

Εργαστήριο Ασφάλειας Πληροφοριακών και Επικοινωνιακών Συστημάτων Πανεπιστήμιο Αιγαίου



Info-Sec-Lab

**Διασφαλίζοντας την Ιδιωτικότητα κατά τη
δημοσίευση και την ανταλλαγή δεδομένων**

Π. Ριζομυλιώτης

Πρόγραμμα εργασιών

15/12: (20.00-21.00, Μυρτώ)

Οικονομικά της Προστασίας της Ιδιωτικότητας

Μετά τις διακοπές

- Πρώτη έκδοση ιστοσελίδας εργασίας
- «Παρουσίαση» στο πρώτο μάθημα.

Ενδεικτική βιβλιογραφία

- Προστασία της Ιδιωτικότητας και Τεχνολογίες Πληροφορικής και Επικοινωνιών, **Τεχνικά και Νομικά Θέματα.**

Κ. Λαμπρινουδάκης, Λ. Μήτρου, Στ. Γκρίτζαλης, Σ. Κάτσικας

Εκδόσεις Παπασωτηρίου

(βασική πηγή της και της παρουσίασης)

- Κεφ. 20^ο : Διασφαλίζοντας την Ιδιωτικότητα κατά τη δημοσίευση και την ανταλλαγή δεδομένων

Γεώργιος Γρατσίας, Δημήτρης Σαχαρίδης, Μανώλης Τερροβίτης, Τίμος Σελλής

Σκοπός Παρουσίασης

- οι πιο διαδεδομένες εγγυήσεις ιδιωτικότητας που προσφέρονται για τη δημοσίευση δεδομένων
 1. **k -ανωνυμία,**
 2. **l -ποικιλομορφία,**
 3. **m -αμεταβλητότητα,**
 4. **k^m -ανωνυμία**

k -ανωνυμία (1)

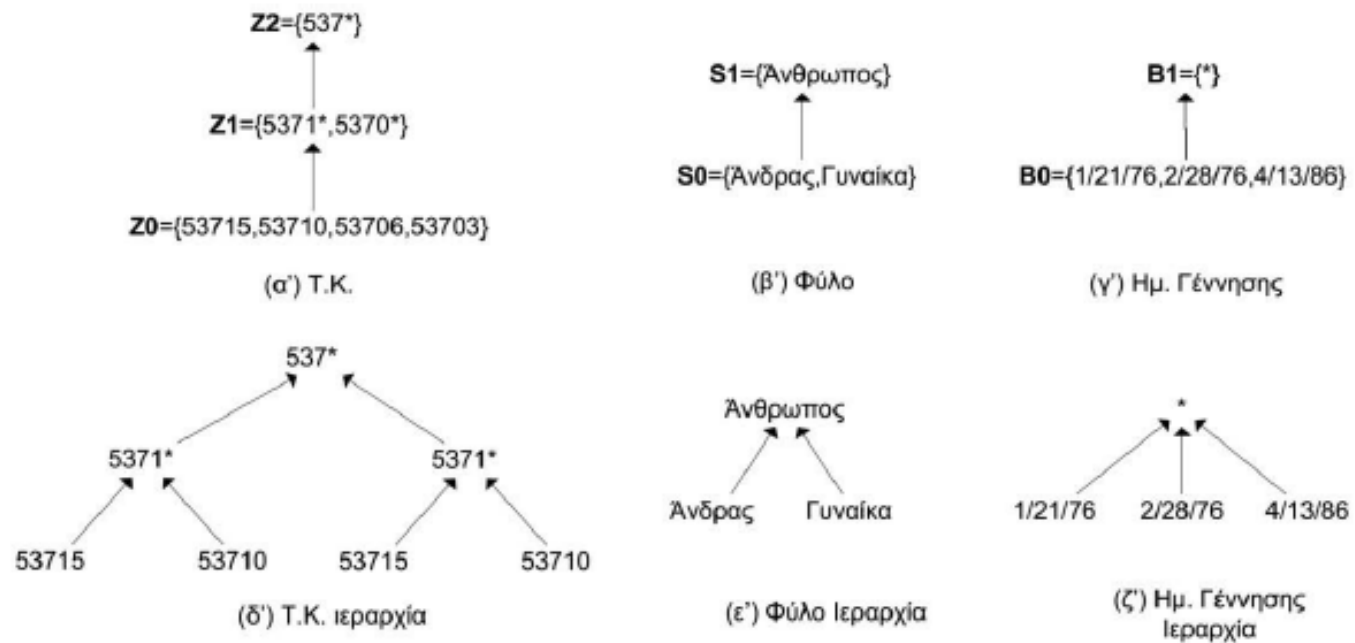
- Σκοπός της k -ανωνυμίας είναι να καταστήσει κάθε μια εγγραφή αδιάκριτη ανάμεσα σε άλλες $k-1$.
- **Ορισμός 1.** Το ελάχιστο σύνολο από γνωρίσματα $Q=Q_1, \dots, Q_d$ με το οποίο ένας πίνακας T μπορεί να συζευχθεί με κάποιες εξωτερικές πληροφορίες για να αναγνωριστούν ατομικές εγγραφές ονομάζεται ψευδό-αναγνωριστικό Σύνολο.
- **Ορισμός 2.** Ένας πίνακας T θα λέμε ότι είναι k -ανώνυμος με βάση ένα σύνολο γνωρισμάτων $Q=Q_1, \dots, Q_d$ αν το μέγεθος κάθε κλάσης ισοδυναμίας στο T με βάση τα $Q_1, \dots, Q_d$ έχει πληθάρημο τουλάχιστον k .
- Στόχος της ανωνυμοποίησης ενός πίνακα T είναι η παραγωγή μίας όψης V του πίνακα T , η οποία θα μετασχηματίζει τα δεδομένα του T , έτσι ώστε η V να είναι k -ανώνυμη με βάση το ψευδό-αναγνωριστικό.

k-ανωνυμία (2)

- Διάφοροι τρόποι *k*-ανωνυμοποίησης.
 - η γενίκευση (Generalization): οι τιμές των ψευδο-αναγνωριστικών (Quasi Identifiers) αντικαθίστανται με γενικότερες ή με βάση συγκεκριμένες ιεραρχίες γενίκευσης.

Δεδομένα Ασθενών			
Ημ. Γέννησης	Φύλο	T.K.	Ασθένεια
1/21/76	Άνδρας	53715	Γρίπη
4/13/86	Γυναίκα	53715	Ηπατίτιδα
2/28/76	Άνδρας	53703	Βρογχίτιδα
1/21/76	Άνδρας	53703	Κάταγμα
4/13/86	Γυναίκα	53706	Διάστρεμμα
2/28/76	Γυναίκα	53706	Οίδημα

k-ανωνυμία (3)



Σχήμα 20.3: Γενίκευση συνόλου τιμών και γενίκευση τιμής για T.K. (α, δ), Φύλο (β, ε) και Ημ. Γέννησης (γ, ζ)

k-ανωνυμία (4)

Χαρακτηριστικά μεθόδων που θα εξετάσουμε.

- *Ολική και Τοπική Κωδικοποίηση (Recoding).* Κάποια μοντέλα επιλέγουν να πετύχουν την *k*-ανωνυμία αντικαθιστώντας κάθε φορά μία τιμή του ψευδοαναγνωριστικού με μία άλλη, πιο γενική, σε όλη την έκταση της βάσης. Σε αυτή την περίπτωση μιλάμε για ολική κωδικοποίηση.
- *Με βάση Ιεραρχία και με βάση Κατατμήσεις.* Τα μοντέλα μπορεί να χρησιμοποιούν κάποια προκαθορισμένη ιεραρχία γενίκευσης τιμών ή να θεωρούν το σύνολο τιμών των γνωρισμάτων ταξινομημένο ορίζοντας γενικεύσεις χωρίζοντας το χώρο σε μοναδικές ξένες κατατμήσεις. Τα τελευταία μοντέλα, είναι προτιμότερα για αριθμητικά δεδομένα, ενώ τα πρώτα για κατηγορικά δεδομένα.

k-ανωνυμία (5)

Αλγόριθμοι

1. Ο αλγόριθμος **Incognito**
2. Ο αλγόριθμος **Mondrian**
3. Ο **Προς-τα-Πάνω αλγόριθμος**

k-ανωνυμία (6)

Ο αλγόριθμος Incognito (βασικές αρχές)

- **Ιδιότητα Γενίκευσης** Έστω ένας πίνακας T και έστω P και Q δύο σύνολα από γνωρίσματα στο T τέτοια ώστε $D_P <_D D_Q$. Αν το T είναι k -ανώνυμος με βάση το P τότε το T είναι k -ανώνυμος με βάση το Q .
- **Ιδιότητα Υποσυνόλου** Έστω ένα πίνακας και ένα σύνολο Q από γνωρίσματα στο T . Αν το T είναι k -ανώνυμο με βάση το Q , τότε το T είναι k -ανώνυμο με βάση κάθε σύνολο P από γνωρίσματα έτσι ώστε $P \text{ subset of } Q$.

k -ανωνυμία (7)

Αλγόριθμος 20.1: Incognito

Είσοδος: Ένας πίνακας T προς k -ανωνυμοποίηση, ένα σύνολο Q του ψευδó-αναγνωριστικού και μία ιεραρχία για κάθε γνώρισμα του ψευδó-αναγνωριστικού.

Έξοδος: Ένα σύνολο από k -ανώνυμων γενικεύσεων γενικού συνόλου.

C_1 = Κόμβοι της ιεραρχίας γενίκευσης συνόλου τιμών των γνωρισμάτων του Q .

E_1 = ακμές της ιεραρχίας γενίκευσης συνόλου τιμών των γνωρισμάτων του Q .

queue = μία άδεια ουρά

for $i=1$ to d

 // C_i and E_i ορίζουν ένα γράφο γενικεύσεων

S_i = ένα αντίγραφο του C_i

 roots = οι όλοι οι κόμβοι στο C_i χωρίς κάποιοι ακμή στο E_i κατευθυνόμενοι προς αυτούς

 Εισαγωγή των roots στην queue, διατηρώντας την queue ταξινομημένοι ως προς το ύψος.

 while queue δεν είναι άδεια do

 node = αφαίρεση του πρώτου στοιχείου της queue

 if node δεν είναι μαρκαρισμένος then

 if το T είναι k -ανώνυμο με βάση τα γνωρίσματα του κόμβου then

 Μάρκαρε όλες τις άμεσες γενικεύσεις του κόμβου

 else

 Διέγραψε τον κόμβο από το S_i

 Εισήγαγε τις άμεσες γενικεύσεις του κόμβου στην ουρά, κρατώντας την ουρά ταξινομημένη ως προς το ύψος.

 end if

 end if

end while

C_{i+1}, E_{i+1} = ΚατασκευήΓράφου(S_i, E_i)

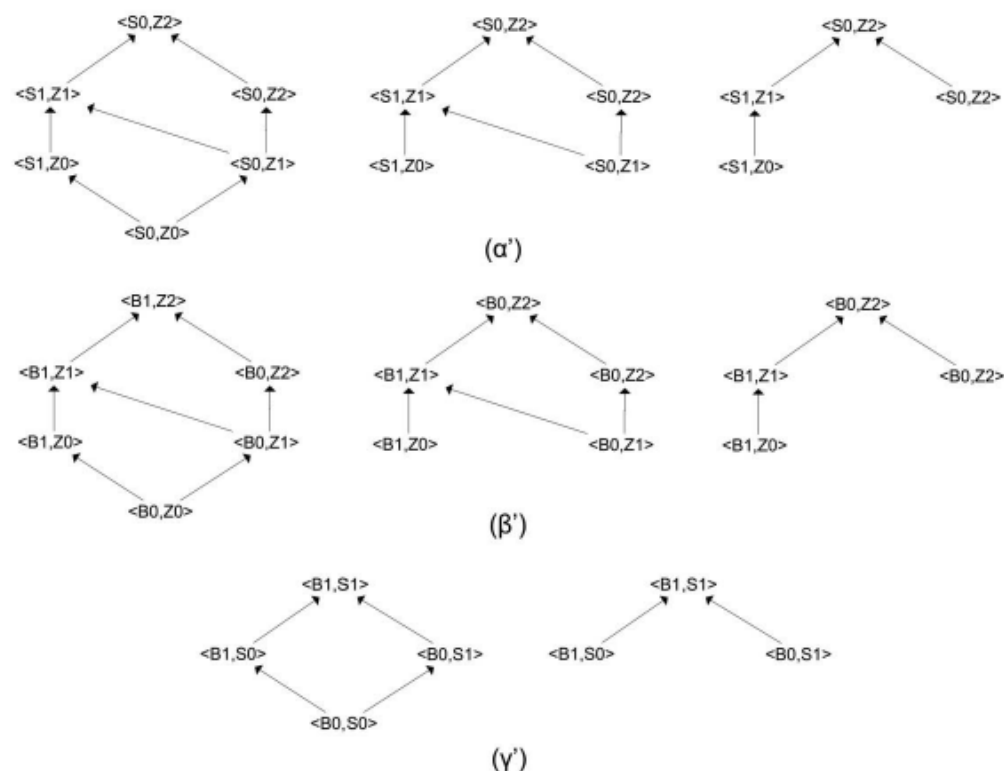
end for

Προβολή όλων των γνωρισμάτων του S_n στο T και στις ιεραρχίες

k-ανωνυμία (8)

Ο αλγόριθμος Incognito

Δεδομένα Ασθενών			
Ημ. Γέννησης	Φύλο	T.K.	Ασθένεια
1/21/76	Άνδρας	53715	Γρίπη
4/13/86	Γυναίκα	53715	Ηπατίτιδα
2/28/76	Άνδρας	53703	Βρογχίτιδα
1/21/76	Άνδρας	53703	Κάταγμα
4/13/86	Γυναίκα	53706	Διάστρεμμα
2/28/76	Γυναίκα	53706	Οίδημα

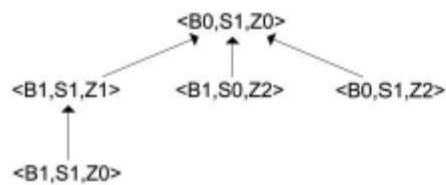


Σχήμα 20.5: Οι γράφοι που κατασκευάζει ο αλγόριθμος στη δεύτερη επανάληψη

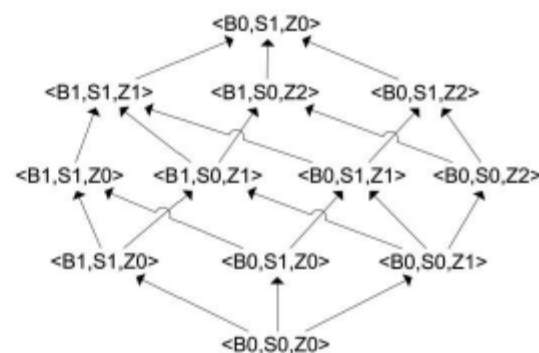
24/1/2012

k-ανωνυμία (9)

Ο αλγόριθμος Incognito



(α') Ο γράφος που κατασκευάζει
ο αλγόριθμος



(β') Όλος ο γράφος

Σχήμα 20.6: Αφαίρεση κόμβων μέσω των φάσεων σύζευξης και κλαδέματος

k-ανωνυμία (1 1)

- Ανωνυμοποίηση με βάση τη χρησιμότητα
 - Ο Προς-τα-Πάνω αλγόριθμος
 - Έστω ένας πίνακας T με ψευδό-αναγνωριστικό $(Q_1, \dots, Q_d)$ και έστω ότι το γνώρισμα Q_i είναι
 - αριθμητικό. Θεωρούμε μία εγγραφή $t=(q_1, \dots, q_i, \dots, q_d)$ η οποία γενικεύεται στην εγγραφή $t'=(q_1, \dots, [z_i, y_i], \dots, q_d)$ έτσι ώστε $y_i \leq q_i \leq z_i$. Για το γνώρισμα Q_i η κανονικοποιημένη ποινή βεβαιότητας (Normalized Certainty Penalty) ορίζεται ως
$$NCP_{Q_i}(t)=(z_i - y_i)/|Q_i|$$
 - κατηγορικό (Categorical Attribute), θεωρούμε μία εγγραφή $t=(q_1, \dots, q_i, \dots, q_d)$ η οποία γενικεύεται στην εγγραφή $t'=(q_1', \dots, u, \dots, q_d')$ όπου u είναι ένας πρόγονος του q_i σε κάποια ιεραρχία γενίκευσης. Για το γνώρισμα Q_i η κανονικοποιημένη ποινή βεβαιότητας ορίζεται ως

$$NCP_{Q_i}(t)=\text{μέγεθος}(u)/|Q_i|$$

k-ανωνυμία (12)

- Για την εγγραφή t η κανονικοποιημένη ποινή βεβαιότητας θα είναι:

$$NCP(t) = \sum_{i=1}^d (w_i NCP_{Q_i}(t)).$$

- Τέλος η συνολική ποινή για ένα υποσύνολο του πίνακα ή για όλο τον πίνακα αθροίζουμε όλες τις ποινές για κάθε εγγραφή

$$NCP(Q) = \sum_{t \in T} NCP(t)$$

k-ανωνυμία (13)

Ο Προς-τα-Πάνω αλγόριθμος

$$O(|T|^2 \log k)$$

Αλγόριθμος 20.3: Προς-τα-Πάνω

Είσοδος: ένας πίνακας T , οι ιεραρχίες των κατηγορικών δεδομένων και τα βάρη.

Εξοδος: μία k -ανώνυμη όψη V

Αρχικοποίηση: Δημιούργησε μία ομάδα για κάθε εγγραφή.

while υπάρχει κάποια ομάδα G έτσι ώστε $|G| < k$ do

 for κάθε ομάδα G έτσι ώστε $|G| < k$ do

 εξέτασε όλες τις άλλες ομάδες και βρες μία ομάδα G' έτσι ώστε το $NCP(G \cup G')$ να ελαχιστοποιείται και συνένωσε το G με το G' .

 end for

for κάθε ομάδα G έτσι ώστε $|G| \geq 2k$ do

 χώρισε την ομάδα σε $|G|/k$ ομάδες έτσι ώστε κάθε ομάδα να έχει τουλάχιστον k εγγραφές.

 end for

end while

return γενίκευσε και εξέδωσε τις τελικές ομάδες

I-ποικιλομορφία (1)

- Στόχος της I-ποικιλομορφίας είναι να διασφαλίσει την ανωνυμία των ευαίσθητων δεδομένων

Ακατέργαστα Δεδομένα Ασθενών				4-Ανώνυμα Δεδομένα Ασθενών			
	Μη-ευαίσθητα		Ευαίσθητο		Μη-ευαίσθητα		Ευαίσθητο
ΑΑ	T.K.	Ηλικία	Ασθένεια	ΑΑ	T.K.	Ηλικία	Ασθένεια
1	13053	28	Έμφραγμα	1	130**	< 30	Έμφραγμα
2	13068	29	Έμφραγμα	2	130**	< 30	Έμφραγμα
3	13068	21	Ίωση	3	130**	< 30	Ίωση
4	13053	23	Ίωση	4	130**	< 30	Ίωση
5	14853	50	Καρκίνος	5	1485*	≥ 40	Καρκίνος
6	14853	55	Έμφραγμα	6	1485*	≥ 40	Έμφραγμα
7	14850	47	Ίωση	7	1485*	≥40	Ίωση
8	14850	49	Ίωση	8	1485*	≥ 40	Ίωση
9	13053	31	Καρκίνος	9	130**	3*	Καρκίνος
10	13053	37	Καρκίνος	10	130**	3*	Καρκίνος
11	13068	36	Καρκίνος	11	130**	3*	Καρκίνος
12	13068	35	Καρκίνος	12	130**	3*	Καρκίνος

l-ποικιλομορφία (2)

- *Βασικά μειονεκτήματα της k-ανωνυμίας:*
 1. Είναι πιθανόν κάποιος επιτιθέμενος να γνωρίζει μέρος του ψευδό-αναγνωριστικού ενός ατόμου. Με αυτό τον τρόπο μπορεί να υπολογίσει τα προσωπικά δεδομένα του αν αυτά έχουν μεγάλη συχνότητα εμφάνισης.
 2. Εκτός από τη γνώση του ψευδό-αναγνωριστικού, ο χρήστης μπορεί να έχει κάποια εξωτερική γνώση.
- **Ορισμός 3.** Έστω ένα πίνακας T . Το σύνολο των γνωρισμάτων $S=S_1, \dots, S_m$ τις οποίες ένα εξωτερικός παράγοντας δε θέλουμε να γνωρίζει ή να μπορεί να βρει με αρκετά μεγάλη πιθανότητα ονομάζεται ευαίσθητο σύνολο και τα γνωρίσματα αυτού ευαίσθητα γνωρίσματα. (*Sensitive Attributes*).
- **Ορισμός 4.** Ένας πίνακας T είναι l -πολύμορφος εάν σε κάθε κλάση ισοδυναμίας Q_l η συχνότερη τιμή που εμφανίζεται στο ευαίσθητο γνώρισμα S δεν εμφανίζεται πάνω από $1/l \times |Q_l|$ φορές.

I-ποικιλομορφία (3)

Παράδειγμα (2-πολύμορφος πίνακας)

ΑΑ	Μη-ευαίσθητα		Ευαίσθητη
	T.K.	Ηλικία	Ασθένεια
1	1305*	≤ 40	Έμφραγμα
4	1305*	≤ 40	Πάθηση Ιού
9	1305*	≤ 40	Καρκίνος
10	1305*	≤ 40	Καρκίνος
5	1485*	> 40	Καρκίνος
6	1485*	> 40	Έμφραγμα
7	1485*	> 40	Πάθηση Ιού
8	1485*	> 40	Πάθηση Ιού
2	1306*	≤ 40	Έμφραγμα
3	1306*	≤ 40	Πάθηση Ιού
11	1306*	≤ 40	Καρκίνος
12	1306*	≤ 40	Καρκίνος

I-ποικιλομορφία (4)

Μεθοδολογία της Ανατομίας

- Για κάθε εγγραφή t του T συμβολίζουμε με $t[i]$ ($1 \leq i \leq d$) την τιμή του γνωρίσματος Q_i και με $t[d+1]$ την τιμή του S . Η ανατομία κατασκευάζει δύο νέους πίνακες, τον QIT και τον ST ώστε να ικανοποιήσει την I -ποικιλομορφία. Πιο συγκεκριμένα, δοθέντος μίας I -πολύμορφης κατάτμησης με m κλάσεις ισοδυναμίας, η ανατομία παράγει ένα πίνακα ψευδό-αναγνωριστικού (QIT) και ένα ευαίσθητων τιμών (ST) με βάση τις παρακάτω ιδιότητες:
- Το QIT έχει το σχήμα $(Q_1, \dots, Q_d, \text{Σύνολο-AA})$.
- Για κάθε κλάση ισοδυναμίας Q_{Ij} ($1 \leq j \leq m$) και για κάθε εγγραφή t του Q_{Ij} , το QIT έχει μία εγγραφή της μορφής: $(t[1], t[2], \dots, t[d], j)$.
- Το ST έχει σχήμα $(\text{Σύνολο-AA}, S, \text{Πλήθος})$.
- Για κάθε κλάση ισοδυναμίας Q_{Ij} ($1 \leq j \leq m$) και για κάθε τιμή v του ευαίσθητου γνωρίσματος S στο Q_{Ij} , το ST έχει μία εγγραφή της μορφής: $(j, v, c_j(v))$, όπου $c_j(v)$ ο αριθμός των εγγραφών t του Q_{Ij} έτσι ώστε $t[d+1]=v$.
- Εκτός από τις προηγούμενες εγγραφές δεν υπάρχουν άλλες στους δύο πίνακες.

I-ποικιλομορφία (5)

Παράδειγμα

Ακατέργαστα Δεδομένα Ασθενών				
ΑΑ	Μη-ευαίσθητα			Ευαίσθητο
	Ηλικία	Φύλο	T.K.	Ασθένεια
1	23	A	11000	πνευμονία
2	27	A	13000	δυσπείψια
3	35	A	59000	δυσπείψια
4	59	A	12000	πνευμονία
5	61	Γ	54000	γρίπη
6	65	Γ	25000	γαστρεντερίτιδα
7	65	Γ	25000	γρίπη
8	70	Γ	30000	βρογχίτιδα

Δεδομένα Ασθενών				
ΑΑ	Μη-ευαίσθητα			Ευαίσθητο
	Ηλικία	Φύλο	T.K.	Ασθένεια
1	[21-60]	A	[10001-60000]	πνευμονία
2	[21-60]	A	[10001-60000]	δυσπείψια
3	[21-60]	A	[10001-60000]	δυσπείψια
4	[21-60]	A	[10001-60000]	πνευμονία
5	[61-70]	Γ	[10001-60000]	γρίπη
6	[61-70]	Γ	[10001-60000]	γαστρεντερίτιδα
7	[61-70]	Γ	[10001-60000]	γρίπη
8	[61-70]	Γ	[10001-60000]	βρογχίτιδα

QIT πίνακας				
ΑΑ	Ηλικία	Φύλο	T.K.	Σύνολο-ΑΑ
1	23	A	11000	1
2	27	A	13000	1
3	35	A	59000	1
4	59	A	12000	1
5	61	Γ	54000	2
6	65	Γ	25000	2
7	65	Γ	25000	2
8	70	Γ	30000	2

ST πίνακας		
1	δυσπείψια	2
1	πνευμονία	2
2	βρογχίτιδα	1
2	γρίπη	2
2	γαστρεντερίτιδα	1

I-ποικιλομορφία (6)

Αλγόριθμος της Ανατομίας (εκτός ύλης)



Αλγόριθμος 20.4: Ανατομία

```
QIT =  $\emptyset$ ; ST =  $\emptyset$ ; gcnt = 0
Κατακερμάτισε τις εγγραφές στο T με βάση τις τιμές του  $A^s$  (ένα καλάθι για κάθε τιμή)
/*Πρώτη Φάση: Δημιουργία μιας νέας κλάσης ισοδυναμίας*/
while υπάρχουν τουλάχιστον 1 μη κενά καλάθια do
    gcnt = gcnt + 1;
    QIgcnt =  $\emptyset$  ;
    BC = Το σύνολο των 1 μεγαλύτερων καλάθιών
    for κάθε καλάθι στο BC do
        αφαίρεση μίας τυχαίας εγγραφής  $t$  από το BC.
        QIgcnt = QIgcnt  $\cup t$ 
    end for
end while
```

I-ποικιλομορφία (7)

Αλγόριθμος της Ανατομίας

```
/*Δεύτερη Φάση: Διαμοιρασμός των υπόλοιπων εγγραφών*/  
for κάθε μη κενό καλάθι do  
  /*Αυτό το καλάθι έχει ακριβώς μία εγγραφή*/  
   $t$  = η μοναδική εγγραφή στο καλάθι  
   $BC'$  = το σύνολο των κλάσεων ισοδυναμίας του προηγούμενου σταδίου που δεν  
  περιέχουν την τιμή  $t[d+1]$   
  στην  $S$   
  ανάθεση της  $t$  σε μία τυχαία κλάση ισοδυναμίας στο  $BC'$   
end for  
/*Δημιουργία του QIT και του ST*/  
for  $j = 1$  to  $gcnt$  do  
  for κάθε εγγραφή  $t \in QI_j$  do  
    εισαγωγή της εγγραφής  $(t[1], \dots, t[d], j)$  στο QIT  
  end for  
  for κάθε μοναδική τιμή  $v$  της  $S$  στο  $QI_j$  do  
     $c_j(v)$  = το πλήθος των εγγραφών στο  $QI_j$  με τιμή  $v$  στην  $S$   
    εισαγωγή της  $(j, v, c_j(v))$  στο ST  
  end for  
end for  
return QIT and ST
```

m-αμεταβλητότητα (1)

- πρόβλεψη για την ύπαρξη εισαγωγών και διαγραφών στη βάση
- **Ορισμός 5** Έστω μία αλληλουχία ανώνυμων όψεων $V(1), \dots, V_i(n)$ μίας βάσης T . Θα λέμε ότι ικανοποιούν την *m*-αμεταβλητότητα αν
 - Κάθε όψη είναι *m*-μοναδική.
 - Κάθε εγγραφή t η οποία ανήκει στο T και υπάρχει σε κάποιες από αυτές τις όψεις ως γενικευμένη ή ανατμημένη εγγραφή t^* , έχει πάντα την ίδια υπογραφή.
 - Είναι πιθανόν σε αυτές τις όψεις να υπάρχουν πλαστές εγγραφές.
- Ένας πίνακας T θα είναι *m*-μοναδικός αν σε κάθε κλάση ισοδυναμίας στον πίνακα υπάρχουν τουλάχιστον *m* εγγραφές και όλες οι εγγραφές σε αυτόν έχουν διαφορετικές τιμές στο ευαίσθητο γνώρισμα

m-αμεταβλητότητα (2)

- *Παράδειγμα*

Δεδομένα Ασθενών 1η Έκδοση			
Όνομα	Ηλικία	T.K.	Ασθένεια
Βασίλης	21	12000	δυσπεψία
Αλίκη	22	14000	βρογχίτιδα
Αντρέας	24	18000	γρίπη
Δαβίδ	23	25000	γαστρεντερίτιδα
Γιώργος	41	20000	γρίπη
Ελένη	36	27000	γαστρεντερίτιδα
Γιάννα	37	33000	δυσπεψία
Καίτη	40	35000	γρίπη
Λίνα	43	26000	γαστρεντερίτιδα
Παύλος	52	33000	δυσπεψία
Στέφανος	56	34000	γαστρεντερίτιδα

Ανώνυμα Δεδομένα: 1η Έκδοση			
Σύνολο-ΑΑ	Ηλικία	T.K.	Ασθένεια
1	21	12000	δυσπεψία
1	22	14000	βρογχίτιδα
2	24	18000	γρίπη
2	23	25000	γαστρεντερίτιδα
3	41	20000	γρίπη
3	36	27000	γαστρεντερίτιδα
4	37	33000	δυσπεψία
4	40	35000	γρίπη
4	43	26000	γαστρεντερίτιδα
5	52	33000	δυσπεψία
5	56	34000	γαστρεντερίτιδα

m-αμεταβλητότητα (3)

- Παράδειγμα

Δεδομένα Ασθενών: 2η Έκδοση			
Όνομα	Ηλικία	T.K.	Ασθένεια
Βασίλης	21	12000	δυσπείψια
Δαβίδ	23	25000	γαστρεντερίτιδα
Έφη	25	21000	γαστρεντερίτιδα
Γιάννα	37	33000	δυσπείψια
Λίνα	43	26000	γαστρεντερίτιδα
Γιώργος	41	20000	γρίπη
Μαίρη	46	30000	γαστρεντερίτιδα
Ρέα	54	31000	δυσπείψια
Στέφανος	56	34000	γαστρεντερίτιδα
Θύμιος	60	44000	γαστρεντερίτιδα
Βίκυ	65	36000	γρίπη

Ανώνυμα Δεδομένα: 2 ^η Έκδοχή			
Σύνολο-ΑΑ	Ηλικία	T.K.	Ασθένεια
1	21	12000	δυσπείψια
1	23	25000	γαστρεντερίτιδα

2	25	21000	γαστρεντερίτιδα
2	37	33000	δυσπείψια
2	43	26000	γαστρεντερίτιδα

3	41	20000	γρίπη
3	46	30000	γαστρεντερίτιδα

4	54	31000	δυσπείψια
4	56	34000	γαστρεντερίτιδα

5	60	44000	γαστρεντερίτιδα
5	65	36000	γρίπη

m-αμεταβλητότητα (4)

- Παράδειγμα

Πλαστά Ανώνυμα Δεδομένα: 2 ^η Έκδοχή				
Όνομα	Σύνολο-AA	Ηλικία	T.K.	Ασθένεια
Βασίλης	1	21	12000	δυσπεψία
c ₁	1	22	14000	βρογχίτιδα
Δαβίδ	2	23	25000	γαστρεντερίτιδα
Έφη	2	25	21000	γαστρεντερίτιδα
Γιάννα	3	37	33000	δυσπεψία
c ₂	3	40	35000	γρίπη
Λίνα	3	43	26000	γαστρεντερίτιδα
Γιώργος	4	41	20000	γρίπη
Μαίρη	4	46	30000	γαστρεντερίτιδα
Ρέα	5	54	31000	δυσπεψία
Στέφανος	5	56	34000	γαστρεντερίτιδα
Θύμιος	6	60	44000	γαστρεντερίτιδα
Βίκυ	6	65	36000	γρίπη

Βοηθητική Σχέση	
Σύνολο-AA	Πλήθος
1	1
3	1

m-αμεταβλητότητα (5)

- Αλγόριθμος
- Ο αλγόριθμος αποτελείται από τρεις φάσεις: διαίρεση, ισορρόπηση και ανάθεση
- Ορίζουμε

$$S_{\cap} = T(n) \cap T(n-1) \text{ και } S_{-} = T(n) - T(n-1)$$

m-αμεταβλητότητα (5)

- $\Pi\chi$

$S_+ = \{ \text{Βασίλης, Δαβίδ, Λίνα, Γιάννα, Γιώργος, Στέφανος} \}$

$S_- = \{ \text{Έφη, Μαίρη, Ρέα, Θύμιος, Βίκυ} \}$

Γιώργος	Δαβίδ		Στέφανος	Βασίλης		Γιάννα		Λίνα
γρίπη	γαστρ.	δυσ.	γαστρ.	δυσ.	βρογ	δυσ.	γρίπη	γαστρ.

(α') Τα περιεχόμενα των καλαθιών μετά τη διαίρεση

Γιώργος	Δαβίδ	Ρέα	Στέφανος	Βασίλης	c_1	Γιάννα	c_2	Λίνα
γρίπη	γαστρ.	δυσ.	γαστρ.	δυσ.	βρογ .	δυσ.	γρίπη	γαστρ.

(β') Τα περιεχόμενα των καλαθιών μετά την ισορρόπηση

Ερωτήσεις??

