



EPL342 –Databases

Lecture 22: Functional Dependencies and Normalization

Functional Dependencies

(Chapter 14.2, Elmasri-Navathe 7ED)

Demetris Zeinalipour

<http://www.cs.ucy.ac.cy/courses/EPL342>



Περιεχόμενο Διάλεξης

- **Συναρτησιακές Εξαρτήσεις (Functional Dependencies FD)**
- **Κανόνες Συμπερασμού για Συναρτησιακές Εξαρτήσεις (Inference Rules - IR)**
 - Τα αξιώματα Armstrong
 - Αποδείξεις με χρήση γνωστών κανόνων
- **Ορισμοί Συναρτησιακών Εξαρτήσεων (FD):**
 - Κλειστότητες F^+ και X^+ (Closure)
 - Ισοδυναμία $F^+=G^+$ (Equivalence)
 - Κάλυψη FD (Cover)
 - Ελάχιστη Κάλυψη FD (Minimal Cover)

Συναρτησιακές Εξαρτήσεις – ΣΕ (Functional Dependencies - FD)



- Στην προηγούμενη διάλεξη καλύψαμε με άτυπο τρόπο κάποιες γενικές **κατευθύνσεις** στο **σχεδιασμό ενός καλού σχεσιακού σχήματος**.
- Σε αυτή την διάλεξη θα μελετήσουμε την έννοια των **Συναρτησιακών Εξαρτήσεων (Functional Dependencies, FD)**
- Με χρήση των FD θα μπορέσουμε να αποτιμήσουμε, στην ερχόμενη διάλεξη, με τυπικό τρόπο την **χρησιότητα (goodness)** ενός σχεσιακού σχήματος.
 - Οι Συναρτησιακές Εξαρτήσεις αποτελούν το βασικό υπόβαθρο στο Relational Design Theory

Παραδείγματα FDs



Παραδείγματα Συναρτησιακών Εξαρτήσεων:

- Το **SSN** προσδιορίζει το όνομα του **Employee**.
 - SSN → ENAME
- Το **PNUMBER** προσδιορίζει το **Project Name** και **Location**
 - PNUMBER → {PNAME, PLOCATION}
- Το **SSN** και το **PNumber** προσδιορίζει τον αριθμό ωρών που εργάζεται ένας employee σε ένα project.
 - {SSN, PNUMBER} → HOURS
- Φαινομενικά, το **αριστερό μέλος** ενός FDs είναι κάποιο πρωτεύων κλειδί. Στην πραγματικότητα, μπορεί να είναι οποιοδήποτε/α key ή non-key γνώρισμα/τα, π.χ.,
 - Credits → Status
 - CarModel → Manufacturer
 - NumberGrade → LetterGrade
 - {Author, Title} → Publication Date
- Τα **FDs** ορίζουν **εξαρτήσεις** μεταξύ γνωρισμάτων, οι οποίες προκύπτουν ρητά από τις προδιαγραφές, και οι οποίες οδηγούν σε **επανάληψη (redundancy)** δεδομένων.

Παραδείγματα FDs



- Η συναρτησιακή εξάρτηση **ZIP** → **TOWN** στο ακόλουθο σχήμα προκαλεί την επανάληψη πληροφορίας (redundancy)
 - Π.χ., θεωρώντας ότι οι διευθύνσεις στην ίδια περιοχή έχουν το ίδιο ταχυδρομικό κώδικα (zip)

<i>SSN</i>	<i>Name</i>	<i>Town</i>	<i>Zip</i>
1234	Joe	Stony Brook	11790
4321	Mary	Stony Brook	11790
5454	Tom	Stony Brook	11790
.....			

Redundancy

Κατ' επέκταση, η πιο πάνω **επανάληψη** οδηγεί σε ανωμαλίες **εισαγωγών, διαγραφών και ενημερώσεων** για αυτό η επανάληψη δεδομένων πρέπει να **ελαχιστοποιηθεί!**

Συναρτησιακές Εξαρτήσεις FDs

- Μια **Συναρτησιακή Εξάρτηση** είναι ένας **περιορισμός** μεταξύ δυο **ομάδων γνωρισμάτων** μιας βάσης δεδομένων.
 - Κάντε την παραδοχή ότι όλα τα γνωρίσματα μιας βάσης αποθηκεύονται σε ένα καθολικό πίνακα $R(A_1, A_2, \dots, A_n)$
- Έστω ότι $R=(A_1, A_2, \dots, A_n)$ και ότι $X \subseteq R$ και $Y \subseteq R$, τότε η Συναρτησιακή Εξάρτηση $X \rightarrow Y$, υποδηλώνει ότι μια ανάθεση τιμών στο σύνολο X προσδιορίζει μοναδικά το σύνολο Y .
 - Δηλαδή, εάν $t1[X]=t2[X]$, τότε $t1[Y]=t2[Y]$
 - Επομένως είναι μια γενίκευση της έννοιας του κλειδιού.
- Θα αναφερόμαστε στα FDs ως ακολούθως:
 - Το X προσδιορίζει συναρτησιακά το Y
 - Το Y είναι συναρτησιακά εξαρτώμενο από το X

Παραδείγματα FDs



- Τα **FDs** είναι ένας επιπλέον **μηχανισμός χαρακτηρισμού** των περιορισμών αναφορικής ακεραιότητας (IC) μεταξύ **γνωρισμάτων** μιας σχέσης
 - Ωστόσο τα FDs **ΔΕΝ** δηλώνονται ρητά σε μια βάση, **απλά** χρησιμοποιούνται ως **εργαλείο** για **Εκλέπτυνση** του Σχήματος κατά την φάση της σχεδίασης.
- Για τα FDs υπάρχουν οι ακόλουθες περιπτώσεις:
 - **A) ΙΣΧΥΕΙ για ΚΑΘΕ στιγμιότυπο βάσης.** π.χ., SSN → Teacher (γενικά όλα τα FDs με αριστερό μέλος KEY ισχύουν ΠΑΝΤΑ.)
 - **B) ΜΠΟΡΕΙ να Ισχύει (σε ΚΑΠΟΙΟ στιγμιότυπο βάσης:** π.χ., στο πιο κάτω στιγμιότυπο τυγχάνει να ισχύει το TEXT → COURSE
 - **C) ΔΕΝ Ισχύει σε ΚΑΠΟΙΟ στιγμιότυπο βάσης:** Αρκεί να βρείτε ένα αντιπαράδειγμα π.χ., πιο κάτω TEACHER -/→ COURSE.

TEACH

Teacher	Course	Text
Smith	Data Structures	Bartram
Smith	Data Management	Martin
Hall	Compilers	Hoffman
Brown	Data Structures	Horowitz

Παραδείγματα FDs



Βρείτε ποια FDs **ισχύουν** στο ακόλουθο στιγμιότυπο βάσης (δηλ., ποια **ΜΠΟΡΕΙ** να **ισχύουν** στην ακόλουθη βάση)

	A	B	C	D
1)	a1	b1	c1	d1
2)	a1	b2	c1	d2
3)	a2	b2	c2	d2
4)	a2	b2	c2	d3
5)	a3	b3	c2	d4

- $A \rightarrow C$; **YES (Maybe)**
- $C \rightarrow A$; **NO (line 5)**
- $B \rightarrow C$; **NO (line 3)**
- $D \rightarrow B$; **YES(Maybe)**
- $AB \rightarrow D$; **NO (line 4)**

Κανόνες Συμπερασμού για FDs (Inference Rules, IR)



- Με βάση ένα σύνολο FDs **F** και τους κανόνες του **Armstrong**, μπορούμε να **συμπεράνουμε (infer)** επιπλέον FDs τα οποία ισχύουν όποτε ισχύει το **F**.
 - Π.χ., Από το $FD = \{SSN \rightarrow Dno, Dno \rightarrow Dname\}$ μπορούμε να συμπεράνουμε, με τους κανόνες Armstrong, ότι $SSN \rightarrow Dname$.
- **Οι Κανόνες Συμπερασμού (IR) Armstrong :**
 - **IR1 (Ανακλαστικός, Reflexive):** **Εάν $X \supseteq Y$ τότε $X \rightarrow Y$**
 - π.χ., Εάν $\{ssn, name\} \supseteq name$ τότε $\{ssn, name\} \rightarrow name$
 - Είναι τετριμμένος κανόνας.
 - **IR2 (Επαυξητικός, Augmentation):** **Εάν $X \rightarrow Y$ τότε $XZ \rightarrow YZ$**
 - π.χ., Εάν $ssn \rightarrow name$ τότε $\{ssn, age\} \rightarrow \{name, age\}$
 - * Το XZ σημαίνει $X \cup Z$, επίσης Εάν $X \rightarrow Y$ τότε $XZ \rightarrow Y$
 - **IR3 (Μεταβατικός, Transitive)** **Εάν $X \rightarrow Y$ και $Y \rightarrow Z$ τότε $X \rightarrow Z$**
 - π.χ., Εάν $ssn \rightarrow Dno$ και $Dno \rightarrow Dname$ τότε $ssn \rightarrow Dname$

Κανόνες Συμπερασμού για FDs (Inference Rules, IR)



- Οι κανόνες Armstrong (IR1, IR2, IR3) είναι **βάσιμοι (sound)** και **πλήρεις (complete)**.
 - **Βάσιμοι (Sound)**: Δηλαδή είναι ορθοί για κάθε στιγμιότυπο εισόδου (δείτε αποδείξεις ορθότητας στο βιβλίο)
 - **Πλήρεις (Complete)**: Με βάσει αυτούς μπορούμε να συνάγουμε ΟΛΟΥΣ τους άλλους κανόνες που μπορεί να συναχθούν.
- Στην επόμενη διαφάνεια δείχνουμε μερικούς άλλους Κανόνες IR, τους οποίους μπορούμε να συνάγουμε από τα αξιώματα Armstrong

Κανόνες Συμπερασμού για FDs (Inference Rules, IR)



Κανόνες που Συνάγονται από τα Αξιώματα Armstrong

- IR4 Διάσπαση (Decomposition):

Εάν $X \rightarrow YZ$ τότε $X \rightarrow Y$ και $X \rightarrow Z$

π.χ., Εάν $ssn \rightarrow \{name, age\}$ τότε $ssn \rightarrow name$ και $ssn \rightarrow age$

Προσοχή: Μόνο το δεξί μέλος διασπάται όχι το αριστερό

π.χ., Εάν $\{ssn, name\} \rightarrow age$ ΔΕΝ ΣΗΜΑΙΝΕΙ ότι $name \rightarrow age$

- IR5 Ένωση (Union), [Αντίθετο της διάσπασης]:

Εάν $X \rightarrow Y$ και $X \rightarrow Z$, τότε $X \rightarrow YZ$

π.χ., Εάν $ssn \rightarrow name$ και $ssn \rightarrow age$ τότε $ssn \rightarrow \{name, age\}$

Προσοχή: $X \rightarrow A$ και $Y \rightarrow B$ ΔΕΝ ΣΗΜΑΙΝΕΙ (με βάση μόνο το IR5) ότι $XY \rightarrow AB$

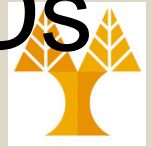
- IR6 Ψευδομετάβαση (Pseudotransitivity):

Εάν $X \rightarrow Y$ και $WY \rightarrow Z$, τότε $WX \rightarrow Z$

π.χ., Εάν $isbn \rightarrow title$ και $\{author, title\} \rightarrow pubdate$ τότε

$\{author, isbn\} \rightarrow pubdate$

Κανόνες Συμπερασμού (IR) για FDs (Αποδείξεις IR4-IR6)



- Οι κανόνες συμπερασμού **IR1-IR3** αποδεικνύονται με **μαθηματικές αποδείξεις** (δείτε βιβλίο) ενώ οι **κανόνες IR4-IR6** και άλλες ασκήσεις με τη **χρήση** των **IR1-IR3**.

- **Απόδειξη IR4 (Διάσπαση):** Εάν $X \rightarrow YZ \models X \rightarrow Y \wedge X \rightarrow Z$

1) $X \rightarrow YZ$ (δεδομένο)

2) $YZ \rightarrow Y$ (IR1: ανακλαστική, το $YZ \supseteq Y$), αντίστοιχα και $YZ \rightarrow Z$ από $YZ \supseteq Z$

3) $X \rightarrow Y$ (IR3: μετάβαση 1-2), αντίστοιχα με (2) $YZ \rightarrow Z$ και (3) έχουμε $X \rightarrow Z$

- **Απόδειξη IR5 (Ένωση):** $(X \rightarrow Y \wedge X \rightarrow Z) \models X \rightarrow YZ$

1) $X \rightarrow Y$ (δεδομένο)

2) $X \rightarrow Z$ (δεδομένο)

3) $XX \rightarrow XY$ (IR2: επαύξηση 1 με X)

4) $X \rightarrow XY$ (απλοποίηση 3, $XX = X$)

5) $XY \rightarrow YZ$ (IR2: επαύξηση 2 με Y)

6) $X \rightarrow YZ$ (IR3: μετάβαση 4-5)

Απόδειξη IR6 (Ψευδομετάβαση):

$X \rightarrow Y \wedge WY \rightarrow Z \models WX \rightarrow Z$

1) $X \rightarrow Y$ (δεδομένο)

2) $WY \rightarrow Z$ (δεδομένο)

3) $WX \rightarrow WY$ (IR2: επαύξηση 1 με W)

4) $WX \rightarrow Z$ (IR3: μετάβαση 3-2)

Ορισμοί Συναρτησιακών Εξαρτήσεων (Κλειστότητες F^+ και X^+)



- **F^+ : Κλειστότητα Συνόλου FD F :** Το σύνολο όλων των συναρτησιακών εξαρτήσεων που προσδιορίζεται από το F (με επαναληπτική εφαρμογή των κανόνων IR1-IR6)
 - π.χ., Εάν $F=\{A \rightarrow B, B \rightarrow C\}$ τότε $F^+=\{A \rightarrow B, B \rightarrow C, A \rightarrow C\}$
- **X^+ : Κλειστότητα Γνωρίσματος X :** Το σύνολο όλων των γνωρισμάτων που εξαρτώνται συναρτησιακά από το X
 - π.χ., Εάν $F=\{A \rightarrow B, B \rightarrow C\}$ τότε $A^+=\{A, B, C\}$, $B^+=\{B, C\}$ και $C^+=\{C\}$

Αλγόριθμος Υπολογισμού του X^+ :

- 1) $X^+ := X$
- 2) repeat
- 3) $old_X^+ := X^+$
- 4) for each FD $Y \rightarrow Z$ in F do
- 5) if $Y \subseteq X^+$ then $X^+ := X^+ \cup Z$
- 6) until ($old_X^+ == X^+$)

Step-by-Step Execution:

- 1) $A^+ := \{A\}$
- 3) $old_A^+ := \{A\}$
- 4-5) $A \subseteq A^+$, so $A^+ := \{A, B\}$
- 4-5) $B \subseteq A^+$, so $A^+ := \{A, B, C\}$
- 6) Now $A^+ == old_A^+$ so quit

- X^+ είναι χρήσιμο εάν θέλουμε να βρούμε κατά πόσο μια FD $X \rightarrow Y$ ανήκει σε κάποιο F^+ (π.χ., βρες εάν το $A \rightarrow C$ είναι στο F^+ $C \in A^+$, άρα είναι!)
- Εάν το X^+ περιέχει όλα τα γνωρίσματα μιας σχέσης τότε το X είναι Candidate Key.

Ορισμοί Συναρτησιακών Εξαρτήσεων (Κλειστότητες F^+ και X^+)



Παράδειγμα Υπολογισμού του X^+

$F = \{ \text{SSN} \rightarrow \text{Ename},$
 $\text{Pnumber} \rightarrow \{ \text{Pname}, \text{Plocation} \}$
 $\{ \text{Ssn}, \text{Pnumber} \} \rightarrow \text{Hours} \}$

- 1) $\{ \text{Ssn} \}^+ = \{ \text{Ssn}, \text{Ename} \}$
- 2) $\{ \text{Pnumber} \}^+ = \{ \text{Pnumber}, \text{Pname}, \text{Plocation} \}$
- 3) $\{ \text{Ssn}, \text{Pnumber} \}^+ = \{ \text{Ssn}, \text{Pnumber}, \text{Ename},$
 $\text{Pname}, \text{Plocation}, \text{Hours} \}$

➔ Το $\{ \text{Ssn}, \text{Pnumber} \}$ είναι **candidate key!**

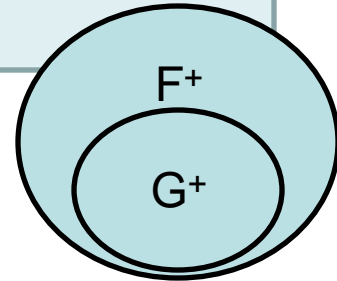
Ορισμοί Συναρτησιακών Εξαρτήσεων (Ισοδυναμία Συνόλου FDs $F^+=G^+$)



- **Κάλυψη FDs (Cover):** Ένα σύνολο FD F , καλύπτει ένα άλλο σύνολο G εάν $G^+ \subseteq F^+$

- Εναλλακτικά, το F καλύπτει το G εάν:
 - Κάθε FD του G μπορεί να συμπεραθεί από το F .
 - π.χ., $\{SSN\}^+=\{SSN, NumGrade, LetterGrade\}$

Το FD $SSN \rightarrow LetterGrade$ καλύπτεται από το $SSN \rightarrow all$



- **Ισοδυναμία FDs (Equivalence):** Δυο σύνολα FDs F και G είναι ισοδύναμα εάν $F^+=G^+$

- Εναλλακτικά, τα F και G είναι **ισοδύναμα** εάν:
 - **ΚΑΘΕ** FD του F μπορεί να συμπεραθεί από το G και
 - **ΚΑΘΕ** FD του G μπορεί να συμπεραθεί από το F
- Συνεπώς, τα F και G είναι **ισοδύναμα** εάν:
 - Το F καλύπτει το G και το G καλύπτει το F .

Παράδειγμα Κάλυψης



Σας δίνεται το σύνολο Συναρτησιακών Εξαρτήσεων F .
Βρείτε εάν τα $AB \rightarrow E$ και $D \rightarrow C$ καλύπτονται από το F

$F: A \rightarrow D$

$AB \rightarrow C$

$D \rightarrow E$

$AC \rightarrow B$

Καλύπτεται το $AB \rightarrow E$ από το F ;
ΝΑΙ, επειδή το $E \in \{AB\}^+$.

Καλύπτεται το $D \rightarrow C$ από το F ;
ΟΧΙ, επειδή το $C \notin D^+$.

Συμπέρασμα: Το X^+ μας επιτρέπει να βρίσκουμε εάν η συναρτησιακή εξάρτηση της μορφής $X \rightarrow Y$ καλύπτεται από το σύνολο FDs F .

X

X^+

A

$\{A, D, E\}$

AB

$\{A, B, C, D, E\}$

(Επομένως AB είναι κλειδί)

D

$\{D, E\}$

AC

$\{A, B, C, D, E\}$

$\Rightarrow AB, AC$ κλειδιά

Ορισμοί Συναρτησιακών Εξαρτήσεων (Απλό Παράδειγμα Ισοδυναμίας)



Αποδείξτε ότι τα σύνολα FDs F και G είναι ισοδύναμα

$F = \{ \text{SSN} \rightarrow \text{Ename}, \text{Pnumber} \rightarrow \{\text{Pname}, \text{Plocation}\}, \{\text{Ssn}, \text{Pnumber}\} \rightarrow \text{Hours} \}$

$G = \{ \text{SSN} \rightarrow \text{Ename}, \text{Pnumber} \rightarrow \{\text{Pname}, \text{Plocation}\}, \{\text{Ssn}, \text{Pnumber}\} \rightarrow \text{Ename}, \text{Pname}, \text{Plocation}, \text{Hours} \}$

Πρέπει να αποδείξουμε τις ακόλουθες δυο περιπτώσεις:

a) Κάθε FD του G καλύπτεται από το F (δηλ., $G^+ \subseteq F^+$)

Τα πρώτα δυο FDs των συνόλων είναι τα ίδια. Για το τρίτο FD, υπολογίζουμε το $\{\text{Ssn}, \text{Pnumber}\}^+$ (στο F) το οποίο είναι $\{\text{Ssn}, \text{Pnumber}, \text{Ename}, \text{Pname}, \text{Plocation}, \text{Hours}\}$

Το $\{\text{Ssn}, \text{Pnumber}\}^+$ (του F) καλύπτει τον τρίτο κανόνα του G εφόσον περιέχει όλα τα γνωρίσματα στο δεξί του μέλος (δηλ., $\text{Ename}, \text{Pname}, \text{Plocation}, \text{Hours}$)

β) Κάθε FD του F καλύπτεται από το G (δηλ., $F^+ \subseteq G^+$)

Αντίστοιχα με το (a) βρίσκουμε ότι ο τρίτος κανόνας του F καλύπτεται από το $\{\text{Ssn}, \text{Pnumber}\}^+$ του G .

Εφόσον το G καλύπτει το F και αντίστροφα, τα FDs είναι ισοδύναμα.