Σχεδιασμός Σχεσιακών Σχημάτων (2)

Σχεδιασμός Σχεσιακών Σχημάτων - Επανάληψη

Σχεδιασμός *καλών* σχεσιακών σχημάτων

- Μη τυπικές - γενικές κατευθύνσεις
- Θεωρία κανονικών μορφών που θα βασίζεται στις συναρτησιακές εξαρτήσεις

Γενικές Κατευθύνσεις

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

Αποσύνθεση (decomposition)

Αλγόριθμος σχεδιασμού

- Αρχικά ένα καθολικό σχήμα σχέσης που περιέχει όλα τα γνωρίσματα
- Προσδιορισμός των συναρτησιακών εξαρτήσεων
- Διάσπαση σε ένα σύνολο από σχήματα που ικανοποιούν κάποιες ιδιότητες

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

γνωρίσματα

$$R = R_1, \cup R_2, ..., \cup R_n$$

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

πλειάδες

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

$$r \subseteq r_1 * r2 * ... * r_n$$

Επιθυμητές Ιδιότητες Αποσύνθεσης

1. Συνενώσεις Άνευ Απωλειών

Η φυσική συνένωση των σχέσεων που προκύπτουν μας δίνει *ακριβώς* την αρχική σχέση (χωρίς επιπρόσθετες πλειάδες)

2. Διατήρηση Εξαρτήσεων

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

3. Αποφυγή Επανάληψης Πληροφορίας

Συνενώσεις Άνευ Απωλειών

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

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

Θεώρημα: Έστω R ένα σχεσιακό σχήμα και F ένα σύνολο από συναρτησιακές εξαρτήσεις στο R. Έστω R_1 και R_2 μια αποσύνθεση του P. Αν μια τουλάχιστον από τις $R_1 \cap R_2 \rightarrow R_1$ **ή** $R_1 \cap R_2 \rightarrow R_2$ **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*

Αποφυγή δημιουργίας πλασματικών πλειάδων

(αδυναμία αναπαράστασης συγκεκριμένης πληροφορίας)

Ταινία

Τίτλος	Έτος	Διάρκεια	Είδος	Όνομα-Ηθοποιού

Ταινία

Τίτλος	Έτος	Διάρκεια	Είδος

Παίζει

Τίτλος	Όνομα-Ηθοποιού

Διατήρηση Εξαρτήσεων

Έλεγχος διατήρησης εξαρτήσεων όταν γίνονται τροποποιήσεις χωρίς να υπολογίζουμε τις αρχικές σχέσεις

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

Ο **περιορισμός του F στο R_i** είναι το σύνολο F_i όλων των συναρτησιακών εξαρτήσεων του F^+ που περιέχουν μόνο γνωρίσματα του R_i.

Έστω $F' = F_1 \cup F_2 \ldots \cup F_n$

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

Παραδείγματα

1. Έστω R(A, B, C, D), F = {A →C, B → C, BD → A}. Η αποσύνθεση του R σε S(A, C) και T(A, B, D) διατηρεί τις εξαρτήσεις ;

2. Έστω R(A, B, C, D, E), F = {A → D, B → E, DE → C}. Η αποσύνθεση του R σε S(A, B, C) και T(A, B, D, E) διατηρεί τις εξαρτήσεις ;

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

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, Εταιρεία-Παραγωγής, Όνομα-Ηθοποιού)

Η σχέση Ταινία δεν είναι σε BCNF

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

Για παράδειγμα η ΣΕ Τίτλος Έτος → Εταιρεία-Παραγωγής

Boyce-Codd Κανονική Μορφή

Παράδειγμα 2

Ταινία2 (Τίτλος, Έτος, Διάρκεια, Είδος, Εταιρεία-Παραγωγής)

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 $X \cap Y = \varnothing$
- 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, Εταιρεία-Παραγωγής, Όνομα-Ηθοποιού)

Τίτλος Έτος $\rightarrow$ Διάρκεια Είδος Εταιρεία-Παραγωγής

Ταινία1(Τίτλος, Έτος, Διάρκεια, Είδος, Εταιρεία-Παραγωγής)

Ταινία2(Τίτλος, Έτος, Όνομα-Ηθοποιού)

Boyce-Codd Κανονική Μορφή

Παράδειγμα 2

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

Πρόβλημα: υπάρχει μια ***μεταβατική*** εξάρτηση

Τίτλος Έτος → Εταιρεία-Παραγωγής

Εταιρεία-Παραγωγής → Διεύθυνση-Εταιρείας

Τίτλος Έτος → Διεύθυνση-Εταιρείας

Ταινία-Εταιρεία1 (Εταιρεία-Παραγωγής, Διεύθυνση-Εταιρείας)
Ταινία-Εταιρεία2 (Τίτλος, Έτος, Διάρκεια, Είδος, Εταιρεία-Παραγωγής)

Boyce-Codd Κανονική Μορφή

- 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

Boyce-Codd Κανονική Μορφή

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

Παράδειγμα

Έστω η σχέση Παίζει(Έργο, Κινηματογράφος, Πόλη) με τους περιορισμούς ότι δεν υπάρχουν κινηματογράφοι με το ίδιο όνομα [~~στην ίδια πόλη~~], κάθε έργο παίζεται μόνο σε ένα κινηματογράφο σε κάθε πόλη και κάθε κινηματογράφος έχει πολλές αίθουσες (παίζει πολλά έργα)

Κινηματογράφος → Πόλη	Κλειδιά;
Έργο Πόλη → Κινηματογράφος	{Έργο, Πόλη} {Κινηματογράφος, Έργο}

Boyce-Codd Κανονική Μορφή

Παίζει(Έργο, Κινηματογράφος, Πόλη)

	Κλειδιά
Κινηματογράφος → Πόλη	{Έργο, Πόλη}
Έργο Πόλη → Κινηματογράφος	{Κινηματογράφος, Έργο}

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

Κινηματογράφος	Πόλη
Οντεόν	Αθήνα
Village Center	Αθήνα

Κινηματογράφος	Έργο
Οντεόν	Safe Sex
Village Center	Safe Sex

Τρίτη Κανονική Μορφή

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 ⇒ 3NF

Παίζει(Έργο, Κινηματογράφος, Πόλη)

Κινηματογράφος → Πόλη

Έργο Πόλη → Κινηματογράφος

Κλειδιά {Έργο, Πόλη}
{Κινηματογράφος, Έργο}

Η σχέση είναι σε 3NF

Είδος ΣΕ: μοναδικά αντικείμενα ή

με βάση πρακτικές

Αλγόριθμος Αποσύνθεσης σε 3NF

- Υπολόγισε το ελάχιστο κάλυμμα F_c της F
- Για κάθε ΣΕ $X \rightarrow Y$ της F_c

  αν κανένα από τα R_i δεν περιέχει τα X, Y

  νέα σχέση με γνωρίσματα $X \cup Y$

- Αν καμία από τις σχέσεις που προέκυψαν δεν περιέχει κάποιο υποψήφιο κλειδί του R δημιούργησε μια νέα σχέση με γνωρίσματα τα γνωρίσματα του κλειδιού

Αλγόριθμος Αποσύνθεσης σε 3NF

- Απώλειες στη συνένωση;
- Διατήρηση εξαρτήσεων;

Παράδειγμα

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

Όνομα-Τραπεζίτη → Όνομα-Υποκαταστήματος Αριθμός-Γραφείου

Όνομα-Πελάτη Όνομα-Υποκαταστήματος → Όνομα-Τραπεζίτη

Κλειδιά {Όνομα-Πελάτη, Όνομα-Υποκαταστήματος} 3NF;

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

Τραπεζίτης2(Όνομα-Πελάτη, Όνομα-Υποκαταστήματος, Όνομα-Τραπεζίτη) *BCNF;*

BCNF - 3NF

- διατήρηση εξαρτήσεων
- απώλειες στη συνένωση
- επανάληψη πληροφορίας λόγω ΣΕ

Ανακεφαλαίωση

- Αποσύνθεση καθολικού σχήματος
  - Επιθυμητές ιδιότητες
    - διατήρηση εξαρτήσεων
    - όχι απώλειες στη συνένωση
    - επανάληψη πληροφορίας λόγω ΣΕ
- Κανονικές Μορφές
  - BCNF
  - 3NF