



Ειδικά Θέματα Βάσεων Δεδομένων

Ενότητα 3: Συναρτησιακές εξαρτήσεις και κανονικοποιήσεις

Δρ. Τσιμπίρης Αλκιβιάδης

Τμήμα Μηχανικών Πληροφορικής ΤΕ



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.



Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «Ανοικτά Ακαδημαϊκά Μαθήματα στο ΤΕΙ Κεντρικής Μακεδονίας» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.



Ενότητα 3

Συναρτησιακές εξαρτήσεις και κανονικοποιήσεις

Δρ. Τιμπίρης Αλκιβιάδης

Περιεχόμενα ενότητας

- Συναρτησιακές εξαρτήσεις
- Μεταβατική συναρτησιακή εξάρτηση
- Κλειδιά και μοναδικότητα
- Κανόνες συναγωγής- Αξιώματα του Armstrong
- Κανονικές μορφές 1KM,2KM,3KM
- BC-KM, 4KM, 5KM
- Άλλες KM
- Από-κανονικοποίηση

Σκοποί ενότητας

Στο κεφάλαιο αυτό θα οριστούν οι συναρτησιακές εξαρτήσεις (τετριμμένες και μη) στις σχεσιακές βάσεις δεδομένων, καθώς και οι κανόνες συναγωγής ή αξιώματα του Armstrong. Η θεωρία της κανονικοποίησης αποτελεί σημαντική ενότητα για την δημιουργία συμπαγών και ακέραιων βάσεων για αυτό θα αναλυθούν όλες οι κανονικές μορφές που μπορούν να εφαρμοσθούν ξεκινώντας από την πρώτη κανονική μορφή και φτάνοντας μέχρι και την κανονική μορφή πεδίου ορισμού/κλειδιού.

Συναρτησιακές εξαρτήσεις

Υποθετικά υπάρχει μία σχέση A και τα B και Δ είναι τυχαία υποσύνολα του συνόλου των γνωρισμάτων A , σε μία τέτοια περίπτωση το Δ είναι συναρτησιακά εξαρτημένο από το B ($B \rightarrow \Delta$) όμως αυτό ισχύει εάν η κάθε τιμή του B αντιστοιχεί σε μία ακριβώς τιμή του Δ . Ποιό κατανοητά δύο συστοιχίες του συνόλου όταν συμφωνούν στην μία τιμή συμφωνούν και στην άλλη.

Πίνακας A

B		Δ		
Υποσύνολο 1	Υποσύνολο 1	Υποσύνολο 3	Υποσύνολο 4	Υποσύνολο 5

Συναρτησιακές εξαρτήσεις

- Στις συναρτησιακές εξαρτήσεις εμπεριέχεται ο όρος εξάρτηση στον οποίο επίσης αποδίδονται δύο όροι **dependence (εξάρτηση)** και **dependency (εξαρτημένο αντικείμενο)**.
- οι συναρτησιακές εξαρτήσεις εμπεριέχουν ενδιαφέρουσες τυπικές ιδιότητες και αντιμετώπιση των προβλημάτων με τυπικό και αυστηρό τρόπο.
- Μια συναρτησιακή εξάρτηση είναι μία συσχέτιση πολλά προς πολλά.
- Στις συναρτησιακές σχέσεις ισχύει ότι εάν το B είναι υποψήφιο κλειδί μίας σχέσης A (ή το πρωτεύον κλειδί) τότε όλα τα γνωρίσματα της Δ πρέπει να είναι συναρτησιακά εξαρτημένα από το B. Επίσης εάν ισχύει η συναρτησιακή εξάρτηση ($B \rightarrow \Delta$) και το B δεν είναι υποψήφιο κλειδί τότε η σχέση A έχει κάποιο πλεονασμό.

Συναρτησιακές εξαρτήσεις

Παράδειγμα:

Στη σχέση Φοιτητές το όνομα του φοιτητή και το ΑΕΜ του είναι συναρτησιακά εξαρτώμενο από το ΑΕΜ. Αυτό γίνεται γιατί Το ΑΕΜ προσδιορίζει μοναδικά τον φοιτητή δεν μπορούν να έχουν δύο φοιτητές το ίδιο ΑΕΜ. Για τον λόγο αυτό το ΑΕΜ ονομάζεται προσδιοριστικό (ή ορίζουσα).

Το σύμβολο \rightarrow σημαίνει
προσδιορίζεται
συναρτησιακά.

$\text{ΑΕΜ} \rightarrow \text{Όνομα} _ \text{Φοιτητή} = \text{Το ΑΕΜ προσδιορίζει μοναδικά το όνομα του φοιτητή.}$

Συναρτησιακές εξαρτήσεις

Τετριμμένες και μη τετριμμένες εξαρτήσεις:

- **Τετριμμένη** λέγεται μία συναρτησιακή εξάρτηση εάν και μόνο εάν το δεξιό μέλος είναι υποσύνολο του αριστερού. Επίσης λέγεται τετριμμένη όταν δεν ικανοποιείται .
- **μη τετριμμένες** (nontrivial) εξαρτήσεις είναι "γνήσιες" δεσμεύσεις ακεραιότητας. Στην τυπική θεωρία των εξαρτήσεων, είναι αδύνατο να θεωρηθεί ότι οι εξαρτήσεις είναι μη τετριμμένες.

Μεταβατική συναρτησιακή εξάρτηση

Οι μεταβατικές συναρτησιακές εξαρτήσεις έχουν τις εξής ιδιότητες:

- Έστω μια σχέση A που περιέχει κάποια γνωρίσματα (B, Γ, Δ) τέτοια ώστε να ισχύουν οι παρακάτω συναρτησιακές εξαρτήσεις: $B \rightarrow \Gamma, \Gamma \rightarrow \Delta$ και αφού ισχύουν αυτές να ισχύει και η $B \rightarrow \Delta$.
- Στο παραπάνω παράδειγμα η $B \rightarrow \Delta$ είναι ένα παράδειγμα μεταβατικής συναρτησιακής εξάρτησης και σημαίνει ότι το Δ εξαρτάται από το B μεταβατικά μέσω του Γ .

Κλειδιά και μοναδικότητα

Το **κλειδί** είναι μία συστοιχία που προσδιορίζει μοναδικά ένα ή περισσότερα γνωρίσματα:

- Προσδιορίζει συναρτησιακά μια συνάρτηση.
- Δεν είναι όλα προσδιοριστικά κλειδιά.

Κανόνες συναγωγής- Αξιώματα του Armstrong

Υποθετικά θεωρήστε ότι τα B , Γ , Δ είναι τυχαία υποσύνολα ενός συνόλου με γνωρίσματα της σχέσης A και ότι το $B\Gamma$ συμβολίζει την ένωση αυτών των δύο τότε:

- Το B είναι υποσύνολο του Γ ($B \rightarrow \Gamma$). **Ανελαστικότητα.**
- Εάν $B \rightarrow \Gamma$ τότε ισχύει $B\Delta \rightarrow \Gamma\Delta$. **Επαύξηση.**
- Εάν $B \rightarrow \Gamma$ και $\Gamma \rightarrow \Delta$ τότε $B \rightarrow \Delta$. **Μεταβατικότητα.**

Κανόνες συναγωγής- Αξιώματα του Armstrong

Από τους κανόνες ανελαστικότητα, επαύξηση και μεταβατικότητα προκύπτουν και οι εξής:

- Αυτοκαθορισμός
- Ανάλυση
- Ένωση
- Σύνθεση

Θεωρία Κανονικοποίησης

Μία σχέση είναι ένα σύνολο από γνωρίσματα με τιμές για κάθε γνώρισμα τέτοιες ώστε να ισχύουν οι παρακάτω ιδιότητες

1. Κάθε όνομα γνωρίσματος είναι μοναδικό.
2. Όλες οι τιμές κάθε γνωρίσματος είναι ίδιου τύπου (ή πεδίου ορισμού).
3. Κάθε τιμή γνωρίσματος είναι ατομική (μία τιμή και όχι ομάδα πολλών τιμών).
4. Τα γνωρίσματα δεν έχουν διάταξη από τα αριστερά προς τα δεξιά..
5. Οι συστοιχίες (σειρές) δεν έχουν διάταξη από επάνω προς τα κάτω.
6. Δεν υπάρχουν δύο ίδιες σειρές (συστοιχίες) σε μία σχέση.

Θεωρία Κανονικοποίησης

- Η διαδικασία που ακολουθούμε είναι :
- 1. Συγκεντρώνουμε τις απαιτήσεις της επιχείρησης και των χρηστών.
- 2. Σχεδιάζουμε το μοντέλο οντοτήτων-συσχετίσεων
- 3. Μετατρέπουμε το διάγραμμα οντοτήτων-συσχετίσεων της επιχείρησης σε ένα σύνολο από σχέσεις (πίνακες) με το σχεσιακό μοντέλο.
- 4. Κανονικοποιούμε τις σχέσεις για να απομακρύνουμε τυχόν ανωμαλίες ενημέρωσης-διαγραφής-εισαγωγής στοιχείων.
- 5. Υλοποιούμε τη βάση δεδομένων δημιουργώντας ένα πίνακα για κάθε κανονικοποιημένη σχέση.

Κανονικοποίηση-Κανονικές μορφές

- Κατηγορίες (ή κλάσεις) **Κανονικές Μορφές** (normal forms).
 - **Κανονική Μορφή**: Μία κλάση σχέσεων απαλλαγμένων από συγκεκριμένα προβλήματα τροποποιήσεων.
1. Πρώτη κανονική μορφή (1NF – 1KM)
 2. Δεύτερη κανονική μορφή (2NF – 2KM)
 3. Τρίτη κανονική μορφή (3NF – 3KM)
 4. Boyce-Codd κανονική μορφή (BCNF – KM BC)
 5. Τέταρτη κανονική μορφή (4NF – 4KM)
 6. Πέμπτη κανονική μορφή (5NF – 5KM)
 7. Κανονική μορφή πεδίου ορισμού κλειδιού (Domain-Key/NF)

Κανονικοποίηση-Κανονικές μορφές

- Αυτές οι κανονικές μορφές είναι αθροιστικές. Μία σχέση που βρίσκεται σε Τρίτη κανονική μορφή είναι επίσης και σε δεύτερη και σε πρώτη.
- Οι τρεις πρώτες κανονικές μορφές (1KM, 2KM, 3KM) ορίστηκαν από τον Codd. Όλες οι κανονικοποιημένες σχέσεις είναι σε 1KM.
- Με άλλα λόγια, "κανονικοποιημένη" και "σε 1KM" σημαίνει ακριβώς το ίδιο πράγμα.
- Μερικές σχέσεις 1KM είναι επίσης σε 2KM, και μερικές σχέσεις 2KM είναι επίσης σε 3KM.

Κανονικοποίηση-Κανονικές μορφές

- Ο Fagin όρισε την τέταρτη κανονική μορφή.
- Μετέπειτα και πάλι ο Fagin όρισε άλλη μία κανονική μορφή, την κανονική μορφή προβολής–σύζευξης (projection join), που αργότερα έγινε γνωστή και ως πέμπτη κανονική μορφή (5KM). ‘
- Μερικές σχέσεις που είναι σε KM-BC είναι επίσης σε 4KM, και μερικές σχέσεις που είναι σε 4KM είναι επίσης σε 5KM.

Πρώτη Κανονική Μορφή 1KM

- Μία σχέση βρίσκεται σε πρώτη κανονική μορφή αν ικανοποιεί όλες τις 6 ιδιότητες του ορισμού της σχέσης
- Εάν υπάρχει καθορισμένο **κλειδί** για τη σχέση τότε ικανοποιείται η απαίτηση της μοναδικότητας των συστοιχιών (σειρών).
- Ένας πίνακας σε πρώτη κανονική μορφή λέγεται κανονικοποιημένος πίνακας και τότε και μόνο τότε αντιστοιχεί σε μία σχέση (οι σχέσεις του σχεσιακού μοντέλου είναι στην 1η κανονική μορφή).

Δεύτερη κανονική μορφή (2NF)

- Μία σχέση βρίσκεται σε δεύτερη κανονική μορφή εάν κάθε ένα από τα γνωρίσματά της που δεν είναι κλειδιά εξαρτώνται συναρτησιακά από ολόκληρο το πρωτεύων κλειδί και όχι μόνο από ένα τμήμα του.
- Οι σχέσεις που έχουν μόνο **ένα γνώρισμα σαν πρωτεύων κλειδί** βρίσκονται αυτόματα και στη **δεύτερη κανονική μορφή**.
- Αυτός είναι ένας λόγος για τον οποίο χρησιμοποιούμε συχνά τεχνητά αναγνωριστικά σαν κλειδιά.

Τρίτη κανονική μορφή (3NF)

- Μία σχέση βρίσκεται σε Τρίτη κανονική μορφή εάν είναι σε δεύτερη και δεν περιέχει **μεταβατικές εξαρτήσεις**.
- Θεωρήστε για παράδειγμα τη σχέση R που έχει γνωρίσματα τα A, B και Γ. Εάν $A \rightarrow B$ και $B \rightarrow \Gamma$ τότε θα ισχύει και $A \rightarrow \Gamma$.

Κανονική Μορφή Boyce/Codd

- Ο 3KM δεν αντιμετωπίζει περιπτώσεις:
 - 1. Έχει δύο ή περισσότερα κλειδιά
 - 2. Τα δύο υποψήφια κλειδιά να είναι σύνθετα
 - 3. Να επικαλύπτονται (να έχουν τουλάχιστον ένα γνώρισμα κοινό)
-
- κάθε σχέση στην **BCNF** είναι επίσης στην 3NF, αλλά δεν ισχύει πάντα το αντίστροφο.

Κανονική Μορφή Boyce/Codd

- Μία σχέση βρίσκεται σε κανονική μορφή Boyce-Codd εάν κάθε προσδιοριστικό της σχέσης είναι ένα υποψήφιο κλειδί.
- ΚΜ-BC εάν και μόνο εάν τα μόνα ορίζοντα μέλη είναι υποψήφια κλειδιά

Κανονική Μορφή Boyce/Codd

- Οποιαδήποτε σχέση που έχει μόνο δύο γνωρίσματα είναι σε BCNF.
- Η κανονική μορφή Boyce-Codd είναι η μεγαλύτερη που μπορούμε να φτάσουμε μέσω των συναρτησιακών εξαρτήσεων.

Κανονική Μορφή Boyce/Codd

- Πχ.
 - Χρηματιστηριακές Συναλλαγές γίνονται σε πολλούς Τύπους Μετοχών
 - Οι Συναλλαγές διαχειρίζονται από έναν ή περισσότερους Χρηματιστές
 - Οι Τύποι Μετοχών μπορούν έχουν έναν ή πολλούς Χρηματιστές
 - Οι Χρηματιστές μπορούν να συναλλάσσονται σε έναν Τύπο Μετοχών
-
- ΚωδΣυναλλαγής, ΤύποςΜετοχής -> Χρηματιστής
 - ΚωδΣυναλλαγής, Χρηματιστής -> ΤύποςΜετοχής
 - Χρηματιστής -> ΤύποςΜετοχής

Κανονική Μορφή Boyce/Codd

ΚωδΣυναλλαγής	ΤύποςΜετοχής	Χρηματιστής
15-03	Κοινές Μετοχές	Παπάς
15-04	Ομόλογα Δημοσίου	Γιώτης
3-103	Κοινές Μετοχές	Δημητρίου
2-234	Προνομιακές Μετοχές	Δήμου
56-117	Κοινές Μετοχές	Παπάς

Κανονική Μορφή Boyce/Codd

Τι θα συνέβαινε εάν διαγραφόταν η εγγραφή με ΚωδΣυναλλαγής 2-234. Θα χανόταν και το γεγονός ότι ο Δήμου διαχειρίζεται τις Προνομιακές Μετοχές

Βήματα για ΚΜ BC:

- ✓Λίστα με όλα τα γνωρίσματα
- ✓Έλεγχος εάν κάθε γνώριμα μπορεί να είναι υποψήφιο κλειδί
- ✓Για τα γνωρίσματα που δεν είναι κλειδιά δημιουργήστε μία σχέση για την κάθε συναρτησιακή εξάρτηση.
- ✓Συσχετίστε το γνώρισμα με την αρχική σχέση

Λύση

ΠινακαςΑ(Χρηματιστής, Τύπος Μετοχής)

ΠίνακαςΒ(ΚωδΣυναλλαγής, Χρηματιστής)

Κανονική Μορφή Boyce/Codd

- Επίπεδα κανονικοποίησης πλήρη διάταξη με την έννοια ότι σε κάθε σχέση που είναι σε $n+1$ κανονική μορφή είναι αυτόματα και σε n ΚΜ ενώ το αντίστροφο δεν ισχύει
- Η αναγωγή σε ΚΜ-BC είναι πάντα δυνατή. Οποιαδήποτε δεδομένη σχέση μπορεί πάντα να αντικατασταθεί με ένα ισοδύναμο σύνολο σχέσεων σε ΚΜ-BC
- Σκοπός της αναγωγής είναι να αποφευχθεί ο πλεονασμός. Και ανωμαλίες ενημέρωσης

Εξάρτηση Πολλαπλών Τιμών

- R σχέση, A,B,C, τυχαία υποσύνολα του συνόλου των γνωρισμάτων της R.
- Τότε το B είναι πολλαπλά εξαρτημένο με A ή $B \twoheadrightarrow A$

ή αλλιώς το A καθορίζει πολλαπλά το B εάν σύνολο τιμών του B που αντιστοιχούν σε δεδομένο ζεύγος (A,C) στο R, εξαρτάται μόνο από την τιμή A και ανεξάρτητο από την τιμή-C

Εξάρτηση Σύζευξης

- R σχέση, A, B, \dots, Z , τυχαία υποσύνολα του συνόλου των γνωρισμάτων της R. τότε R ικανοποιεί εξάρτηση σύζευξης $*(A, B, \dots, Z)$ εάν και μόνο εάν η R είναι ίση με τη σύζευξη των προβολών της πάνω στα A, B, \dots, Z
- Η τέταρτη κανονική μορφή εξετάζει άλλου είδους εξαρτήσεις, τις εξαρτήσεις πολλαπλών τιμών. Επίσης η πέμπτη κανονική μορφή απομακρύνει εξαρτήσεις σύζευξης.

ΤΕΤΑΡΤΗ Κανονική Μορφή

- Μία σχέση είναι σε 4KM εάν είναι σε BC-NF και δεν περιέχει εξαρτήσεις πολλαπλών τιμών (*multivalued dependencies*)
- 4KM εάν και μόνο εάν οι εξαρτήσεις πολλαπλών τιμών που ικανοποιεί είναι στην πραγματικότητα συναρτησιακές εξαρτήσεις που ξεκινούν από υποψήφια κλειδιά

ΤΕΤΑΡΤΗ Κανονική Μορφή

- Ο εργαζόμενος συμμετέχει σε πολλά έργα.
- Ο εργαζόμενος επιβλέπει περισσότερους του ενός εργαζομένους.

ΠΡΟΪΣΤΑΜΕΝΟΣ	ΕΡΓΟ	ΕΠΙΒΛ-ΕΡΓΑΖΟΜ
ΠΑΠΑΣ	A	ΓΙΑΝΝΗΣ
ΠΑΠΑΣ	B	ΣΟΦΙΑ
ΠΑΠΑΣ	A	ΣΟΦΙΑ
ΠΑΠΑΣ	B	ΓΙΑΝΝΗΣ
ΔΗΜΟΥ	A	ΚΩΣΤΑΣ

ΤΕΤΑΡΤΗ Κανονική Μορφή

•ΛΥΣΗ

ΠΡΟΪΣΤΑΜΕΝΟΣ	ΕΡΓΟ
ΠΑΠΑΣ	A
ΠΑΠΑΣ	B
ΔΗΜΟΥ	Δ
ΔΗΜΟΥ	A
ΔΗΜΟΥ	B

ΠΡΟΪΣΤΑΜΕΝΟΣ	ΕΞΑΡΤΩΜΕΝΟΣ
ΠΑΠΑΣ	ΓΙΑΝΝΗΣ
ΠΑΠΑΣ	ΣΟΦΙΑ
ΔΗΜΟΥ	ΚΩΣΤΑΣ
ΔΗΜΟΥ	ΝΙΚΗ
ΔΗΜΟΥ	ΣΤΕΛΛΑ

Πέμπτη Κανονική Μορφή

- 5KM εάν και μόνο εάν οι μόνες εξαρτήσεις σύζευξης που ικανοποιεί είναι στην πραγματικότητα συναρτησιακές εξαρτήσεις που ξεκινούν από υποψήφια κλειδιά

Άλλες ΚΜ

ΚΜ ΠΟΚ(ΠΕΔΙΟΥ ΟΡΙΣΜΟΥ-ΚΛΕΙΔΙΩΝ)

- Μια δέσμευση που λέει ότι οι τιμές ενός δεδομένου γνωρίσματος παίρνονται από ένα καθορισμένο πεδίο ορισμού
- Μια δέσμευση που ορίζει ότι κάποιο γνώρισμα ή συνδυασμός γνωρισμάτων είναι υποψήφιο κλειδί

ΚΜ ΠΕΡΙΟΡΙΣΜΟΥ- ΕΝΩΣΗΣ

Δίνει απαντήσεις σε ερωτήσεις που η θεωρία της κανονικοποίησης δεν δίνει απαντήσεις. Κακοί σχεδιασμοί.

Απο-κανονικοποίηση

- Υπάρχουν περιπτώσεις όπου μπορεί γίνει να απο-κανονικοποίηση των σχέσεων ώστε να επιτευχθεί καλύτερη απόδοση της βάσης(De-Normalization)