

Προγραμματισμός Σημασιολογικού Ιστού

Ενότητα 3: Εισαγωγή στην Οργάνωση των Σημασιολογικών Δεδομένων

Μ.Στεφανιδάκης

27-2-2016

Ποιο το κατάλληλο μοντέλο δεδομένων;

- ▶ Τα σημασιολογικά δεδομένα πρέπει να εκτεθούν “ώς έχουν” (raw)
 - ▶ Ποιο είναι το κατάλληλο μοντέλο οργάνωσης;
 - ▶ Και πώς θα εμπεριέχονται και τα μεταδεδομένα;
- ▶ Ας ξεκινήσουμε με ένα απλό μοντέλο: δεδομένα σε μορφή πίνακα (**tabular data**)
 - ▶ Η σημασιολογία των δεδομένων προκύπτει **έμμεσα από τη θέση τους** (γραμμή-στήλη)

Τύπος	Αριθμός
Λεωφορεία	58.519
ΙΧ	2.062.880
Οδοποιητικά	7.642
Εκχιονιστικά	6
Ποδήλατα	8.540.192

Η έμμεση σημασιολογία του πίνακα

- ▶ Κάθε γραμμή αντιστοιχεί σε μία βασική **οντότητα** (entity) δεδομένων
 - ▶ Έμμεση πληροφορία
- ▶ Κάθε στήλη αντιστοιχεί σε μια **ιδιότητα** (attribute)
 - ▶ Ρητή πληροφορία, αν υπάρχει περιγραφή στηλών

The diagram illustrates the semantic meaning of a table. An arrow labeled "Οντότητα_i" points to the first column of the table. A curved arrow labeled "Ιδιότητα_j" points from the first row to the second column. The table itself is as follows:

Τύπος	Αριθμός
Λεωφορεία	58.519
ΙΧ	2.062.880
Οδοποιητικά	7.642
Εκχιονιστικά	6
Ποδήλατα	8.540.192

Μοντέλο και Μορφή Αποθήκευσης

- ▶ Η μορφή πίνακα είναι ένα μοντέλο οργάνωσης δεδομένων (**data model**)
 - ▶ Προσδιορίζει τον τρόπο δόμησης της πληροφορίας
- ▶ Η δομημένη πληροφορία όμως
 - ▶ Πρέπει να **αποθηκευτεί** ως ακολουθία bytes
 - ▶ Και να **μεταδοθεί** μεταξύ παραγωγού και καταναλωτή
- ▶ Συνεπώς, πέρα από το μοντέλο των δεδομένων, είναι απαραίτητο ένα μορφότυπο σειριοποίησης (**serialization format**) των δεδομένων

Η μορφή αποθήκευσης CSV

- ▶ Comma Separated Values
 - ▶ Ένα ..μη πρότυπο πρότυπο (τουλάχιστον μέχρι πρόσφατα)
 - ▶ Με πολλές “διαλέκτους” (σχεδόν κάθε εφαρμογή έχει τη δική της!)
 - ▶ Για εξαγωγή-εισαγωγή μεγάλων σετ δεδομένων σε μορφή πίνακα από-σε βάσεις δεδομένων
- ▶ Έλεγχος Ιδιοτήτων:
 - ▶ Ανοικτό πρότυπο: **NAI**
 - ▶ Χρήση στο Web: **NAI** (αν και όχι τόσο συχνά)
 - ▶ Ευκολία προγραμματισμού: **NAI** (βιβλιοθήκες για πολλές γλώσσες)
 - ▶ Ρητά μεταδεδομένα: **OXI** (προαιρετικά, ονόματα στηλών στην πρώτη γραμμή)

Προσθήκη ρητού αναγνωριστικού (id)

- ▶ Οικείο σχήμα από τις σχεσιακές βάσεις...
- ▶ Κάθε βασική οντότητα διαθέτει μοναδικό αναγνωριστικό
 - ▶ Θεωρήστε προς το παρόν ότι αυτός είναι ο μοναδικός πίνακας στον κόσμο!

id	Τύπος	Αριθμός
1	Λεωφορεία	58.519
2	ΙΧ	2.062.880
3	Οδοποιητικά	7.642
4	Εκχιονιστικά	6
5	Ποδήλατα	8.540.192

Το μετα-μοντέλο ΕΑV

- ▶ Μια νέα μεταμόρφωση: το μοντέλο **Entity-Attribute-Value**
 - ▶ Μορφή **τριάδας** (triple): (**Οντότητα**, **Ιδιότητα**, **Τιμή**)
 - ▶ Η Οντότητα συμβολίζεται με το ρητό αναγνωριστικό της
 - ▶ Ως Ιδιότητες μπαίνουν οι (ρητές ή εννοούμενες) επικεφαλίδες των στηλών
 - ▶ Ως Τιμές χρησιμοποιούνται τα περιεχόμενα των κελιών στις διασταυρώσεις γραμμών-στηλών
 - ▶ Όλα τα μεταδεδομένα (ιδιότητες) δηλώνονται ρητά
 - ▶ Ο “εφιάλτης” του σχεσιακού μοντέλου!!!
 - ▶ Πλήρης απο-κανονικοποίηση (denormalization)

Παράδειγμα μετασχηματισμού

- ▶ Έστω ο πίνακας (δείχνεται μια γραμμή μόνο)

id	Engine	Weight(kg)	Tracks(mm)	Boom(m)
...
38rb	V8	60000	850	18
...

Δεδομένα κατά το μοντέλο EAV

- ▶ Η μία αυτή γραμμή παράγει από μόνη της τις εξής τριάδες

Entity	Attribute	Value
38rb	Engine	V8
38rb	Weight(kg)	60000
38rb	Tracks(mm)	850
38rb	Boom(m)	18

- ▶ Κάθε άλλη γραμμή του αρχικού πίνακα θα μετασχηματιστεί επίσης στις αντίστοιχες τριάδες!