ΣΧΕΔΙΑΣΜΟΣ ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΚΑΝΟΝΙΚΟΠΟΙΗΣΗ

Β’ μέρος

Σκοπός: Να βρούμε θεωρία ώστε…

- Να αποφασίζουμε αν μια σχέση R είναι σε «καλή» μορφή
- Σε περίπτωση που η R δεν είναι σε καλή μορφή, να την αποσυνθέσουμε σε ένα σύνολο σχέσεων $\{R_1, R_2, ..., R_n\}$ τέτοιων ώστε
  - Κάθε σχέση να είναι σε καλή μορφή
  - Η αποσύνθεση να μην έχει απώλειες πληροφορίας
- Η θεωρία βασίζεται σε **συναρτησιακές εξαρτήσεις**

Συναρτησιακές εξαρτήσεις

- Περιορισμοί στο σύνολο των αποδεκτών σχέσεων
- Απαιτούν η τιμή για συγκεκριμένο σύνολο γνωρισμάτων να ορίζει μονοσήμαντα την τιμή κάποιου άλλου συνόλου γνωρισμάτων
- Γενίκευση της έννοιας του κλειδιού

Συναρτησιακές εξαρτήσεις

- K κλειδί της R: $K \to R$
- K υποψήφιο κλειδί της R
  - $K \to R$, και
  - Δεν υπάρχει $\alpha \subset K$ ώστε $\alpha \to R$
- Εκφράζουν και περιορισμούς που δεν μπορούν να εκφραστούν με κλειδιά. Π.χ.

Loan-info-schema = (customer-name, loan-number, branch-name, amount).

Ισχύει:

loan-number $\to$ *amount*
loan-number $\to$ *branch-name*

Δεν ισχύει:

loan-number $\to$ *customer-name*

Κλειστότητα Συνόλου Συναρτησιακών Εξαρτήσεων

- Δεδομένου ενός συνόλου συναρτησιακών εξαρτήσεων *F*, υπάρχουν άλλες εξαρτήσεις που προκύπτουν λογικά από την *F*.
  - π.χ. αν $A \rightarrow B$ και $B \rightarrow C$, τότε $A \rightarrow C$
- Το σύνολο των συναρτησιακών εξαρτήσεων που προκύπτουν από το σύνολο F λέγεται κλειστότητα του F και συμβολίζεται F^+.
- Μπορούμε να βρούμε το F^+ εφαρμόζοντας τα αξιώματα του Armstrong:
  - αν $\beta \subseteq \alpha$, τότε $\alpha \rightarrow \beta$ **(ανακλαστικότητα-reflexivity)**
  - αν $\alpha \rightarrow \beta$, τότε $\gamma\,\alpha \rightarrow \gamma\,\beta$ **(προσαύξηση-augmentation)**
  - αν $\alpha \rightarrow \beta$ και $\beta \rightarrow \gamma$, τότε $\alpha \rightarrow \gamma$ **(μεταβατικότητα-transitivity)**
- Έγκυροι και πλήρεις κανόνες

Επιπλέον απλούστεροι κανόνες

- Επιπλέον κανόνες που προκύπτουν από τα αξιώματα Armstrong και μπορούν να χρησιμοποιηθούν για τον υπολογισμό του F^+
  - αν $\alpha \rightarrow \beta$ *και* $\alpha \rightarrow \gamma$, τότε $\alpha \rightarrow \beta\gamma$ **(ένωση-union)**
  - αν $\alpha \rightarrow \beta\gamma$, τότε $\alpha \rightarrow \beta$ και $\alpha \rightarrow \gamma$ **(αποσύνθεση-decomposition)**
  - αν $\alpha \rightarrow \beta$ *και* $\gamma\beta \rightarrow \delta$, τότε $\alpha\gamma \rightarrow \delta$ **(ψευδομεταβατικότητα-pseudotransitivity)**

Υπολογισμός του F^+

- Αλγόριθμος:

$F^+ = F$
repeat
    for each functional dependency f in F^+
        apply reflexivity and augmentation rules on f
        add the resulting functional dependencies to F^+
    for each pair of functional dependencies f_1 and f_2 in F^+
        if f_1 and f_2 can be combined using transitivity
            then add the resulting functional dependency to F^+
until F^+ does not change any further

«ΑΡΓΟΣ» Αλγόριθμος

Παραδείγματα

- $R = (A, B, C, G, H, I)$
  $F = \{ A \rightarrow B$
  $A \rightarrow C$
  $CG \rightarrow H$
  $CG \rightarrow I$
  $B \rightarrow H\}$
- Συναρτησιακές εξαρτήσεις στην F^+
  - $A \rightarrow H$
    - » Με μεταβατικότητα $A \rightarrow B$ *και* $B \rightarrow H$
  - $AG \rightarrow I$
    - » Προσαυξηστικά από το $A \rightarrow C$ με το G, παίρνουμε το $AG \rightarrow CG$ και μετά με μεταβατικότητα με το $CG \rightarrow I$
  - $CG \rightarrow HI$
    - » Από το $CG \rightarrow H$ *και το* $CG \rightarrow I$: “κανόνας ένωσης”

Κλειστότητα συνόλων ιδιοτήτων

- Δεδομένου συνόλου γνωρισμάτων α, ορίζουμε την κλειστότητα (*closure*) του α στο F (α^+) ως το σύνολο όλων των γνωρισμάτων που εξαρτώνται συναρτησιακά από το α στο F:

$$\alpha \rightarrow \beta \text{ is in } F^+ \Leftrightarrow \beta \subseteq \alpha^+$$

Υπολογισμός Κλεισίματος Συνόλου Γνωρισμάτων (α^+)

- Αλγόριθμος υπολογισμού α^+ στο F

result := α;
    while (changes to *result*) **do**
        for each $\beta \rightarrow \gamma$ **in** *F* **do**
            begin
                if $\beta \subseteq result$ **then** $result := result \cup \gamma$
            end

«ΓΡΗΓΟΡΟΣ» Αλγόριθμος

Παραδείγματα του Κλεισίματος Συνόλου Γνωρισμάτων (α^+)

- $R = (A, B, C, G, H, I)$
- $F = \{A \rightarrow B$
  $A \rightarrow C$
  $CG \rightarrow H$
  $CG \rightarrow I$
  $B \rightarrow H\}$
- $(AG)^+$
  1. *result* = *AG*
  2. *result* = *ABCG* ($A \rightarrow C$ and $A \rightarrow B$)
  3. *result* = *ABCGH* ($CG \rightarrow H$ and $CG \subseteq AGBC$)
  4. *result* = *ABCGHI* ($CG \rightarrow I$ and $CG \subseteq AGBCH$)
- Είναι το *AG* υποψήφιο κλειδί?
  1. Είναι το AG κλειδί?
     1. *Η Ερώτηση AG* $\rightarrow$ *R? Είναι ισοδύναμη με την « είναι το* $(AG)^+ \supseteq R$»?
  2. Είναι κάποιο υποσύνολο του AG κλειδί?
     1. Είναι το $A \rightarrow R$? == Είναι το $(A)^+ \supseteq R$
     2. Είναι το $G \rightarrow R$? == Είναι το $(G)^+ \supseteq R$

Χρήσεις του Κλεισίματος Συνόλου Γνωρισμάτων (α^+)

- ΕΛΕΓΧΟΣ ΚΛΕΙΔΙΟΥ:
  - Για τον έλεγχο αν το α είναι κλειδί, υπολογίζουμε το α^+, και ελέγχουμε αν το α^+ περιέχει όλα τα γνωρίσματα του *R*.
- ΕΛΕΓΧΟΣ ΣΥΝΑΡΤΗΣΙΑΚΩΝ ΕΞΑΡΤΗΣΕΩΝ
  - Για τον έλεγχο αν η συναρτησιακή εξάρτηση $\alpha \rightarrow \beta$ ισχύει (με άλλα λόγια, αν είναι στο F^+), έλεγξε αν $\beta \subseteq \alpha^+$.
  - Δηλαδή, υπολογίζουμε το α^+ με το γνωστό αλγόριθμο, και ελέγχουμε αν περιέχει το β.
  - Είναι ένα εύκολος και γρήγορος έλεγχος
- ΥΠΟΛΟΓΙΣΜΟΣ ΤΗΣ ΚΛΕΙΣΤΟΤΗΤΑΣ ΤΟΥ F
  - Για κάθε $\gamma \subseteq R$, υπολογίζουμε το γ^+, και για κάθε $S \subseteq \gamma^+$, παίρνουμε μια συναρτησιακή εξάρτηση $\gamma \rightarrow S$.

Κανονική Κάλυψη

- Ένα σύνολο συναρτησιακών εξαρτήσεων μπορεί να έχει πλεονάζουσες εξαρτήσεις που προκύπτουν από άλλες
  - Π.χ.:η $A \rightarrow C$ στο F $\{A \rightarrow B,\ \ B \rightarrow C,\ \ A \rightarrow C\}$
  - Μέρη εξαρτήσεών είναι πλεονάζοντα
    - » Π.χ το $\{A \rightarrow B,\ \ B \rightarrow C,\ \ A \rightarrow CD\}$ μπορεί να απλοποιηθεί στο $\{A \rightarrow B,\ \ B \rightarrow C,\ \ A \rightarrow D\}$
    - » Π.χ. το $\{A \rightarrow B,\ \ B \rightarrow C,\ \ AC \rightarrow D\}$ μπορεί να απλοποιηθεί σε $\{A \rightarrow B,\ \ B \rightarrow C,\ \ A \rightarrow D\}$
- Διαισθητικά μια κανονική κάλυψη του F είναι ένα ελάχιστο σύνολο από συναρτησιακές εξαρτήσεις ισοδύναμο με το F, που δεν έχει πλεονάζουσες εξαρτήσεις ή πλεονάζοντα μέρη εξαρτήσεων.

Εξωτερικά Γνωρίσματα

- Αν F σύνολο συναρτησιακών εξαρτήσεων που περιέχει εξάρτηση $\alpha \rightarrow \beta$ τότε
  - Το γνώρισμα A είναι εξωτερικό στο α αν $A \in \alpha$ το F συνεπάγεται λογικά το $(F - \{\alpha \rightarrow \beta\}) \cup \{(\alpha - A) \rightarrow \beta\}$.
  - Το γνώρισμα A είναι εξωτερικό στο β αν $A \in \beta$ και το $(F - \{\alpha \rightarrow \beta\}) \cup \{\alpha \rightarrow (\beta - A)\}$ συνεπάγεται λογικά το F

Εξωτερικά Γνωρίσματα

- $F = \{A \rightarrow C, AB \rightarrow C\}$
  - *Το B* είναι εξωτερικό στο $AB \rightarrow C$ γιατί το F $\{A \rightarrow C, AB \rightarrow C\}$ συνεπάγεται λογικά το $A \rightarrow C$ (δλδ το αποτέλεσμα της διαγραφής του *B* από το $AB \rightarrow C$).
- $F = \{A \rightarrow C, AB \rightarrow CD\}$
  - *Το C* είναι εξωτερικό του $AB \rightarrow CD$ γιατί το $\{A \rightarrow C, AB \rightarrow D\}$ συνεπάγεται λογικά το F (δλδ το $AB \rightarrow CD$) ακόμα και μετά τη διαγραφή του *C*

Έλεγχος αν ένα γνώρισμα είναι εξωτερικό

- Θεωρούμε σύνολο συναρτησιακών εξαρτήσεων F και την εξάρτηση $\alpha \rightarrow \beta$ στο F.
- Για να ελέγξουμε αν ένα γνώρισμα $A \in \alpha$ είναι εξωτερικό στο α
  1. Υπολογίζουμε το $(\{\alpha\} - A)^+$ στο F
  2. ελέγχουμε ότι το $(\{\alpha\} - A)^+$ περιέχει το A
- Για να ελέγξουμε αν ένα γνώρισμα $A \in \beta$ είναι εξωτερικό στο β
  1. Υπολογίζουμε α^+ στο
     $$F' = (F - \{\alpha \rightarrow \beta\}) \cup \{\alpha \rightarrow (\beta - A)\},$$
  2. ελέγχουμε ότι το α^+ περιέχει το A

Κανονική κάλυψη

- Μια κανονική κάλυψη του F είναι ένα σύνολο από συναρτησιακές εξαρτήσεις F_c τέτοιο ώστε
  – *Το F* συνεπάγεται λογικά όλες τις εξαρτήσεις του F_c
  – *Το* F_c συνεπάγεται λογικά όλες τις εξαρτήσεις του F
  – Δεν υπάρχει συναρτησιακή εξάρτηση στο F_c που να περιέχει εξωτερικό γνώρισμα
  – Η αριστερή πλευρά κάθε συναρτησιακής εξάρτησης στο F_c είναι μοναδική

Αλγόριθμος εύρεσης κανονικής κάλυψης

$F_c = F$

repeat

Use the union rule to replace any dependencies in *F*

$\alpha_1 \rightarrow \beta_1$ and $\alpha_1 \rightarrow \beta_2$ with $\alpha_1 \rightarrow \beta_1 \beta_2$

Find a functional dependency $\alpha \rightarrow \beta$ with an extraneous attribute either in α or in β

If an extraneous attribute is found, delete it from $\alpha \rightarrow \beta$

until F_c does not change

Σημείωση: Ο κανόνας της ένωσης μπορεί να είναι εφαρμόσιμος μετά τη διαγραφή κάποιων εξωτερικών γνωρισμάτων, οπότε θα πρέπει να εφαρμοστεί ξανά

Παράδειγμα

- $R = (A, B, C)$
  $F = \{A \rightarrow BC$
  $B \rightarrow C$
  $A \rightarrow B$
  $AB \rightarrow C\}$
- Combine $A \rightarrow BC$ and $A \rightarrow B$ into $A \rightarrow BC$
  - Set is now $\{A \rightarrow BC, B \rightarrow C, AB \rightarrow C\}$
- A is extraneous in $AB \rightarrow C$
  - Check if the result of deleting A from $AB \rightarrow C$ is implied by the other dependencies
    - » Yes: in fact, $B \rightarrow C$ is already present!
  - Set is now $\{A \rightarrow BC, B \rightarrow C\}$
- C is extraneous in $A \rightarrow BC$
  - Check if $A \rightarrow C$ is logically implied by $A \rightarrow B$ and the other dependencies
    - » Yes: using transitivity on $A \rightarrow B$ *and* $B \rightarrow$ C.
      - Can use attribute closure of A in more complex cases
- Η ελάχιστη κάλυψη είναι: $A \rightarrow B$
  $B \rightarrow C$

Σχεσιακή Αποσύνθεση

- ΤΟ ΑΡΧΙΚΟ ΣΗΜΕΙΟ όλων των αλγορίθμων είναι ένα ***σχήμα καθολικής σχέσης R*** που περιέχει ***όλα*** τα γνωρίσματα της Βάσης
- ΣΤΟΧΟΣ του σχεδιασμού είναι **μια αποσύνθεση (decomposition) D του R** σε **m** Σχήματα Σχέσεων $R_1, R_2, R_3, \ldots R_m$ όπου κάθε R_i περιέχει ένα υποσύνολο γνωρισμάτων του **R** και κάθε γνώρισμα του **R** πρέπει να παρουσιάζεται σε **τουλάχιστον ένα** Σχήμα Σχέσης R_i **(ιδιότητα διατήρησης Γνωρισμάτων)**

- Στην ουσία, όταν κάνουμε αποσύνθεση του R, σκοπεύουμε να αποθηκεύσουμε στιγμιότυπα των Σχέσεων που προκύπτουν από την αποσύνθεση, αντί για στιγμιότυπα του R.

Αποσύνθεση

- Αποσύνθεση της σχέσης *Lending-schema (branch-name, branch-city, assets, customer-name, loan-number, amount) σε* :

Branch-schema = (branch-name, branch-city, assets)

Loan-info-schema = (customer-name, loan-number, branch-name, amount)

- Όλα τα γνωρίσματα του αρχικού σχήματος (*R*) πρέπει να εμφανίζονται στην αποσύνθεση (R_1, R_2):

$$R = R_1 \cup R_2$$

- Αποσύνθεση χωρίς απώλειες
για όλες της πιθανές σχέσεις *r* στο σχήμα *R*

$$r = \prod_{R1}(r) \bowtie \prod_{R2}(r)$$

- Μια αποσύνθεση του R σε R_1 and R_2 είναι χωρίς απώλειες αν και μόνο αν υπάρχει τουλάχιστον μια συναρτησιακή εξάρτηση στο F^+:
  - $R_1 \cap R_2 \rightarrow R_1$
  - $R_1 \cap R_2 \rightarrow R_2$

(το $R_1 \cap R_2$ κλειδί του R_1 ή του R_2)

Παράδειγμα αποσύνθεσης με απώλειες

- Οδηγεί σε λάθος πληροφορίες (πλασματικές πλειάδες)
- Π.χ. $R = (A, B)$

$$R_1 = (A) \qquad R_2 = (B)$$

A	*B*
α	1
α	2
β	1

r

A
α
β

$\Pi_A(r)$

B
1
2

$\Pi_{B(r)}$

$\Pi_A (r) \bowtie \Pi_B (r)$

A	*B*
α	1
α	2
β	1
β	2

Αποσύνθεση

- Κατά την αποσύνθεση του σχήματος R με σύνολο συναρτησιακών εξαρτήσεων F σε R_1, R_2,.., R_n θέλουμε
  - Αποσύνθεση χωρίς απώλειες
  - Αποφυγή πλεονασμού: Οι σχέσεις R_i να είναι κατά προτίμηση σε κανονική μορφή Boyce-Codd ή σε 3NF.
  - Διατήρηση εξαρτήσεων: Αν F_i το σύνολο των εξαρτήσεων του F^+ που περιλαμβάνουν μόνο γνωρίσματα του R_i.
    - » $(F_1 \cup F_2 \cup \ldots \cup F_n)^+ = F^+$
    - » Αλλιώς ο έλεγχος αν οι ενημερώσεις τηρούν τις συναρτησιακές εξαρτήσεις απαιτούν joins που είναι ακριβά

Παράδειγμα

- $R = (A, B, C)$
  $F = \{A \rightarrow B, B \rightarrow C)$
  – Μπορεί να αποσυντεθεί με 2 τρόπους
- $R_1 = (A, B), \quad R_2 = (B, C)$
  – Χωρίς απώλειες:
  $$R_1 \cap R_2 = \{B\} \text{ and } B \rightarrow BC$$
  – Διατηρεί εξαρτήσεις
- $R_1 = (A, B), \quad R_2 = (A, C)$
  – Χωρίς απώλειες :
  $$R_1 \cap R_2 = \{A\} \text{ and } A \rightarrow AB$$
  – Δε διατηρεί εξαρτήσεις
  (δε γίνεται να ελέγξεις το $B \rightarrow C$ χωρίς να υπολογίσεις το
  $R_1 \bowtie R_2$)

Έλεγχος διατήρησης εξαρτήσεων

- Για να ελέγξουμε αν η εξάρτηση $\alpha \to \beta$ διατηρείται στην αποσύνθεση του R σε $R_1, R_2, \ldots, R_n$ χρησιμοποιούμε την κλειστότητα γνωρισμάτων
  - *result* = α
    while (changes to *result*) do
      for each R_i in the decomposition
        $t = (result \cap R_i)^+ \cap R_i$
        $result = result \cup t$
  - If *result* contains all attributes in β, then the functional dependency $\alpha \to \beta$ is preserved.
- Το εφαρμόζουμε για όλες τις εξαρτήσεις του F
- Απαιτεί πολυωνυμικό χρόνο αντί του εκθετικού για τον υπολογισμό του F^+ και $(F_1 \cup F_2 \cup \ldots \cup F_n)^+$

Κανονική μορφή Boyce-Codd

Ένα σχήμα σχέσης R είναι σε BCNF σε σχέση με ένα σύνολο συναρτησιακών εξαρτήσεων F όταν για όλες τις εξαρτήσεις στο F^+ της μορφής $\alpha \rightarrow \beta$, όπου $\alpha \subseteq R$ and $\beta \subseteq R$, ισχύει τουλάχιστον ένα από τα παρακάτω

- $\alpha \rightarrow \beta$ είναι τετριμμένη (i.e., $\beta \subseteq \alpha$)
- α είναι κλειδί της R

Παράδειγμα

- $R = (A, B, C)$
  $F = \{A \rightarrow B$
  $\quad B \rightarrow C\}$
  κλειδί = $\{A\}$
- R δεν είναι σε BCNF
- Decomposition $R_1 = (A, B)$, $R_2 = (B, C)$
  - R_1 και R_2 είναι σε BCNF
  - Lossless-join decomposition
  - Dependency preserving

Έλεγχος για BCNF

- Για μη τετριμμένη σχέση $\alpha \to \beta$
  1. Υπολογίζω α^+
  2. Τσεκάρω ότι περιλαμβάνει όλα τα γνωρίσματα του R (κλειδί του R)
- Απλοποιημένος έλεγχος: Αρκεί να ελέγξω τις εξαρτήσεις στο F αντί όλες τι εξαρτήσεις στο F^+.
- Να χρησιμοποιήσω το F αντί του F^+ όμως είναι λάθος για έλεγχο σχέσης από αποσύνθεση του R
  – E.g. Consider R (A, B, C, D), with $F = \{ A \to B, B \to C\}$
    » Decompose R into $R_1(A,B)$ and $R_2(A,C,D)$
    » Neither of the dependencies in F contain only attributes from (A,C,D) so we might be mislead into thinking R_2 satisfies BCNF.
    » In fact, dependency $A \to C$ in F^+ shows R_2 is not in BCNF.

Αλγόριθμος αποσύνθεσης σε BCNF

result := $\{R\}$;
done := false;
compute F^+;
while (not *done*) **do**
  if (there is a schema R_i in *result* that is not in BCNF)
    then begin
      let $\alpha \to \beta$ be a nontrivial functional
        dependency that holds on R_i
        such that $\alpha \to R_i$ is not in F^+,
        and $\alpha \cap \beta = \emptyset$;
      $result := (result - R_i) \cup (R_i - \beta) \cup (\alpha, \beta);$
    end
    else *done* := **true;**

Note: each R_i is in BCNF, and decomposition is lossless-join.

Παράδειγμα αποσύνθεσης BCNF

- *R = (branch-name, branch-city, assets, customer-name,*
  loan-number, amount)

  F = {branch-name $\rightarrow$ *assets branch-city*

  loan-number $\rightarrow$ *amount branch-name}*

  Key = *{loan-number, customer-name}*

- Decomposition
  – R_1 = (*branch-name, branch-city, assets)*
  – R_2 = (*branch-name, customer-name, loan-number, amount)*
  – R_3 = (*branch-name, loan-number, amount)*
  – R_4 = (*customer-name, loan-number)*

- Final decomposition

  R_1, R_3, R_4

BCNF and Dependency Preservation

Δεν είναι πάντα δυνατόν να βρεθεί αποσύνθεση BCNF που διατηρεί εξαρτήσεις

- $R = (J, K, L)$
  $F = \{JK \rightarrow L$
  $L \rightarrow K\}$
  δύο υποψήφια κλειδιά, *JK* and *JL*
- *R* δεν είναι σε BCNF
- Καμία αποσύνθεση του *R* δε διατηρεί την εξάρτηση

$$JK \rightarrow L$$

3NF: Κίνητρο

- Υπάρχουν περιπτώσεις που
  - Η BCNF δε διατηρεί εξαρτήσεις
  - Είναι σημαντικός ο αποδοτικός έλεγχος καταπάτησης των συναρτησιακών εξαρτήσεων μετά από ενημερώσεις
- Λύση: Χαλαρώνουμε περιορισμούς με 3NF
  - Επιτρέπει κάποιον πλεονασμό αλλά
  - Οι συναρτησιακές εξαρτήσεις μπορούν να ελεγχθούν χωρίς join
  - Υπάρχει πάντα μια αποσύνθεση σε 3NF χωρίς απώλειες και με διατήρηση εξαρτήσεων

3NF

- Ένα σχήμα R είναι σε 3NF αν για όλα τα

$$\alpha \rightarrow \beta \text{ στο } F^{+}$$

ισχύει τουλάχιστον ένα από τα παρακάτω
  - $\alpha \rightarrow \beta$ είναι τετριμμένη
  - α είναι κλειδί της *R*
  - Κάθε γνώρισμα *A* στο $\beta - \alpha$ εμπεριέχεται σε κάποιο υποψήφιο κλειδί της *R*

  (Σημείωση: διαφορετικά γνωρίσματα μπορεί να είναι σε διαφορετικά κλειδιά)
- Μια σχέση σε BCNF είναι πάντα και 3NF
- Η Τρίτη συνθήκη είναι η ελάχιστη χαλάρωση του BCNF για να πετύχουμε διατήρηση εξαρτήσεων

Παράδειγμα

- Example
  - $R = (J, K, L)$
    $F = \{JK \rightarrow L, L \rightarrow K\}$
  - Two candidate keys: JK and JL
  - R is in 3NF

    $JK \rightarrow L$ — JK is a superkey
    $L \rightarrow K$ — K is contained in a candidate key
  - BCNF decomposition has (JL) and (LK)
    - Testing for $JK \rightarrow L$ requires a join

Έλεγχος για 3NF

- Πρέπει να ελέγξουμε μόνο τις εξαρτήσεις στο *F*, όχι στο F^+
- Χρησιμοποιούμε την κλειστότητα γνωρισμάτων για κάθε $\alpha \rightarrow \beta$, αν το α είναι υπερκλειδί
- Αν το α δεν είναι υπερκλειδί πρέπει να δούμε αν εμπεριέχεται σε κάποιο υποψήφιο κλειδί του R
  – Ο έλεγχος για 3NF έχει αποδειχτεί NP-hard
  – Παραδόξως, η αποσύνθεση σε 3NF λύνεται σε πολυωνυμικό χρόνο

Αλγόριθμος αποσύνθεσης σε 3NF

Let F_c be a canonical cover for F;
i := 0;
for each functional dependency $\alpha \rightarrow \beta$ in F_c **do**
if none of the schemas R_j, $1 \leq j \leq i$ contains $\alpha\ \beta$
then begin $i := i + 1$;
$R_i := \alpha\ \beta$ **end**
done
if none of the schemas R_j, $1 \leq j \leq i$ contains a candidate key for R
then begin
$i := i + 1$;
R_i := any candidate key for R;
end
return $(R_1, R_2, ..., R_i)$

Αλγόριθμος αποσύνθεσης σε 3NF

- Ο αλγόριθμος εγγυάται
  - Κάθε σχήμα R_i είναι σε 3NF
  - Η αποσύνθεση διατηρεί τις εξαρτήσεις και δεν έχει απώλειες

Example

- Relation schema:

  Banker-info-schema = (branch-name, customer-name,
  banker-name, office-number)

- The functional dependencies for this relation schema are:

  banker-name $\rightarrow$ *branch-name office-number*

  customer-name branch-name $\rightarrow$ *banker-name*

- The key is: {*customer-name*, *branch-name*}
- The **for** loop in the algorithm causes us to include the following schemas in our decomposition:

  Banker-office-schema = (banker-name, branch-name, office-number)

  Banker-schema = (customer-name, branch-name, banker-name)

- Since *Banker-schema* contains a candidate key for *Banker-info-schema*, we are done with the decomposition process.

Σύγκριση BCNF και 3NF

- Είναι πάντα εφικτό να αποσυνθέσεις μια σχέση σε σχέσεις 3NF ώστε
  – Η αποσύνθεση να μην έχει απώλειες
  – Να διατηρούνται οι εξαρτήσεις
- Είναι πάντα εφικτό να αποσυνθέσεις μια σχέση σε σχέσεις BCNF ώστε
  – Η αποσύνθεση να μην έχει απώλειες

Comparison of BCNF and 3NF (Cont.)

- Example of problems due to redundancy in 3NF
  - $R = (J, K, L)$
    $F = \{JK \rightarrow L, L \rightarrow K\}$

J	L	K
j_1	l_1	k_1
j_2	l_1	k_1
j_3	l_1	k_1
null	l_2	k_2

A schema that is in 3NF but not in BCNF has the problems of

- repetition of information (e.g., the relationship l_1, k_1)
- need to use null values (e.g., to represent the relationship l_2, k_2 where there is no corresponding value for J).

Design Goals

- Στόχος μιας σχεσιακής βάσης είναι
  - BCNF.
  - Αποσύνθεση χωρίς απώλειες
  - Διατήρηση εξαρτήσεων
- Αν δεν μπορούμε να το επιτύχουμε τότε ένα από τα 2
  - BCNF χωρίς διατήρηση εξαρτήσεων
  - 3NF
- Η SQL δεν υποστηρίζει άμεσα λειτουργικές εξαρτήσεις άλλες από τα υπερκλειδιά

  Έμμεσα υποστηρίζονται μέσω assertions
- Οπότε ακόμα κι αν είχαμε αποσύνθεση με διατήρηση εξαρτήσεων δε θα μπορούσαμε να ελέγξουμε αποδοτικά με SQL μια εξάρτηση της οποίας το αριστερό μέρος δεν είναι κλειδί

Επιπρόσθετες Εξαρτήσεις

- Με τη χρήση των FD μπορούμε να φτάσουμε μέχρι 3NF και BCNF

- Άλλες εξαρτήσεις: ***πλειότιμες εξαρτήσεις* (Multivalued) (MVD), *εξαρτήσεις συνένωσης* (Join Dependencies), *εξαρτήσεις εγκλεισμού* (Inclusion Dependencies), ...**
Αυτές οι εξαρτήσεις μας πάνε σε κανονικές μορφές πέραν των 3NF και BCNF (δηλαδή, 4NF και 5NF)

- ***Πλειότιμες Συναρτήσεις:*** (άτυπος ορισμός) Ένα σύνολο γνωρισμάτων Χ *καθορίζει πολλαπλά* ένα σύνολο Υ **αν η τιμή του** Χ **καθορίζει ένα σύνολο τιμών για το** Υ
(*ανεξάρτητα από τα άλλα γνωρίσματα στη Σχέση*)

Πλειότιμες Εξαρτήσεις

- Μια MVD παρουσιάζεται ως $X \rightarrow \rightarrow Y$
- Υπάρχουν ***βάσιμα και πλήρη*** σύνολα κανόνων συμπερασμού για MVDs
- Μια FD είναι ειδική περίπτωση των MVD
- Ένα Σχεσιακό Σχήμα R είναι στην ***τέταρτη κανονική μορφή (4NF)*** σε σχέση με ένα σύνολο συναρτησιακών και πλειότιμων εξαρτήσεων F αν για κάθε μη τετριμμένη πλειότιμη εξάρτηση $X \rightarrow \rightarrow Y$ στην F^+, το X είναι υποψήφιο κλειδί του R
- Δεν υπάρχει **αποδοτικός αλγόριθμος** για την αποσύνθεση του R σε 4NF Σχέσεις, έτσι ώστε η αποσύνθεση να μην έχει απώλειες σε σχέση με ένα σύνολο F από FD και MVD στο R

Εξαρτήσεις Συνένωσης

- Μια ***εξάρτηση συνένωσης*** JD(R_1, R_2, R_3, ... R_m) είναι ένας δομικός περιορισμός στην R που προσδιορίζει ότι *κάθε νόμιμο στιγμιότυπο* r(R) **πρέπει να έχει μια lossless join αποσύνθεση** στα R_1, R_2, R_3, ... R_m
- Μια MVD είναι ειδική περίπτωση των JD (όπου m=2)
- Ένα Σχεσιακό Σχήμα R είναι στην ***πέμπτη κανονική μορφή (5NF)*** σε σχέση με ένα σύνολο συναρτησιακών, πλειότιμων και εξαρτήσεων συνένωσης F αν για κάθε μη τετριμμένη JD(R_1, R_2, ... R_m) στην F^+, κάθε R_i είναι υπέρ-κλειδί του R
- Η 5NF ονομάζεται επίσης PJNF (Project-Join Normal Form)

Εξαρτήσεις Εγκλεισμού

- FD, MVD και JD ορίζονται **εντός του ιδίου Σχεσιακού Σχήματος** R (δεν συσχετίζουν γνωρίσματα που βρίσκονται σε διαφορετικές σχέσεις)
- Υπάρχουν και άλλες εξαρτήσεις, όπως οι *εξαρτήσεις εγκλεισμού* που χρησιμοποιούνται για την παράσταση της **αναφορικής ακεραιότητας και των ιεραρχιών class / subclass μεταξύ δύο σχέσεων** R και S
- Μια ***εξάρτηση εγκλεισμού*** **R.X < S.Y** προσδιορίζει ότι σε κάθε χρονικό σημείο, αν τα r(R) και s(S) είναι στιγμιότυπα σχέσεων των R και S, τότε $\Pi_X(\mathbf{r(R)}) \subseteq \Pi_Y(\mathbf{s(S)})$

Πρακτικά Θέματα για την Κανονικοποίηση

- 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 - **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 υποψήφιο κλειδί αποτελείται ακριβώς από ένα γνώρισμα, τότε είναι και σε 5NF* (Fagin, 1991)

ΣΧΟΛΙΑ

- Η διαδικασία Κανονικοποίησης έχει και **μειονεκτήματα**:
  - 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, μια πρακτικά χρήσιμη προσπάθεια να γίνουν με τυπικό και συστηματικό τρόπο πράγματα που τα κάνουμε συνήθως διαισθητικά.