)** σε σχέση με ένα σύνολο F συναρτησιακών εξαρτήσεων αν για όλες τις μη τετριμμένες ΣΕ,  $X \rightarrow Y$ , στο  $F^+$  ισχύει:

-- X είναι **υπερκλειδί (δηλαδή υποψήφιο κλειδί ή υπερσύνολο υποψήφιου κλειδιού)** 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.

# Κανονικές Μορφές

Πλεονασμός (επανάληψη πληροφορίας)

## Ταινία

<u>Τίτλος</u>	<u>Έτος</u>	Διάρκεια	Είδος	<u>Όνομα-Ηθοποιού</u>
---------------	-------------	----------	-------	-----------------------

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

# Αλγόριθμος Διάσπασης σε BCNF

- Βρες μια ΣΕ που παραβιάζει τον BCNF ορισμό,

έστω  $X \rightarrow Y$  και  $X \cap Y = \emptyset$

- Διάσπαση 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*

*Διάσπαση χωρίς απώλειες;*

# Παραδείγματα

## Παράδειγμα 1

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

Τίτλος Έτος → Διάρκεια Είδος

Ταινία1(Τίτλος, Έτος, Διάρκεια, Είδος)

Ταινία2(Τίτλος, Έτος, Όνομα-Ηθοποιού)

## Παράδειγμα 2

Δείξτε ότι οποιαδήποτε σχέση με δύο γνωρίσματα είναι σε BCNF



# BCNF

Μπορεί να χρειαστεί παραπάνω από μία διάσπαση

Διάσπαση του αρχικού σχήματος R σε δύο σχήματα

- $R_1$  με γνωρίσματα  $X \cup Y$  και
- $R_2$  με γνωρίσματα  $R - Y$

Διάσπαση της  $R_2$  αν χρειάζεται

Πιθανών συνεχείς διασπάσεις

Αφού παίρνουμε σχέσεις με αυστηρά μικρότερο αριθμό γνωρισμάτων, η διαδικασία τερματίζει (στη χειρότερη περίπτωση όταν 2 γνωρίσματα)

# BCNF

Παραβίαση του BCNF σημαίνει ότι υπάρχει  $X \rightarrow A$  όπου το  $X$  δεν είναι υπερκλειδί

**Περίπτωση 1:**  $X$  είναι γνήσιο υποσύνολο κάποιου υποψήφιου κλειδιού  
(μερική εξάρτηση)

# Παραδείγματα Μερικής Εξάρτησης

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

## Λογαριασμός

Όνομα-Υποκαταστήματος	<u>Αριθμός-Λογαριασμού</u>	Ποσό	<u>Όνομα-Πελάτη</u>
-----------------------	----------------------------	------	---------------------

## Πελάτης

<u>Όνομα-Πελάτη</u>	Οδός	Πόλη	<u>Αριθμός-Δανείου</u>
---------------------	------	------	------------------------



Διεύθυνση πελάτη

# BCNF

**Περίπτωση 2:** X δεν είναι γνήσιο υποσύνολο κάποιου υποψήφιου κλειδιού

Τότε έστω K (υποψήφιο κλειδί)

$K \rightarrow X \rightarrow A$  (μεταβατική εξάρτηση)

- Δε μπορώ να συνδυάσω μια τιμή του X με μια τιμή του K χωρίς να συνδυάσω μια τιμή A με μια τιμή X
- Δε μπορώ να εισάγω τιμή του X, χωρίς να ξέρω και το «σωστό» A

# Παράδειγμα Μεταβατικής Εξάρτησης

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

Ισχύει

Εταιρεία-Παραγωγής → Διεύθυνση-Εταιρείας

Πρόβλημα: υπάρχει μια *μεταβατική* εξάρτηση

Τίτλος Έτος → Εταιρεία-Παραγωγής

Τίτλος Έτος → Διεύθυνση-Εταιρείας

*Για να αντιστοιχήσουμε μια ταινία σε εταιρεία πρέπει να ξέρουμε τη διεύθυνση!*

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

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

# BCNF Διάσπαση και Διατήρηση Εξαρτήσεων

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

Παράδειγμα

Έστω η σχέση Παίζει(Έργο, Κινηματογράφος, Πόλη) με τους περιορισμούς:

Κινηματογράφος → Πόλη

Έργο Πόλη → Κινηματογράφος

Υποψήφια Κλειδιά;

{Έργο, Πόλη} {Κινηματογράφος, Έργο}

# BCNF Διάσπαση και Διατήρηση Εξαρτήσεων

Παίζει(Έργο, Κινηματογράφος, Πόλη)  $F = \{ \text{Κινηματογράφος} \rightarrow \text{Πόλη}, \text{Έργο} \text{ Πόλη} \rightarrow \text{Κινηματογράφος} \}$

Υποψήφια Κλειδιά {Έργο, Πόλη} {Κινηματογράφος, Έργο}

Διάσπαση σε:  $R_1\{\text{Κινηματογράφος}, \text{Πόλη}\}$  και  $R_2\{\text{Κινηματογράφος}, \text{Έργο}\}$

Διατηρεί τις εξαρτήσεις;

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

Δε μπορώ κοιτάζοντας μόνο την  $R_2$  (ή την  $R_1$ ) να δω ότι η εισαγωγή της δεύτερης πλειάδας παραβιάζει μια ΣΕ (πρέπει να κάνω συνένωση!)

# 3NF

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

- $X$  είναι υπερκλειδί του σχήματος  $R$ ,
- κάθε γνώρισμα  $A$  του  $Y - X$  περιέχεται σε κάποιο υποψήφιο κλειδί

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

Δηλαδή, επιπρόσθετα επιτρέπει συναρτησιακές εξαρτήσεις που στο δ.μ. έχουν πρωτεύοντα γνωρίσματα

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



# Παράδειγμα

Παίζει(Έργο, Κινηματογράφος, Πόλη)

$F = \{\text{Κινηματογράφος} \rightarrow \text{Πόλη}, \text{Έργο Πόλη} \rightarrow \text{Κινηματογράφος}\}$

Υποψήφια Κλειδιά  $\{\text{Έργο, Πόλη}\} \{\text{Κινηματογράφος, Έργο}\}$

Η σχέση είναι σε 3NF

# Παράδειγμα

Παίζει(Έργο, Κινηματογράφος, Πόλη)

$F = \{\text{Κινηματογράφος} \rightarrow \text{Πόλη}, \text{Έργο Πόλη} \rightarrow \text{Κινηματογράφος}\}$

Υποψήφια Κλειδιά {Έργο, Πόλη} {Κινηματογράφος, Έργο}

Η σχέση είναι σε 3NF

Υπάρχει μια μεταβατική εξάρτηση

Αλλά απαιτούμε να είναι σε πρωτεύον γνώρισμα

# Διάσπαση σε 3NF

## Αλγόριθμος (Από) σύνθεσης σε 3NF

1. Υπολόγισε το ελάχιστο κάλυμμα  $F_c$  του  $F$  της αρχικής  $R$
2. 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  $X \cup Y$   
(δηλαδή, το  $X$  και όλα τα γνωρίσματα που εξαρτώνται από το  $X$ )
3. Αν κανένα από τα σχήματα που δημιουργούνται δεν περιέχει κλειδί της  $R$ , δημιούργησε ένα σχήμα σχέσης που να περιέχει τα γνωρίσματα που σχηματίζουν κλειδί (όχι απώλεια)

# 3NF

## Αλγόριθμος Αποσύνθεσης σε 3NF

- Απώλειες στη συνένωση;
- Διατήρηση εξαρτήσεων;

# Κανονικές Μορφές (επανάληψη)

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

# Σχεδιασμός Σχεσιακών Σχημάτων

- Η διαδικασία κανονικοποίησης έχει και μειονεκτήματα:
  - ο Δεν είναι δημιουργική
  - ο Συνήθως η κανονικοποίηση γίνεται αφού έχουμε κάποιο σχήμα (μας λέει αν είναι «καλό» ή «κακό»)
  - ο 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 διαισθητικά.

# Σχεδιασμός Σχεσιακών Σχημάτων

- Υπάρχουν **εμπορικά εργαλεία**, που με είσοδο ένα σύνολο σχημάτων σχέσεων/γνωρισμάτων και ένα σύνολο 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:
  - Διάφορα θεωρητικά αποτελέσματα, πχ: Αν μια σχέση είναι σε *3NF* και κάθε υποψήφιο κλειδί αποτελείται ακριβώς από ένα γνώρισμα, τότε είναι και σε *5NF* (Fagin, 1991)

# Ερωτήσεις;



# Ασκήσεις

1. Έστω το σχεσιακό σχήμα  $R(A, B, C, D)$  στο οποίο ισχύει μόνο η συναρτησιακή εξάρτηση  $A \rightarrow B$ .

(i) Δώστε το υποψήφια (υποψήφιο) κλειδιά.

(ii) Είναι σε BCNF ή όχι και γιατί. Αν όχι διασπάστε τη σχέση σε σχέσεις που να είναι σε BCNF (χρησιμοποιώντας το σχετικό αλγόριθμο) και δώστε τα το υποψήφια (υποψήφιο) κλειδιά για κάθε μια από τις σχέσεις που προκύπτουν.

2. Έστω μια σχεσιακή βάση με σχήμα  $S(E, F, G)$  και το στιγμιότυπο με 2 πλειάδες:  $\{(123, \text{smith}, \text{main-street}), (123, \text{johanson}, \text{forbes})\}$ . Για κάθε μία από τις συναρτησιακές εξαρτήσεις (i)-(iii) παρακάτω εξηγήστε αν μπορείτε ή όχι να πείτε αν ισχύουν ή όχι.

(i)  $E \rightarrow F$

(ii)  $EF \rightarrow G$

(iii)  $F \rightarrow G$

# Ασκήσεις

3. Έστω ότι στο σχεσιακό σχήμα  $R = (P, Q, S, T, U, V)$  ισχύει το σύνολο των συναρτησιακών εξαρτήσεων  $F = \{Q \rightarrow ST, P \rightarrow T, PS \rightarrow T, QU \rightarrow V\}$ .

- (i) Υπάρχει κάποια εξάρτηση που είναι περιττή. Εξηγείστε.
- (ii) Ισχύει ή όχι  $Q \rightarrow S$ .
- (iii) Είναι το  $\{Q, P\}$  κλειδί ή όχι;
- (iv) Είναι το  $\{Q, P, V, U\}$  υποψήφιο κλειδί ή όχι;

Εκτός ύλης

# Πλειότιμες Εξαρτήσεις

Υπάρχει επανάληψη πληροφορίας που δεν μπορεί να εκφραστεί με απλές ΣΕ

# Πλειότιμες Εξαρτήσεις

- Προκύπτουν όταν δυο γνωρίσματα είναι ανεξάρτητα το ένα από το άλλο

## Παράδειγμα

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

Ας υποθέσουμε ότι για κάθε ηθοποιό είναι πιθανόν να υπάρχουν πολλές διευθύνσεις (πλειότιμο γνώρισμα), και ένα ηθοποιός παίζει σε πολλές ταινίες

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

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

# Πλειότιμες Εξαρτήσεις

*Παράδειγμα (συνέχεια)*

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

Όλες οι εξαρτήσεις είναι τετριμμένες

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

# Παράδειγμα

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

Όνομα →→ Οδός Πόλη

Ο ηθοποιός C. Fisher έχει 2 διευθύνσεις:

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

# Πλειότιμες Εξαρτήσεις

$$X \twoheadrightarrow Y$$

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  $X$ :

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

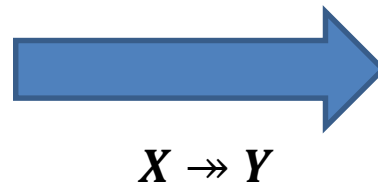
- η  $t_3$  συμφωνεί με την  $t_1$  στο  $Y$ :  $t_3[Y] = t_1[Y]$
- η  $t_3$  συμφωνεί με την  $t_2$  στο  $R - X - Y$ :  $t_3[R - X - Y] = t_2[R - X - Y]$
- η  $t_4$  συμφωνεί με την  $t_2$  στο  $Y$ :  $t_4[Y] = t_2[Y]$
- η  $t_4$  συμφωνεί με την  $t_1$  στο  $R - X - Y$ :  $t_4[R - X - Y] = t_1[R - X - Y]$



# Πλειότιμες Εξαρτήσεις

$X = \{x\}$ ,  $Y = \{y\}$ :

x	y	z
1	0	1
1	1	0



x	y	z
1	0	1
1	1	0
1	0	0
1	1	1

# Πλειότιμες Εξαρτήσεις

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

Όνομα

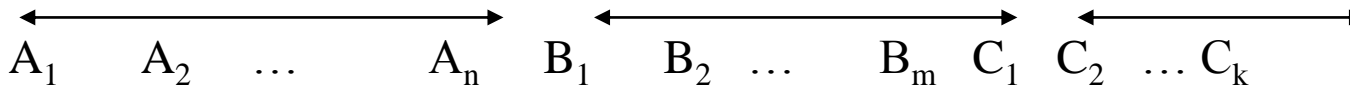
Πόλη Οδός

Τίτλος Έτος

X

Y

R - X - Y



$a_1$	$a_2$	...	$a_n$	$b_1$	$b_2$	...	$b_m$	$c_1$	$c_2$	...	$c_k$	←	$t_1$
$a_1$	$a_2$	...	$a_n$	$b'_1$	$b'_2$	...	$b'_m$	$c'_1$	$c'_2$	...	$c'_k$	←	$t_2$
												←	$t_3$
												←	$t_4$