Συναρτησιακές Εξαρτήσεις

- Μία **συναρτησιακή εξάρτηση** (*functional dependency*) είναι ένας περιορισμός μεταξύ δύο συνόλων γνωρισμάτων.

  Εστω $A_1, A_2, \ldots, A_n$ όλα τα γνωρίσματα μιας σχέσης R. Αν X και Y είναι υποσύνολα του $\{A_1, A_2, \ldots, A_n\}$, τότε η συναρτησιακή εξάρτηση $X \rightarrow Y$ καθορίζει ότι για οποιεσδήποτε δύο πλειάδες t_1, t_2 της R, αν $t_1[X] = t_2[X]$, τότε πρέπει επίσης να ισχύει ότι $t_1[Y] = t_2[Y]$.

- Δεδομένης μιας συναρτησιακής εξάρτησης $X \rightarrow Y$, λέμε ότι το σύνολο X προσδιορίζει συναρτησιακά το σύνολο Y ή ότι το σύνολο Y εξαρτάται συναρτησιακά από το σύνολο X.

- Αν η συναρτησιακή εξάρτηση $X \rightarrow Y$ ισχύει σε μια σχέση R, τότε δεν είναι δυνατόν να υπάρχουν πλειάδες οι οποίες συμφωνούν στις τιμές όλων των γνωρισμάτων στο X και συγχρόνως δεν συμφωνούν στην τιμή κάποιου από τα γνωρίσματα του Y.
- Αν το X είναι κλειδί της R, τότε η εξάρτηση $X \rightarrow Y$ ισχύει για κάθε υποσύνολο Y των γνωρισμάτων της R.
- Η εξάρτηση $X \rightarrow Y$ δεν συνεπάγεται την $Y \rightarrow X$.

Παράδειγμα: Εστω ότι οι παρακάτω σχέσεις έχουν το περιεχόμενο που φαίνεται στους αντίστοιχους πίνακες. Προσδιορίστε τις συναρτησιακές εξαρτήσεις οι οποίες ισχύουν.

T1

A	B
x_1	y_1
x_2	y_2
x_3	y_1
x_4	y_1
x_5	y_2
x_6	y_2

T2

A	B
x_1	y_1
x_2	y_4
x_1	y_1
x_3	y_2
x_2	y_4
x_4	y_3

T3

A	B
x_1	y_1
x_2	y_4
x_1	y_1
x_3	y_2
x_2	y_4
x_4	y_4

T1: $A \rightarrow B$
$B \not\rightarrow A$

T2: $A \rightarrow B$
$B \rightarrow A$

T3: $A \rightarrow B$
$B \not\rightarrow A$

Παράδειγμα: Συναρτησιακές Εξαρτήσεις στο σχήμα **emp_info**

1. **emp_id** → **emp_name emp_phone dept_name**
2. **dept_name** → **dept_phone dept_mgrname**
3. **skill_id** → **skill_name**
4. **emp_idskill_id** → **skill_date skill_lvl**

Λογικές Συνέπειες Συναρτησιακών Εξαρτήσεων

- Κανόνας Εγκλεισμού (inclusion rule):

  Αν X, Y είναι σύνολα γνωρισμάτων από το σχήμα της σχέσης R και $Y \subseteq X$, τότε $X \rightarrow Y$.

- Μια συναρτησιακή εξάρτηση $X \rightarrow Y$ λέγεται τετριμμένη αν ισχύει για κάθε σχέση R της οποίας το σχήμα περιέχει X και Y.

- Τετριμμένες εξαρτήσεις εμφανίζονται σαν αποτέλεσμα της εφαρμογής του κανόνα εγκλεισμού.

  Θεώρημα: Αν $X \rightarrow Y$ είναι τετριμμένη συναρτησιακή εξάρτηση, πρέπει να ισχύει ότι $Y \subseteq X$.

Απόδειξη: Υποθέστε ότι $Y \supset X$. Δημιοθργήστε μια σχέση με όλα τα γνωρίσματα των X και Y και θεωρήστε ένα γνώρισμα A του $Y - X$. Εφόσον $A \in Y$ και $A \notin X$, είναι δυνατόν να κατασκευάσομε δύο πλειάδες u και v, οι οποίες έχουν κοινές τιμές σε όλα τα γνωρίσματα στο X αλλά έχουν δοαφορετικές τιμές στο A. Τότε όμως η τετριμμένη εξάρτηση δεν ισχύει. Αρα, δεν μπορεί να υπάρχει τέτοιο γνώρισμα A στο $Y - X$. Επομένως, $Y \subseteq X$.

- Από ένα μικρό αριθμό κανόνων συναπαγωγής και ένα αρχικό σύνολο συναρτησιακών εξαρτήσεων μπορεί να εξαχθεί ένας αριθμός πρόσθετων συναρτησιακών εξαρτήσεων.

Αξιώματα Armstrong

Εστω ότι τα σύνολα γνωρισμάτων X, Y, Z περιέχονται στο σχήμα της σχέσης R. Τότε ισχύουν οι παρακάτω κανόνες:

1. **Κανόνας Εγκλεισμού** : αν $Y \subseteq X$, τότε $X \rightarrow Y$
2. **Κανόνας Μεταβατικότητας** : αν $X \rightarrow Y$ και $Y \rightarrow Z$, τότε $X \rightarrow Z$
3. **Κανόνας Επαύξησης**: αν $X \rightarrow Y$, τότε $XZ \rightarrow YZ$

Συνέπειες των Αξιωμάτων

Θεώρημα: Αν W, X, Y, Z, B περιέχονται στο σχήμα της R, τότε

1. **Κανόνας Ενωσης**: αν $X \rightarrow Y$ και $X \rightarrow Z$, τότε $X \rightarrow YZ$
2. **Κανόνας Αποσύνθεσης**: αν $X \rightarrow YZ$, τότε $X \rightarrow Y$ και $X \rightarrow Z$
3. **Κανόνας Ψευδομεταβατικότητας**: αν $X \rightarrow Y$ και $WY \rightarrow Z$, τότε $XW \rightarrow Z$
4. **Κανόνας Συσσώρευσης**: αν $X \rightarrow YZ$ και $Z \rightarrow B$, τότε $X \rightarrow YZB$

Παράδειγμα: Βρείτε ένα ελάχιστο σύνολο συναρτησιακών εξαρτήσεων οι οποίες ικανοποιούνται στον ακόλουθο πίνακα:

T

A	B	C	D
a_1	b_1	c_1	d_1
a_1	b_1	c_2	d_2
a_2	b_1	c_1	d_3
a_2	b_1	c_3	d_4

1. Σ.Ε. με ένα γνώρισμα στο αριστερό μέλος
   - Οι τετριμμένες εξαρτήσεις $A \rightarrow A, B \rightarrow B, C \rightarrow C, D \rightarrow D$ δεν περιλαμβάνονται στο ελάχιστο σύνολο.
   - $A \rightarrow B, C \rightarrow B, D \rightarrow B$ καθώς όλες οι τιμές του B είναι ίδιες
   - Τα γνωρίσματα A, C, D έχουν τουλάχιστον δύο διακεκριμένες τιμές. Αρα, $B \not\rightarrow A, B \not\rightarrow C$ και $B \not\rightarrow D$

- Ολες οι τιμές του D είναι διαφορετικές. Αρα, $D \rightarrow A, D \rightarrow B, D \rightarrow C$.
- Τα γνωρίσματα A, B, C έχουν τουλάχιστον δύο επαναλαμβανόμενες τιμές. Αρα, $A \not\rightarrow D, B \not\rightarrow D, C \not\rightarrow D$
- $A \not\rightarrow C$ και $C \not\rightarrow A$, εξ'αιτίας των πλειάδων 1,2 και 1,3 αντίστοιχα.
- Αρα, ισχύουν οι ακόλουθες Σ.Ε.: $A \rightarrow B, C \rightarrow B, D \rightarrow A, D \rightarrow B, D \rightarrow C$
- Από τον κανόνα ένωσης: $A \rightarrow B, C \rightarrow B, D \rightarrow ABC$

2. Σ.Ε. με ζεύγος γνωρισμάτων στο αριστερό μέλος
    - Εξ'αιτίας της $D \rightarrow ABC$, κάθε ζεύγος που περιέχει το D προσδιορίζει συναρτησιακά όλα τα υπόλοιπα γνωρίσματα (κανόνας επαύξησης). Αυτές οι εξαρτήσεις είναι συνεπαγωγές εξαρτήσεων που ήδη ανήκουν στο ζητούμενο σύνολο.

- Ζεύγη γνωρισμάτων που περιλαμβάνουν το B στο αριστερό μέλος δίνουν είτε τετριμμένες εξαρτήσεις είτε συνεπαγωγές εξαρτήσεων.
- $AC \rightarrow ABCD$ καθώς το ζεύγος AC έχει διακεκριμένες τιμές σε κάθε πλειάδα. Η μόνη νέα εξάρτηση είναι η $AC \rightarrow D$. Οι εξαρτήσεις $AC \rightarrow A, AC \rightarrow C, AC \rightarrow B$ ήδη εξάγονται από άλλες εξαρτήσεις.

3. Δεν υπάρχουν εξαρτήσεις με 3 ή 4 γνωρίσματα στο αριστερό μέλος οι οποίες ανήκουν σε αυτό το σύνολο.

Το ελάχιστο σύνολο εξαρτήσεων είναι
$\{A \rightarrow B, C \rightarrow B, D \rightarrow ABC, AC \rightarrow D\}$.

Κλείσιμο, Κάλυψη, Ελάχιστη Κάλυψη

- Οι κανόνες συνεπαγωγής ΣΕ παράγουν ένα σύνολο το οποίο είναι πολύ μεγαλύτερο από το αρχικό σύνολο των ΣΕ.
- Δεδομένου ενός συνόλου F από ΣΕ, το **κλείσιμο** (*closure*) F^+ του F ορίζεται σαν το σύνολο των ΣΕ οι οποίες συνεπάγονται από το F.
- **Παράδειγμα**: Θεωρείστε το σύνολο των ΣΕ:

  $F = \{A \to B, B \to C, C \to D, D \to E, E \to F, F \to G, G \to H\}$

  Από $A \to B$ και $B \to C$, συνεπάγεται λόγω μεταβατικότητας η $A \to C$.

  Επίσης: $A \to D, A \to E, A \to F, A \to G, A \to BC, A \to EF$ etc.

  Συνολικά, το F^+ θα περιέχει $2^8 - 1$ FDs με A στο αριστερό μέλος.

  Παρόμοια για εξαρτήσεις με B στο αριστερό μέλος.

- Το μέγεθος του κλεισίματος ενός συνόλου ΣΕ μεγαλώνει εκθετικά με αυτό του αρχικού συνόλου.
- Χρειαζόμαστε εναν τρόπο για να αναφερόμαστε στο σύνολο των εξαρτήσεων που συνεπάγονται από ένα αρχικό σύνολο χωρίς να πρέπει να υπολογίσομε το κλείσιμο του συνόλου αυτού.
- Για κάθε σύνολο ΣΕ μπορούμε να βρούμε ένα "ισοδύναμο" σύνολο το οποίο είναι ελάχιστο.
- Ενα σύνολο F από ΣΕ για μια σχέση R **καλύπτει** ένα άλλο σύνολο G από ΣΕ για την R, αν το σύνολο G μπορεί να εξαχθεί με την εφαρμογή των κανόνων συνεπαγωγής στις ΣΕ του F, δηλαδή, αν $G \subseteq F^+$.

- Αν το F καλύπτει το G και το G καλύπτει το F τότε τα F και G λέγονται ισοδύναμα ($F \equiv G$).
- Αν $F \equiv G$ τότε $F^+ = G^+$
- **Παράδειγμα:** Θεωρείστε τα σύνολα ΣΕ

  $F = \{B \rightarrow CD, AD \rightarrow E, B \rightarrow A\}$

  $G = B \rightarrow CDE, B \rightarrow ABC, AD \rightarrow E\}$

  Δείξτε ότι το F καλυπτει το G.

  Πρέπει να δείξομε ότι κάθε ΣΕ του G μπορεί να εξαχθεί από το F με χρήση των κανόνων συνεπαγωγής.

  – Η ΣΕ $AD \rightarrow E$ είναι ήδη στο F.

  – Από την $B \rightarrow CD$ και την $B \rightarrow A$, εξάγομε την $B \rightarrow ACD$ (κανόνας ένωσης).
    Από $B \rightarrow ACD$ και $B \rightarrow B$ εξάγομε την $B \rightarrow ABCD$ (κανόνας ένωσης).

Από $B \rightarrow ABCD$ εξάγομε την $B \rightarrow AD$ (κανόνας αποσύνθεσης).
Από $B \rightarrow AD$ και $AD \rightarrow E$ εξάγομε την $B \rightarrow E$ (αξίωμα μεταβατικότητας).
Από $B \rightarrow ABCD$ και $B \rightarrow E$ εξάγομε την $B \rightarrow ABCDE$ (κανόνας ένωσης).
Από $B \rightarrow ABCDE$ εξάγομε την $B \rightarrow CDE$ και την $B \rightarrow ABC$ (κανόνας αποσύνθεσης).

- Ο υπολογισμός του κλεισίματος ενός συνόλου από ΣΕ μας δίνει τα σύνολα όλων των γνωρισμάτων τα οποία προσδιορίζονται συναρτησιακά από άλλα σύνολα γνωρισμάτων.
- Ο υπολογισμός αυτός έχει κόστος εκθετικό ως προς το μέγεθος του αρχικού συνόλου των ΣΕ.
- Δεδομένου ενός συνόλου F από ΣΕ σε μια σχέση R και ενός συνόλου X από γνωρίσματα του σχήματος της R, το κλείσιμο του X, X^+ είναι το μέγιστο σύνολο γνωρισμάτων Y, για τα οποία $X \rightarrow Y \ \in F^+$.
- Ο υπολογισμός του κλεισίματος ενός συνόλου γνωρισμάτων έχει κόστος πολυωνυμικό ως πρός το μέγεθος του συνόλου αυτού (πιο αποδοτικός από τον υπολογισμό κλεισίματος συνόλου ΣΕ.)

- Ο ακόλουθος αλγόριθμος υπολογίζει το X^+ δεδομένου ενός συνόλου ΣΕ F.

$I := 0; X[I] := X;$
Repeat
    $I := I + 1;$
    $X[I] := X[I-1];$
    For all $Z \rightarrow W$ in F
        If $Z \subseteq X[I]$ then
            $X[I] := X[I] \cup W;$
Until $X[I] = X[I-1];$
Return $X^+ = X[I];$

- Ο υπολογισμός του X^+ βασίζεται στην εφαρμογή του κανόνα συσσώρευσης.

Παράδειγμα: $X = B$, $F = \{B \rightarrow CD, AD \rightarrow E, B \rightarrow A\}$

Initially: $X[0] = B$

In the repeat-loop:

$X[1] = B$

In the for-loop:

$X[1] = BCD$ (because of $B \rightarrow CD$)

$X[1] = ABCD$ (because of $B \rightarrow A$)

$X[2] = ABCD$

In the for-loop:

$X[2] = ABCDE$ (because of $AD \rightarrow E$)

$X[3] = ABCDE$

After the for-loop: $X[3] = X[2]$

Return $X^+ = ABCDE$

- Χρειαζόμαστε μια μέθοδο κατασκευής μιας κάλυψης για ένα σύνολο από ΣΕ, και επιπλέον η κάλυψη πρέπει να είναι ελάχιστη.
- Ο ακόλουθος αλγόριθμος κατασκευάζει μια ελάχιστη κάλυψη M ενός συνόλου ΣΕ.

1. Δημιουργούμε ένα ισοδύναμο σύνολο H από ΣΕ με ένα μόνο γνώρισμα στο δεξί μέλος.

   $H := \emptyset$;
   For all $X \rightarrow Y$ in F
       For all A in Y
           $H := H \cup \{X \rightarrow A\}$;

   Στο τέλος του 1ου βήματος, $H \equiv F$

2. Αφαιρούμε από το H τις ΣΕ οι οποίες αν αφαιρεθούν δεν επηρεάζουν το H^+.

For all $X \to A$ in H

$J := H - \{X \to A\};$

Dete mine X^+ unde J;

If $A \in X^+$ then

$H := J;$

Το 2ο βήμα μετατρέπει το H σε ένα μικρότερο αλλά ισοδύναμο σύνολο.

3. Αντικατέστησε ΣΕ με άλλες οι οποίες έχουν λιγότερα γνωρίσματα στο αριστερό μέλος εφόσον δεν επηρεάζεται το H^+.

For all $X \rightarrow A$ in H
    For all $B \in X$
        $Y := X - \{B\};$
        $J := (H - \{X \rightarrow A\}) \cup \{Y \rightarrow A\};$
        Generate Y^+ under J and Y^+ under H;
        If Y^+ under $J = Y^+$ under H then
            $H := J;$

Καθώς το H^+ δεν αλλάζει, το σύνολο που προκύπτει είναι ισοδύμανο με το αρχικό.

4. Εφάρμοσε τον κανόνα της ένωσης στις ΣΕ με κοινό αριστερό μελος. Υποθέτομε ότι όλες οι ΣΕ είναι "unmarked" στην αρχή αυτού του βήματος.

$M := \emptyset$;
For all $X \rightarrow A$ in H
    If marked, continue;
    Mark $X \rightarrow A$;
    $Y := \{A\}$;
    For all successive $X \rightarrow B$ in H
        Mark $X \rightarrow B$;
        $Y := Y \cup \{B\}$;
    $M := M \cup \{X \rightarrow Y\}$;
Return M;

Το αποτέλεσμα είναι η ελάχιστη κάλυψη M.

Παράδειγμα: Κατασευάστε την ελάχιστη κάλυψη M για το σύνολο $F = \{A \rightarrow AC, B \rightarrow ABC, D \rightarrow ABC\}$

1. $H = \{A \rightarrow A, A \rightarrow C, B \rightarrow A, B \rightarrow B, B \rightarrow C, D \rightarrow A, D \rightarrow B, D \rightarrow C\}$

2. (a) $A \rightarrow A$ τετριμμένη - μπορεί να αφαιρεθεί

   (b) $A \rightarrow C$ δεν μπορεί να αφαιρεθεί καθώς δεν υπάρχει άλλη ΣΕ με A στο αριστερό μέλος

   (c) $B \rightarrow A$ δεν μπορεί να αφαιρεθεί καθώς $B^+ = BC$ για το $J = H - \{B \rightarrow A\}$

   (d) $B \rightarrow B$ τετριμμένη - μπορεί να αφαιρεθεί

   (e) $B \rightarrow C$ μπορεί να αφαιρεθεί καθώς $B^+ = ABC$ για το $J = H - \{B \rightarrow C\}$

   (f) $D \rightarrow A$ μπορεί να αφαιρεθεί καθώς $D^+ = DBA$ για το $J = H - \{D \rightarrow A\}$

(g) $D \to B$ δεν μπορεί να αφαιρεθεί καθώς $D^+ = DC$ για το $J = H - \{D \to B\}$

(h) $D \to C$ μπορεί να αφαιρεθεί καθώς $D^+ = DBAC$ για το $J = H - \{D \to C\}$

Μετά το βήμα 2, $H = \{A \to C, B \to A, D \to B\}$

3. Τα βήματα 3 και 4 δεν μεταβάλλουν το H.

Αρα, $M = \{A \to C, B \to A, D \to B\}$

Example: Consider a relation R(A,B,C,D,E,G) containing data about factories, with the attributes

A - manager id,	D - employee salary
B - factory name,	E - employee taxes
C - employee id,	G - employee bonus

The data satisfy the following constrains:

- Every manager manages one factory, but a factory can have several managers.

- Every employee works for one factory, but a factory may have many employees.

- The taxes of an employee are determined by the salary

- The bonus of an employee is determined be the salary and the factory policy.

Thus, the set of dependencies is:

$A \rightarrow B$, $C \rightarrow BD$, $C \rightarrow G$, $CD \rightarrow E$, $BCDE \rightarrow G$

The minimal cover for this set of dependencies is:

$A \rightarrow B$, $C \rightarrow BD$, $D \rightarrow E$, $E \rightarrow G$

i.e.,

manager_id → factory_name
employee_id → factory_name, employee_salary
employee_salary → employee_taxes
employee_taxes → employee_bonus