



ΕΙΣΑΓΩΓΗ ΣΤΙΣ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ

1^η διάλεξη

Δομή Διαλέξεων

- **1^η διάλεξη**
 - Έννοιες και ορισμοί Βάσεων Δεδομένων
 - Σχισιακές Βάσεις Δεδομένων
 - Πίνακες
 - Συσχετίσεις
 - Περιορισμοί ακεραιότητας
 - Ερωτήματα
- **2^η Διάλεξη**
 - Επόμμενη γενιά Βάσεων Δεδομένων (Διαδικτυακές, Σημαντικός Ιστός)
 - Ιατρικές Βάσεις Δεδομένων
 - Ηλεκτρονικός Φάκελος Ασθενών
 - Βάσεις Δεδομένων και ευαίσθητα προσωπικά δεδομένα
 - Διαδικτυακές Ιατρικές Βάσεις Δεδομένων και Διεθνή Πρότυπα

Δομή Φροντιστηρίων

- 1^ο φροντιστήριο
 - Δημιουργία Βάσης Δεδομένων
 - Ορισμός πινάκων και περιορισμών
 - Εισαγωγή Πληροφορίας

Γενικές Έννοιες

- Ορισμός
 - Η **βάση δεδομένων** είναι μια **αυστηρά τυποποιημένη συλλογή** από σχετικά μεταξύ τους δεδομένα και παρέχει έναν **αυτόματο, κεντρικό και κοινό τρόπο χειρισμού τους**.
- Δεδομένα
 - Στοιχεία και Έννοιες, **τυποποιημένα σε καθορισμένη μορφή**, κατάλληλα για **επεξεργασία** από ανθρώπους και μηχανές
 - Τύπο Δεδομένων: κείμενο, γραφικά, εικόνες, ήχος, βίντεο
- Σύστημα Διαχείρισης ΒΔ (ΣΔΒΔ)
 - **Λογισμικό** που επιτρέπει τη δημιουργία και διαχείριση Βάσεων Δεδομένων

Παραδείγματα ΒΔ

- Μητρώο Φοιτητών
- Ηλεκτρονικός Ιατρικός Φάκελος Ασθενών
- Ψηφιακές Βιβλιοθήκες
- Δεδομένα γονιδιωματικής
- Πρωτεϊνική Δομή
- Ταξονομίες Ασθενειών

Μοντέλα Βάσεων Δεδομένων

Μοντέλο	Δεκαετία εμφάνισης
<u>Απλά Αρχεία Κειμένου</u>	'50
Ιεραρχικό	'60
Δικτυακό	'70
<u>Σχισιακό</u>	'80
Αντικειμενοστραφές	'90
Σχισιακό Αντικειμενοστραφές	'90
<u>Διαδικτυακό (web enabled)</u>	'90
Δομημένη Πληροφορία (XML)	'90
<u>Σημασιολογικό (RDF – Semantic Web)</u>	'00
No-SQL	'00
Big Data – Map Reduce	'10

Απλά Αρχεία

- Τα δεδομένα αποθηκεύονται ως εγγραφές σε αρχεία
- Απλή δομή εγγραφών και συγκεκριμένο αριθμό πεδίων
- Πιθανή χρήση ευρετηρίων
- Δεν υπάρχει μηχανισμός συσχέτισης δεδομένων μεταξύ αρχείων
- Χρήση συγκεκριμένου λογισμικού για την πρόσβαση στα δεδομένα

Σχεσιακές ΒΔ

- Τα δεδομένα οργανώνονται σε **πίνακες** με στήλες και γραμμές που αποθηκεύουν πληροφορία για μια **οντότητα**.
- Κάθε **γραμμή** αναπαριστά μια **εγγραφή** που είναι ένα **στιγμιότυπο** της οντότητας
- Κάθε **στήλη** αναπαριστά ένα **πεδίο** που είναι ένα **χαρακτηριστικό** μιας οντότητας
- Υπάρχουν συσχετίσεις μεταξύ των πινάκων που αναπαριστούν σχέσεις μεταξύ οντοτήτων (π.χ. ο ιατρός που διεξήγαγε μια εξέταση)
- Η πρόσβαση γίνεται μέσω της πρότυπης γλώσσας Standard Query Language (SQL) είτε τοπικά (στον Η/Υ) είτε απομακρυσμένα (μέσω δικτύου Η/Υ)

Παράδειγμα Αποθηκευμένης Πληροφορίας

hema_id	hema_name	hema_name_uk	hema_icd	icd_category
1	Χρόνια μυελογενής λευχαιμία Ph1(+)(t(9:22)(q34;q11...	Chronic myelogenous leukemia, Philadelphia chromos...	9875/3	1
2	Χρόνια ουδετεροφιλική λευχαιμία	Chronic neutrophilic leukemia	9963/3	1
3	Χρόνια ηωσινοφιλική λευχαιμία / Υπερηωσινοφιλικό σ...	Chronic eosinophilic leukemia/hypereosinophilic sy...	9964/3	1
4	Χρόνια ιδιοπαθής μυελοϊνώση (μυελοσκληήρυνση)	Chronic idiopathic myelofibrosis	9961/3	1
5	Ιδιοπαθής (αληθής) πολυκυτταραιμία	Polycythemia vera	9950/3	1
6	Ιδιοπαθής θρομβοκυττάρρωση	Essential thrombocythemia	9962/3	
7	Αταξινόμητο μυελοϋπερπλαστικό νόσημα	Myeloproliferative disease, unclassifiable	9975/3	
8	Χρόνια μυελομονοκυτταρική λευχαιμία	Chronic myelomonocytic leukemia	9945/3	
9	Ατυπη χρόνια μυελογενής λευχαιμία Ph1(-)	Atypical chronic myelogenous leukemia	9876/3	
10	Νεανική μυελομονοκυτταρική λευχαιμία	Juvenile myelomonocytic leukemia	9946/3	
11	Αταξινόμητο μυελοδυσπλαστικό/	Myelodysplastic / Myeloproliferative disease,	9975/3	

patient_id	patient_AT	patient_insurance	patient_occupation	patient_gender
62	ΘΓ24/10/1962ΚΑ-0062		282	male
61	ΕΜ07/04/1983ΣΑ-0061		282	female
60	ΚΤ11/12/1948ΝΣ-0060		284	male
59	ΑΠ26/10/1975ΒΜ-0059		155	female
58	ΣΠ20/11/1984ΠΕ-0058		282	male
57	ΚΜ30/05/1939ΓΦ-0057		282	male
56	ΓΑ22/07/1961ΧΧ-0056		282	male
55	ΕΣ01/03/1933ΓΣ-0055		282	male
54	ΖΜ14/02/1922ΛΘ-0054		291	female
53	ΧΛ01/02/1960ΑΠ-0053		282	male
16	ΓΜ06/04/2004ΠΑ-0016		279	female
17	ΚΚ20/12/2006ΔΕ-0017		280	male
18	ΓΠ05/12/1944ΕΑ-0018		281	male
19	ΠΑ18/04/1939ΓΑ-0019		282	male
20	ΠΚ10/05/1925ΕΒ-0020		283	male

Ορισμός Πίνακα

#	Column	Type	Collation	Attributes	Null	Default	Extra
1	<u>patient_id</u>	int(11)			No	None	AUTO_INCREMENT
2	patient_AT	varchar(21)	greek_general_ci		No	aa/00/00/yyyy/bb-cccc	
3	patient_insurance	varchar(255)	greek_general_ci		No		
4	patient_occupation	int(11)			No	0	
5	patient_gender	enum('male', 'female')	greek_general_ci		No	male	
6	patient_fname	varchar(255)	greek_general_ci		No		
7	patient_lname	varchar(255)	greek_general_ci		No		
8	patient_fathername	varchar(255)	greek_general_ci		No		
9	patient_mothername	varchar(255)	greek_general_ci		No		
10	patient_address	varchar(255)	greek_general_ci		No		
11	patient_region	int(11)			No	0	
12	patient_birth_place	int(11)			No	0	
13	patient_phone	varchar(255)	greek_general_ci		No		
14	patient_fhistory	varchar(255)	greek_general_ci		No	a	

Κλειδιά Πίνακα

- Ένα πεδίο (ή σύνολο πεδίων) ενός πίνακα λέγεται **κλειδί** αν μπορεί να **διαχωρίζει τις διαφορετικές εγγραφές** του πίνακα (π.χ. ΑΦΜ, ΑΜ φοιτητή).
- Το επιλεγόμενο κλειδί καλείται **πρωτεύον** (primary), ενώ τα υπόλοιπα καλούνται δευτερεύοντα (secondary) ή υποψήφια (candidate) ή εναλλακτικά (alternative).
- **Σύνθετο** (composite) όταν απαιτούνται περισσότερα του ενός χαρακτηριστικά για να συνθέσουν ένα κλειδί.

Συσχετίσεις μεταξύ πινάκων

- 1 προς 1 (1-1)
 - Στην περίπτωση αυτή κάθε εγγραφή του ενός πίνακα αντιστοιχεί σε μια μόνο εγγραφή στον άλλο πίνακα.
- 1 προς N (1-N)
 - Στην περίπτωση αυτή κάθε εγγραφή του ενός πίνακα αντιστοιχεί σε πολλές εγγραφές του άλλου πίνακα αλλά δεν ισχύει το αντίστροφο όπου κάθε εγγραφή του δεύτερου πίνακα αντιστοιχεί σε μία μόνο εγγραφή του πρώτου
 - **Παράδειγμα**
 - Ένας **Ασθενής** μπορεί να κάνει πολλές **Εξετάσεις**.
 - Μια **Εξέταση** γίνεται για έναν μόνο **Ασθενή**.
- M προς N (M-N)
 - Στην περίπτωση αυτή κάθε εγγραφή του ενός πίνακα αντιστοιχεί σε πολλές εγγραφές του άλλου πίνακα και αντίστροφα.
 - **Παράδειγμα**
 - Ένας **Ασθενής** μπορεί να τον εξετάσουν πολλοί **Γιατροί**.
 - Ένας **Γιατρός** μπορεί να εξετάσει πολλούς **Ασθενείς**.

Παράδειγμα συσχέτισης



Ένας Ασθενής διεξάγει πολλές (**N**) Εξετάσεις

Μια Εξέταση διεξάγεται για **1** Ασθενή

Παράδειγμα συσχέτισης

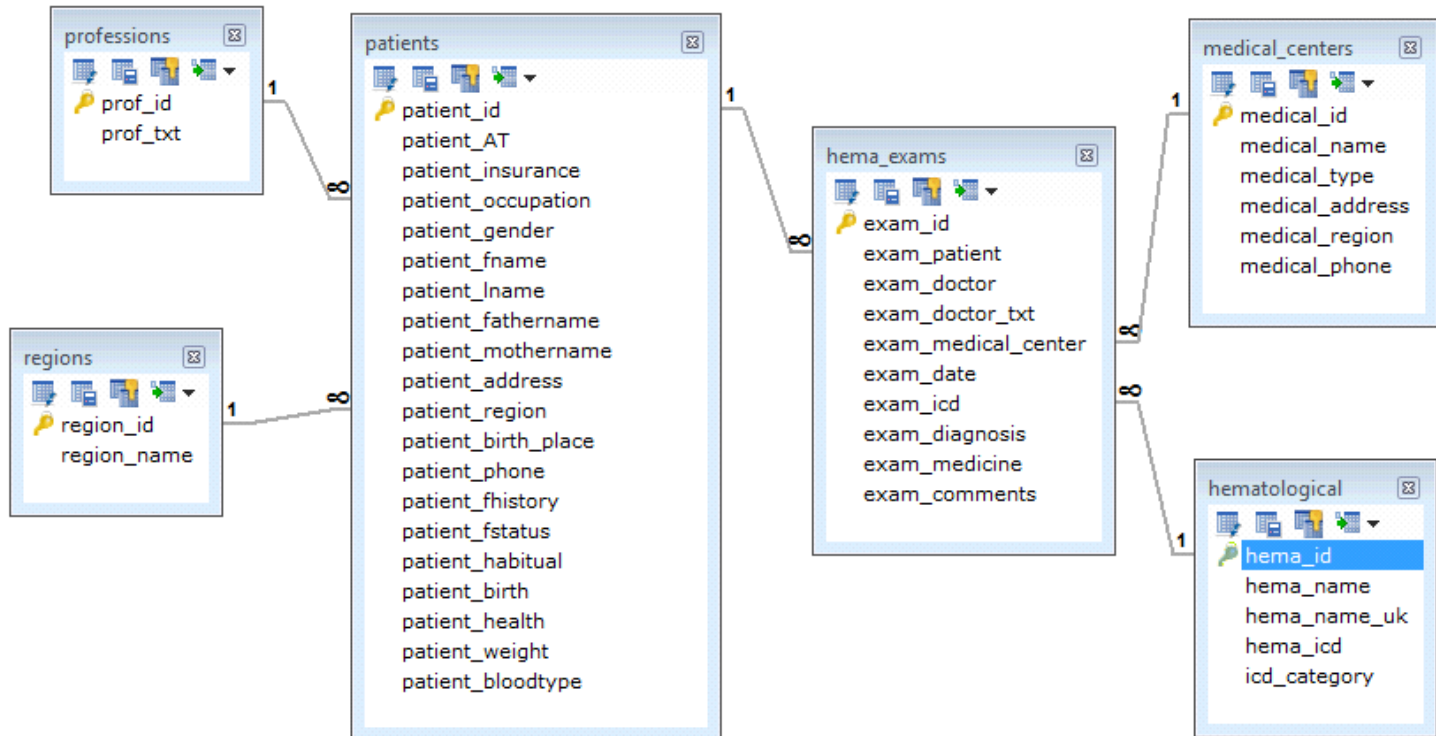
Πρωτεύον Κλειδί

Αρ. Φ.	Όνοματεπώνυμο	Διεύθυνση	Τηλέφωνο	Φύλο	Ημ. Γεν.	Συνήθειες
1524	Γιάννης Φαντομάς	Παναχαϊκού 28	2610123456	Άρρεν	1/1/1970	Μη καπνιστής
1525	Ελένη Ανυπάρχτου	Πτολεμαίου 76	2101234567	Θήλυ	2/2/1954	Καπνιστής
1526	Αναστασία Ρίου	Πάρου 132	2310123456	Θήλυ	3/3/1981	Καπνιστής

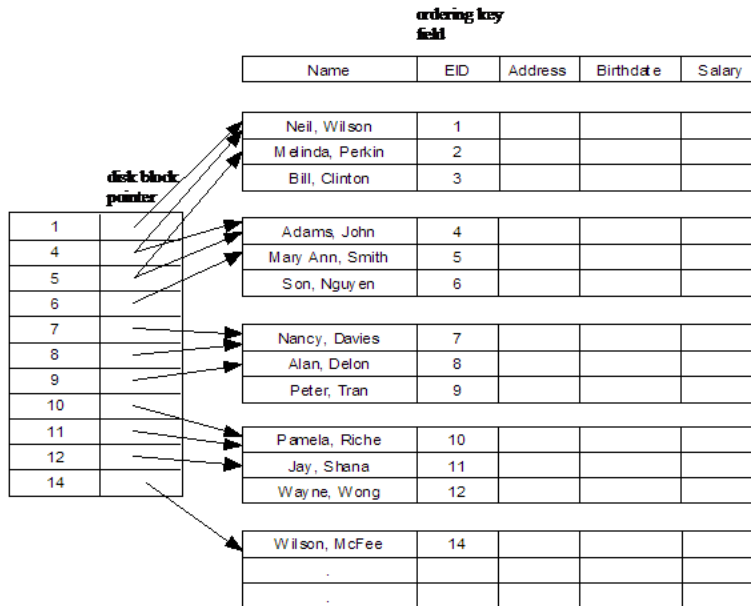
Ξένο Κλειδί

Αρ. Φ.	Ημ.	HCT	WBCC	LDH	PLT	TKE
1524	10/1/2008	40	12.000	248	198	33
1525	15/1/2008	37	1.500	77	135	15
1526	16/1/2008	35	2.300	114	126	20
1524	10/2/2008	41	9.000	257	205	35
1524	10/3/2008	40	10.000	212	199	32

Σχήμα ΒΔ



Ευρετήρια



- Ένα ευρετήριο (index) είναι μια βοηθητική δομή αρχείου που κάνει πιο αποδοτική την αναζήτηση μιας εγγραφής σε έναν πίνακα
- Το ευρετήριο καθορίζεται (συνήθως) σε ένα πεδίο του πίνακα που καλείται πεδίο ευρετηριοποίησης (indexing field)

Χρήση ευρετηρίων

- Οι λειτουργίες ενημέρωσης γίνονται γενικά πιο αργές, γιατί απαιτούν ενημέρωση και του ευρετηρίου
 - **Αλλά έχουμε αποδοτικότερες αναζητήσεις**
- Το αρχείο ευρετηρίου καταλαμβάνει **μικρότερο χώρο από το ίδιο το αρχείο δεδομένων** (οι καταχωρήσεις είναι μικρότερες και λιγότερες)
- Κάνοντας αναζήτηση στο ευρετήριο βρίσκουμε τον δείκτη στον πίνακα όπου αποθηκεύεται η εγγραφή που θέλουμε
- **Πρωτεύον ευρετήριο** (primary index): ορισμένο στο πρωτεύον κλειδί του πίνακα

Περιορισμοί ακεραιότητας

- Οι **περιορισμοί ακεραιότητας** (integrity constraints) πρέπει να ισχύουν πάντα για κάθε στιγμιότυπο της ΒΔ (σε κάθε σχέση της ΒΔ). Τότε ένα στιγμιότυπο ονομάζεται έγκυρο (valid).
- Οι περιορισμοί ακεραιότητας **επαληθεύονται** κάθε φορά που πραγματοποιούνται **αλλαγές στα δεδομένα** (εισαγωγή, διαγραφή, ενημέρωση).
- Οι περιορισμοί ακεραιότητας είναι:
 - Κενές τιμές (null values),
 - Ακεραιότητα οντοτήτων (entity integrity),
 - Ακεραιότητα αναφορών (referential integrity), και
 - Σημασιολογικοί περιορισμοί (semantics)

Περιορισμοί ακεραιότητας

- 1. Περιορισμός ακεραιότητας οντοτήτων (entity integrity constraint).** Η χρήση του πρωτεύοντος κλειδιού προϋποθέτει πάντα ότι στο αντίστοιχο πεδίο δεν είναι δυνατή η καταχώρηση μηδενικής ή κενής τιμής.
- 2. Περιορισμός αναφορικής ακεραιότητας (referential integrity constraint).** διατηρείται η συνέπεια μεταξύ των εγγραφών των συσχετιζόμενων πινάκων.
- 3. Περιορισμοί σημασιολογικής ακεραιότητας (semantic integrity constraint).** καθορισμός του πεδίου ορισμού (domain) των πεδίων κάθε πίνακα της βάσης.

Ερωτήματα

- Τα ερωτήματα είναι «εργαλεία» με τα οποία από έναν ή περισσότερους συσχετισμένους πίνακες δημιουργούμε νέες δυναμικές δομές δεδομένων δηλαδή νέους δυναμικούς πίνακες. Οι πίνακες που δημιουργούνται με μια ερώτηση είναι δυναμικοί πίνακες, δηλαδή αποθηκεύεται μόνο η δομή του πίνακα και όχι το περιεχόμενό του.
- **Παράδειγμα**
 - Εμφάνισε τα ονοματεπώνυμα και ημερομηνία των ασθενών που ξεκινάνε από «Μ».
 - Εμφάνισε τα στοιχεία των γιατρών που δουλεύουν στη παθολογική κλινική.

Είδη Ερωτημάτων

- Τα ερωτήματα χωρίζονται σε δύο βασικές ομάδες: τα ερωτήματα επιλογής και τα ερωτήματα ενέργειας.
 - Ένα **ερώτημα επιλογής** ανακτά απλά τα δεδομένα και τα καθιστά διαθέσιμα για χρήση. Μπορείτε να προβάλετε τα αποτελέσματα ενός ερωτήματος στην οθόνη και να τα εκτυπώσετε στο πρόχειρο.
 - Ένα **ερώτημα ενέργειας**, όπως υποδηλώνει και το όνομα, εκτελεί μια εργασία με τα δεδομένα. Τα ερωτήματα ενέργειας μπορούν να χρησιμοποιηθούν για τη δημιουργία νέων πινάκων, την προσθήκη δεδομένων σε υπάρχοντες πίνακες, την ενημέρωση ή τη διαγραφή δεδομένων.

Ερωτήματα Επιλογής

- Ένα ερώτημα μας επιτρέπει να καθορίσουμε:
 - Ποια πεδία θέλουμε να βλέπουμε.
 - Τη σειρά με την οποία θα εμφανίζονται τα πεδία.
 - Κριτήρια φιλτραρίσματος για κάθε πεδίο.
 - Τη σειρά με την οποία θα ταξινομείται κάθε πεδίο.
- **Παράδειγμα**
 - Εμφανίστε το **όνομα και επώνυμο** των **γιατρών** που **ανήκουν στην ορθοπαιδική κλινική** ταξινομώντας τα σε **αύξουσα σειρά σύμφωνα με το επώνυμό τους**.

SQL

- **SQL (Structured Query Language)** είναι η τυποποιημένη “standard” γλώσσα στις Σχεσιακές Βάσεις.
- Η πρώτη χρήση ήταν στο πρότυπο σύστημα της IBM, που ονομάστηκε SYSTEM-R, το οποίο ανεπτύχθη στα ερευνητικά εργαστήρια της εταιρείας (San Jose, California) στα μέσα της δεκαετίας το 1970. Η SQL έχει υποστεί πολλές τροποποιήσεις.

Παράδειγμα ερωτημάτων SQL

- **Ανάκτηση**

```
SELECT *  
FROM Departments  
WHERE budget > 1000000
```

- **Εισαγωγές**

```
INSERT INTO  
Employees (ssn, name)  
VALUES (83542, 'Aris')
```

- **Ενημερώσεις**

```
UPDATE Departments  
SET budget = 2000000  
WHERE did = 12345
```

- **Διαγραφές**

```
DELETE  
FROM Employees  
WHERE name = "Aris"
```