

# Κεφάλαιο 24

## *Κατηγοριοποίηση*

Τεχνητή Νοημοσύνη - Β' Έκδοση

Ι. Βλαχάβας, Π. Κεφαλάς, Ν. Βασιλειάδης, Φ. Κόκκορας, Η. Σακελλαρίου

# Κατηγοριοποίηση

- ❖ Προσδιορισμός της κατηγορίας στην οποία ανήκει ένα αντικείμενο.
  - ❑ **Είσοδος:** ένα σύνολο δεδομένων που περιγράφουν το αντικείμενο.
  - ❑ **Έξοδος:** η κατηγορία στην οποία ανήκει
- ❖ Η επιλογή κατηγορίας γίνεται από προκαθορισμένο σύνολο
  - ❑ Κάθε αντικείμενο δεν ανήκει σε μια μοναδική κατηγορία.
  - ❑ Κάθε αντικείμενο δεν ανήκει σίγουρα σε κάποια κατηγορία.
- ❖ Συνήθως οι κατηγορίες είναι οργανωμένες σε **ιεραρχίες**
  - ❑ Οι υποκατηγορίες έχουν τις ιδιότητες των υπερκατηγοριών
  - ❑ Οι κατηγορίες του ίδιου επίπεδου έχουν αλληλοαναιρούμενες ιδιότητες
- ❖ Εφαρμογές:
  - ❑ Π.χ. **διάγνωση, διαμόρφωση, επιδιόρθωση βλαβών, κλπ.**

# Ευριστική Κατηγοριοποίηση

## Heuristic Classification

### ❖ Εξαντλητική Κατηγοριοποίηση:

- ❑ Απλή σύγκριση "επιφανειακών" χαρακτηριστικών του αντικειμένου
- ❑ Στις απλές περιπτώσεις αρκεί
- ❑ Όταν υπάρχουν πολλές ιδιότητες και πολύπλοκη ιεραρχία κατηγοριών:
  - Τα επιφανειακά χαρακτηριστικά δεν επαρκούν για την κατάταξη
  - Η εξαντλητική σύγκριση δεν είναι πρακτική

### ❖ Ευριστική κατηγοριοποίηση:

- ❑ Προσπαθεί να κατατάξει τα αντικείμενα σε κατηγορίες στα φύλλα της ιεραρχίας χωρίς να περάσει από όλα τα επίπεδα και να κάνει όλες τις συγκρίσεις
- ❑ Η κατάταξη γίνεται πιο γρήγορα, αλλά με μικρότερη ακρίβεια
- ❑ Χρησιμοποιεί εμπειρική γνώση για αντικείμενα, κατηγορίες και συσχετίσεις

# Φάσεις Ευριστικής Κατηγοριοποίησης

## ❖ Αφαίρεση ή γενίκευση των αντικειμένων-δεδομένων (data abstraction).

- ❑ Επικέντρωση μόνο στα σημαντικά χαρακτηριστικά ενός δεδομένου.

**ΕΑΝ** ένα βακτήριο ζει σε περιβάλλον στο οποίο δεν υπάρχει ελεύθερο οξυγόνο  
**ΤΟΤΕ** πρόκειται για αναερόβιο βακτήριο

- ❑ Απλοποίηση ποσοτικών δεδομένων.

**ΕΑΝ** ο ασθενής είναι ενήλικος, **ΚΑΙ**  
ο αριθμός των λευκών αιμοσφαιρίων είναι μικρότερος από  $2500/\text{cm}^3$   
**ΤΟΤΕ** ο αριθμός των λευκών αιμοσφαιρίων είναι μικρός

- ❑ Ιεραρχική οργάνωση των εννοιών.

**ΕΑΝ** το άτομο είναι πατέρας  
**ΤΟΤΕ** το άτομο είναι άντρας

## ❖ Ευριστική ταυτοποίηση (heuristic match) των γενικευμένων αντικειμένων σε μια γενικότερη περιγραφή ενός συνόλου κατηγοριών.

- ❑ Π.χ., ο πυρετός (γενίκευση υψηλής θερμοκρασίας) **μπορεί** να είναι ένδειξη μόλυνσης, η οποία εξειδικεύεται σε πολλές διαφορετικές μορφές.

## ❖ Επιλογή μιας συγκεκριμένης κατηγορίας-λύσης από το γενικό σύνολο κατηγοριών (solution refinement).

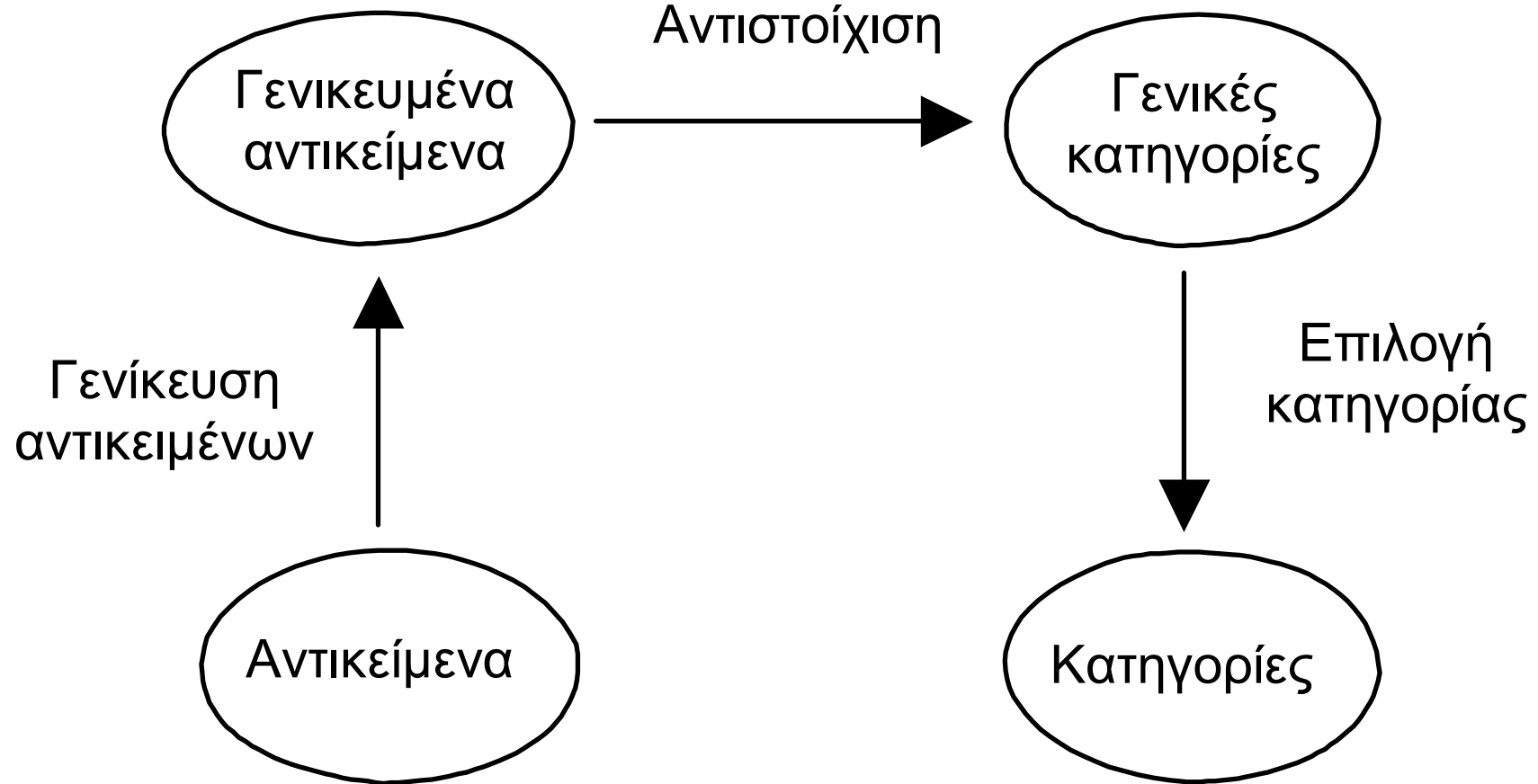
- ❑ Π.χ., το είδος της μόλυνσης πρέπει να διαγνωστεί με ακρίβεια ώστε να δοθεί η σωστή θεραπεία.

# Συστήματα Ευριστικής Κατηγοριοποίησης

## Φάσεις Λειτουργίας

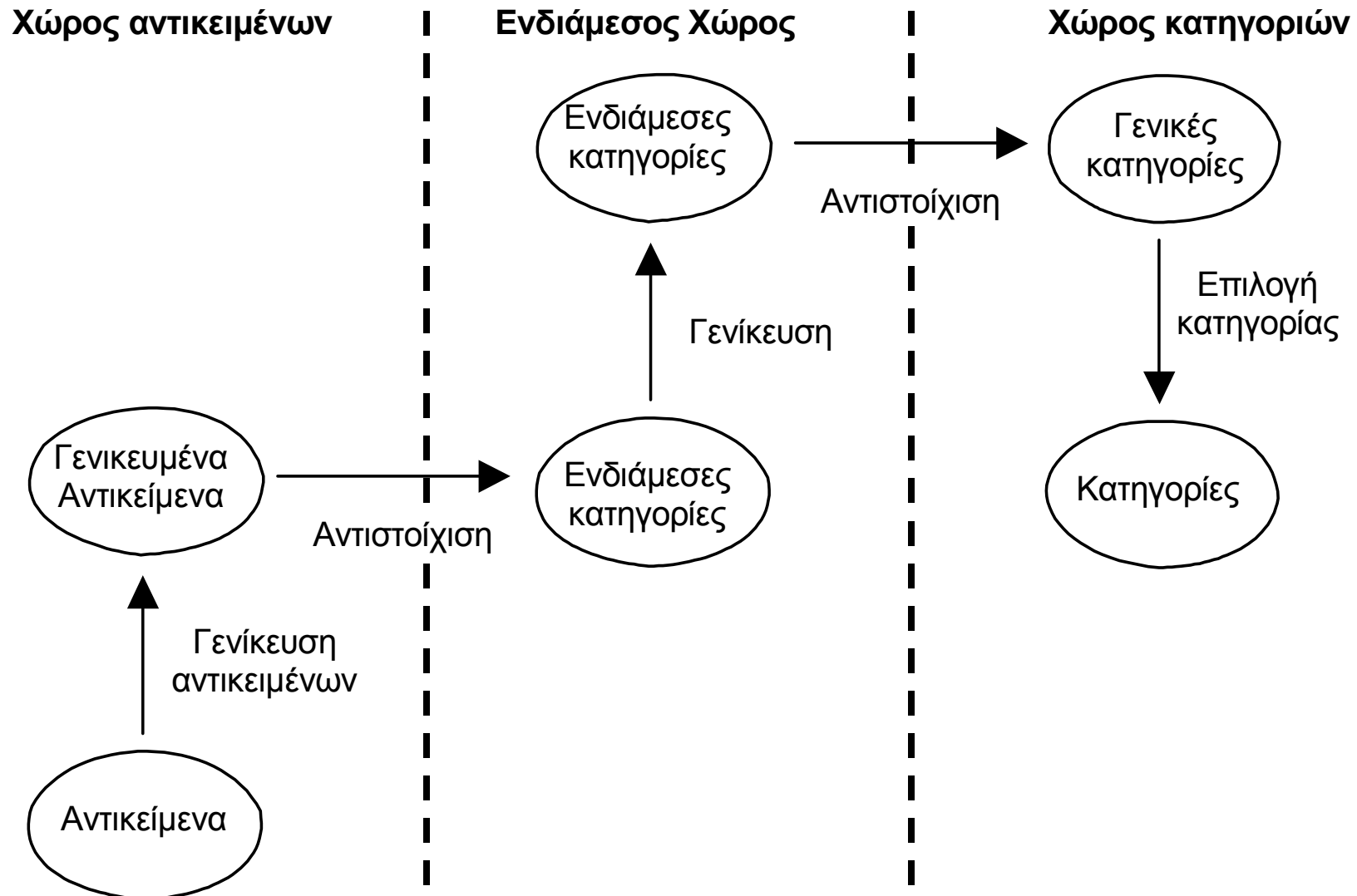
Χώρος αντικειμένων

Χώρος κατηγοριών



# Πολυβάθμια Συστήματα Κατηγοριοποίησης

## Multistage Classification Systems



# Συζευκτικό Μοντέλο Κατηγοριοποίησης (1/2)

- ❖ **Χώρος δεδομένων  $D$ :** Πεπερασμένο σύνολο χαρακτηριστικών  $\{D_i\}$  του αντικειμένου που πρέπει να καταταχθεί σε μια κατηγορία.
  - ❑  $D_i \in \{0, 1, "?\}$
  - ❑ Το σύνολο τιμών των  $D_i$  ονομάζεται **διάνυσμα τιμών**.
- ❖ **Χώρος κατηγοριών-λύσεων  $S$ :** Πεπερασμένο σύνολο λύσεων  $\{S_j\}$ .
- ❖ Για κάθε υποψήφια λύση  $S_j$  υπάρχει ένα πρότυπο που προσδιορίζει τις συνθήκες **συνέπειας** μεταξύ λύσεων και δεδομένων.
  - ❑  $C(S_j, D_i) = 1$ ,  $S_j$  είναι συνεπές με  $D_i = 1$  (δεν μπορεί να είναι λύση αν  $D_i = 0$ )
  - ❑  $C(S_j, D_i) = 0$ ,  $S_j$  είναι συνεπές με  $D_i = 0$  (δεν μπορεί να είναι λύση αν  $D_i = 1$ )
  - ❑  $C(S_j, D_i) = ?$ , οι τιμές του  $D_i$  δεν έχουν σχέση με τη συνέπεια των λύσεων  $S_j$
- ❖ Μια υποψήφια λύση  $S_j$  είναι **ασυνεπής** με ένα διάνυσμα τιμών και απορρίπτεται, αν τουλάχιστον ένα από τα δεδομένα είναι ασυνεπές με αυτήν.
- ❖ Μια λύση είναι **συνεπής (consistent)** αν δεν υπάρχει τιμή που να είναι ασυνεπής.

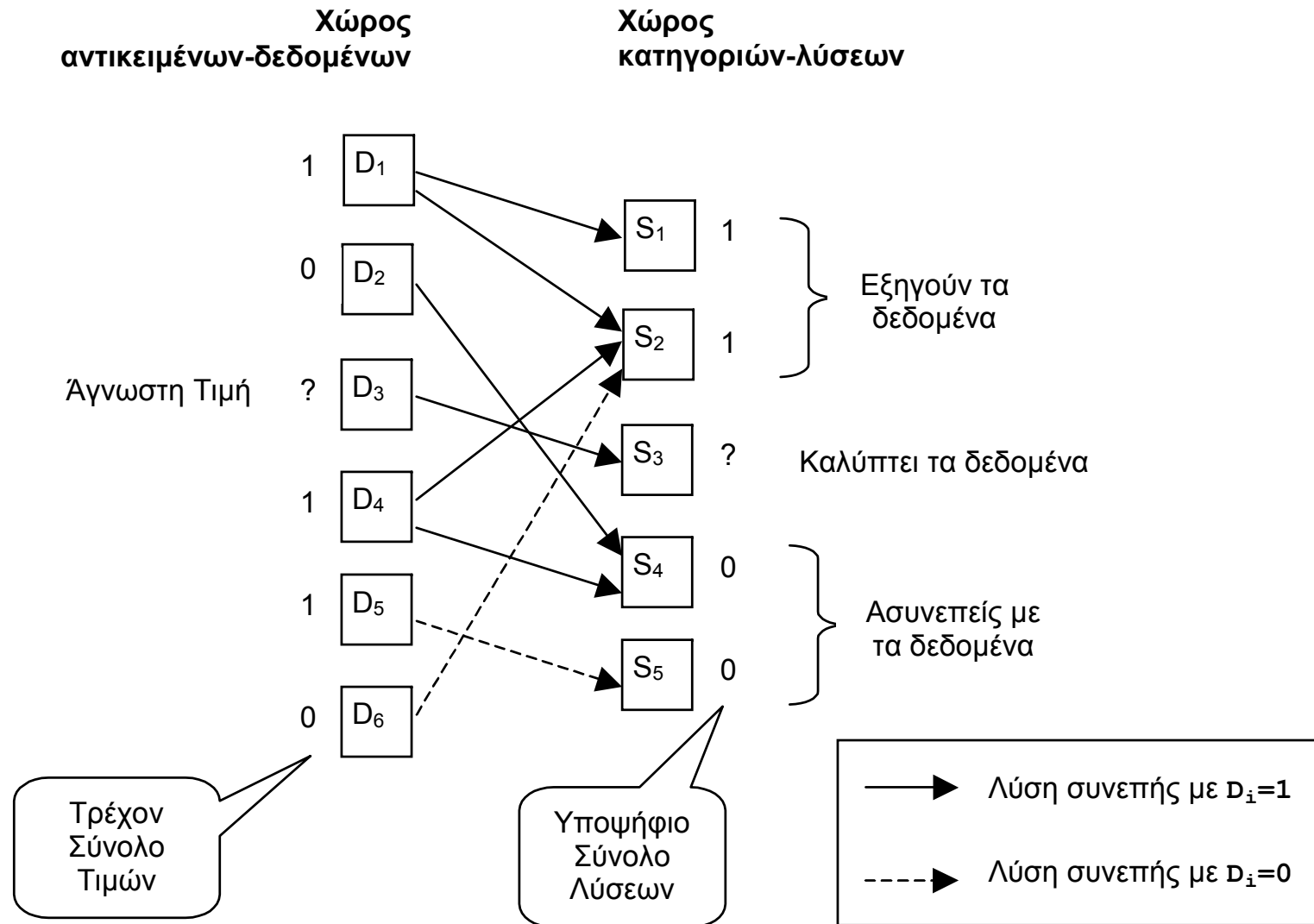
## Συζευκτικό Μοντέλο Κατηγοριοποίησης (2/2)

- ❖ Η λύση  $S_j$  καλύπτει το δεδομένο  $D_i$  αν ισχύει μια από τις περιπτώσεις:
  - A.  $D_i=1$  και  $C(S_j, D_i)=1$ , ή
  - B.  $D_i=0$  και  $C(S_j, D_i)=0$ .
- ❖ Μια λύση  $S_j$  ταιριάζει ή εξηγεί τα δεδομένα, αν όλα τα δεδομένα τα σχετικά με την  $S_j$  είναι γνωστά και όλες οι τιμές τους είναι συνεπείς με αυτήν.
- ❖ Ο προσδιορισμός της συνέπειας μιας λύσης γίνεται με τη διάδοση των τιμών από τα δεδομένα στις υποψήφιες λύσεις.
- ❖ Η κατάσταση μιας λύσης  $S_j$ :
  - Αν όλες οι τιμές στο  $S_j$  είναι 1, τότε το  $S_j$  ταιριάζει ή εξηγεί τα δεδομένα.
  - Αν κάποια τιμή στο  $S_j$  είναι 0, τότε το  $S_j$  είναι ασυνεπές και απορρίπτεται.
  - Αν όλες οι τιμές στο  $S_j$  είναι 1 και ?, τότε το  $S_j$  είναι συνεπές ή καλύπτει τα δεδομένα.
- ❖ Υπάρχει πιθανότητα κάποια διανύσματα τιμών να μην ταιριάζουν με καμία κατηγορία ή να ταιριάζουν με περισσότερες από μια κατηγορίες.



# Συζευκτικό Μοντέλο Κατηγοριοποίησης

## Παράδειγμα



# Μέθοδοι Κατηγοριοποίησης

- ❖ **Βασικός στόχος:** Ο αποκλεισμός των εναλλακτικών μονοπατιών στο χώρο αναζήτησης που αποτελείται από υποψήφιας κατηγορίες των δεδομένων εισόδου.

## Κ1: Παραγωγή και Δοκιμή

1. Θέσε την κενή λίστα ως λίστα των κατηγοριών-λύσεων
  2. Πάρε τα δεδομένα εισόδου και γενίκευσέ τα
  3. Για κάθε υποψήφια γενική κατηγορία
    - i. Έλεγξε αν τα δεδομένα εισόδου ανήκουν στην υποψήφια γενική κατηγορία
    - ii. Εάν ναι, πρόσθεσε την υποψήφια γενική κατηγορία στη λίστα των κατηγοριών
  4. Ανέφερε τις λύσεις από τη λίστα των κατηγοριών-λύσεων
- ❖ Το σύνολο των κατηγοριών (λύσεων) είναι αρκετά μικρό
    - ❑ Η εξαντλητική σύγκριση είναι πρακτικά εφικτή
  - ❖ Όλα τα απαραίτητα δεδομένα μπορούν να αποκτηθούν στην αρχή της διαδικασίας
    - ❑ Η απουσία κάποιων δεδομένων σημαίνει πως αυτά δεν υφίστανται
  - ❖ Ελάχιστα συστήματα χρησιμοποιούν αυτή τη μέθοδο

## Κ2: Από τα Δεδομένα σε Πιθανές Λύσεις

- ❖ Μειώνει τον υπολογιστικό χρόνο όταν υπάρχει μεγάλος αριθμός λύσεων
  - ❑ Επικεντρώνεται μόνο σε εκείνες που πιθανώς ικανοποιούν τα δεδομένα
  - ❑ Καταγράφει ποιες λύσεις έχει ήδη δοκιμάσει
  - ❑ Ασχολείται με κάθε υποψήφια λύση μόνο μια φορά
- ❖ Διαδικασία γενίκευσης δεδομένων (*data abstractor*)
  - ❑ Από μόνη της ένα μικρό σύστημα κατηγοριοποίησης.
  - ❑ Για κάθε σύνολο δεδομένων μπορεί να υπάρχουν πολλές δυνατότητες γενίκευσης.
- ❖ Διαδικασία ανάκλησης υποψήφιων λύσεων (*candidate retriever*)
  - ❑ Μπορούν εν δυνάμει να εξηγήσουν ένα δεδομένο.
  - ❑ Συζευκτικό μοντέλο κατηγοριοποίησης: λύσεις που είναι συνεπείς ή καλύπτουν τα δεδομένα
  - ❑ Δεν πρέπει να επιστρέφονται πολύ λίγες λύσεις (θα χαθούν κάποιες)
  - ❑ Δεν πρέπει να επιστρέφονται πάρα πολλές (δεν βελτιώνει την Κ1)
- ❖ Διαδικασία ελέγχου λύσεων (*solution tester*)
  - ❑ Βαθμολογεί τις υποψήφιες κατηγορίες-λύσεις και επιλέγει μία ή περισσότερες
  - ❑ Κριτήρια: π.χ. προϋπάρχουσες πιθανότητες, κλπ.

## Κ2: Από τα Δεδομένα σε Πιθανές Λύσεις Αλγόριθμος

1. Θέσε την κενή λίστα ως λίστα των κατηγοριών-λύσεων.
2. Πάρε τα δεδομένα εισόδου.
3. Γενίκευσε τα δεδομένα με τη διαδικασία `data_abstractor`.
4. Βρες τις υποψήφιας γενικές κατηγορίες με τη διαδικασία `candidate_retriever`.
5. Για κάθε υποψήφια γενική κατηγορία:
  - i. Έλεγξε αν τα δεδομένα εισόδου ανήκουν στην υποψήφια γενική κατηγορία (`solution_tester`)
  - ii. Εάν ναι, πρόσθεσε την υποψήφια γενική κατηγορία στη λίστα των κατηγοριών.
6. Ανέφερε τις λύσεις από τη λίστα των κατηγοριών-λύσεων.

# Κ3: Ιεραρχική Κατηγοριοποίηση Καθοδηγούμενη από τις Λύσεις

- ❖ Ιεραρχική τεχνική παραγωγής και δοκιμής
- ❖ Υπόθεση: οι πιθανές κατηγορίες-λύσεις είναι ιεραρχικά διατεταγμένες
  - ❑ Διασχίζει την ιεραρχία από πάνω προς τα κάτω και πρώτα σε πλάτος
  - ❑ Οι τελικές κατηγορίες-λύσεις βρίσκονται στα φύλλα του δένδρου
  - ❑ Οι λύσεις των ενδιάμεσων επιπέδων είναι μερικές ή γενικές
- ❖ Σε κάθε επίπεδο συγκρίνει τις υποψήφιες λύσεις με τα δεδομένα και απορρίπτει κλαδιά
  - ❑ Μπορεί να ζητήσει πρόσθετα δεδομένα για λεπτομερέστερη διάκριση
- ❖ Έλεγχος λύσεων σε κατώτερο επίπεδο, ακολουθώντας τα κλαδιά που δεν έχουν απορριφθεί
  - ❑ Αν μια κατηγορία απορριφθεί, απορρίπτονται και όλες οι υποκατηγορίες

# Κ3: Ιεραρχική Κατηγοριοποίηση Καθοδηγούμενη από τις Λύσεις Αλγόριθμος

1. Θέσε την κενή λίστα ως λίστα των κατηγοριών-λύσεων.
2. Για κάθε επίπεδο της ιεραρχίας των κατηγοριών-λύσεων, επανέλαβε τα ακόλουθα:
  - i. Πάρε τα δεδομένα εισόδου που χρειάζονται για να γίνει διάκριση των υποψήφιων λύσεων σε αυτό το επίπεδο.
  - ii. Γενίκευσε τα δεδομένα με τη διαδικασία `data_abstractor`.
  - iii. Για κάθε υποψήφια γενική κατηγορία αυτού του επιπέδου:
    - a. Έλεγξε αν τα δεδομένα εισόδου ανήκουν στην υποψήφια γενική κατηγορία με τη διαδικασία `solution_tester`.
    - b. Εάν όχι, τότε απέριψε την υποψήφια γενική κατηγορία.
    - c. Εάν ναι, τότε έλεγξε αν η υποψήφια γενική κατηγορία είναι τερματική στην ιεραρχία
      - 1) Εάν όχι, τότε αφαίρεσε την υποψήφια γενική κατηγορία από τη λίστα και πρόσθεσε τις κατηγορίες των κόμβων-παιδιών του δένδρου.
      - 2) Εάν ναι, τότε πρόσθεσε την υποψήφια γενική κατηγορία στη λίστα των λύσεων
3. Βαθμολόγησε τις λύσεις από τη λίστα των κατηγοριών-λύσεων και επέστρεψέ τις.

# Κ3: Ιεραρχική Κατηγοριοποίηση Καθοδηγούμενη από τις Λύσεις

## Προϋποθέσεις Εφαρμογής

- ❖ Οι κατηγορίες πρέπει να μπορούν να διαταχθούν σε ιεραρχία
- ❖ Σε κάθε επίπεδο πρέπει να υπάρχει ένα υποσύνολο των δεδομένων που μπορεί να κάνει διάκριση μεταξύ των γενικών κατηγοριών του συγκεκριμένου επιπέδου.
- ❖ Το δένδρο πρέπει να είναι όσο πιο ισορροπημένο γίνεται
  - ❑ Ο διαχωρισμός των πληθυσμών πρέπει να είναι όσο το δυνατόν μεγαλύτερος
  - ❑ Κατάλληλη επιλογή του υποσυνόλου των δεδομένων
  - ❑ Σε κάθε βήμα αποκλείονται πολλές λύσεις (αποδοτικότητα)
  - ❑ Ομοιότητα με τα δένδρα κατηγοριοποίησης ή απόφασης.
- ❖ **Στόχος:** να βρεθούν οι πιο ειδικές κατηγορίες που είναι συνεπείς με τα δεδομένα
  - ❑ Οι τερματικοί κόμβοι (φύλλα) του δένδρου
  - ❑ **Τερματισμός:** αν φθάσει στους τερματικούς κόμβους ή αν δεν υπάρχουν άλλα κλαδιά προς εξερεύνηση

## Κ4: Ιεραρχική Κατηγοριοποίηση Καθοδηγούμενη από τα Δεδομένα

- ❖ Η Κ3 ανεξάρτητα από την περίπτωση ξεκινά ζητώντας πάντα τα ίδια δεδομένα
  - ❑ Εξαιρετικά χρονοβόρο σε προβλήματα διάγνωσης βλαβών ή ασθενειών
  - ❑ Πάντα ξεκινάμε από τα δεδομένα-συμπτώματα
- ❖ Η μέθοδος Κ4 δεν ζητάει από το χρήστη δεδομένα που δεν έχουν σχέση με την περίπτωση που αντιμετωπίζεται
- ❖ Η μέθοδος:
  - ❑ Υποθέτει ότι όλα τα σχετικά δεδομένα έχουν δοθεί στην αρχή (όπως η Κ2)
  - ❑ Προχωρά στην επιλογή των υποψήφιων κατηγοριών-λύσεων με αυτά
  - ❑ Αποκλείει από την ιεραρχία όλα εκείνα τα κλαδιά που δεν είναι συμβατά με τα δεδομένα που δόθηκαν (όπως η Κ3)
  - ❑ Αποκλείει επίσης κάποιες από τις αρχικές υποψήφιες λύσεις, οι οποίες δεν επιβεβαιώνουν τα δεδομένα



## Κ4: Ιεραρχική Κατηγοριοποίηση Καθοδηγούμενη από τα Δεδομένα Αλγόριθμος

1. Θέσε την κενή λίστα ως λίστα των κατηγοριών-λύσεων.
2. Πάρε τα δεδομένα εισόδου.
3. Γενίκευσε τα δεδομένα με τη διαδικασία `data_abstractor`.
4. Βρες τις υποψήφιες γενικές κατηγορίες με τη διαδικασία `candidate_retriever`.
5. Για κάθε επίπεδο της ιεραρχίας των κατηγοριών το οποίο ξεκινάει από τις υποψήφιες γενικές κατηγορίες (και όχι από την κορυφή της ιεραρχίας), επανέλαβε τα ακόλουθα:
  - i. Πάρε τα δεδομένα εισόδου που χρειάζονται για να γίνει διάκριση των υποψήφιων λύσεων σε αυτό το επίπεδο.
  - ii. Γενίκευσε τα δεδομένα με τη διαδικασία `data_abstractor`.
  - iii. Πρόσθεσε νέες υποψήφιες γενικές κατηγορίες λόγω των καινούριων δεδομένων.
  - iv. Για κάθε υποψήφια γενική κατηγορία αυτού του επιπέδου:
    - a. Έλεγξε αν τα δεδομένα εισόδου ανήκουν στην υποψήφια γενική κατηγορία με τη διαδικασία `solution_tester`.
    - b. Εάν όχι, τότε απέριψε την υποψήφια γενική κατηγορία.
    - c. Εάν ναι, τότε έλεγξε αν η υποψήφια γενική κατηγορία είναι τερματική στην ιεραρχία
      - 1) Εάν όχι, τότε αφαίρεσε την υποψήφια γενική κατηγορία από τη λίστα και πρόσθεσε τις κατηγορίες των κόμβων-παιδιών του δένδρου.
      - 2) Εάν ναι, τότε πρόσθεσε την υποψήφια γενική κατηγορία στη λίστα των λύσεων.
6. Βαθμολόγησε τις λύσεις από τη λίστα των κατηγοριών-λύσεων και επέστρεψέ τις.

# Κ4: Ιεραρχική Κατηγοριοποίηση Καθοδηγούμενη από τα Δεδομένα Πλεονεκτήματα

- ❖ Η μέθοδος Κ4 βελτιώνει την Κ3
  - ❑ Θεωρεί ότι όσα δεδομένα δόθηκαν στην αρχή είναι τα μόνα σχετικά με την περίπτωση
  - ❑ Δεν εξετάζει καθόλου τις γενικές κατηγορίες οι οποίες δεν εξηγούν τα αρχικά δεδομένα
- ❖ Όταν φτάσει στη γενίκευση των αρχικών δεδομένων, η Κ4 λειτουργεί ανάστροφα
  - ❑ Από τις λύσεις προς τα δεδομένα (όπως η Κ3)
  - ❑ Ρωτά για επιπλέον δεδομένα για να αποκλείσει κάποιες από τις αρχικές υποψήφιας λύσεις
  - ❑ Κάποιες από τις γενικές κατηγορίες που είχαν αρχικά αποκλειστεί μπορεί με τα καινούρια δεδομένα να θεωρηθούν υποψήφιας.
- ❖ Η μέθοδος κινείται σε κάθε κύκλο:
  1. προς τα πάνω, γενικεύοντας τα νέα δεδομένα που αποκτά
  2. προς τα κάτω, αποκλείοντας κάποια κλαδιά του δένδρου λόγω νέων δεδομένων
- ❖ Η Κ4 είναι ευέλικτη και πιο αποδοτική
  - ❑ Επικεντρώνεται στην εξέταση **μόνο** των λύσεων που είναι απολύτως σχετικές με τα δεδομένα

# Μελέτη Περίπτωσης - Το Σύστημα DENDRAL

- ❖ Ενδιάμεσος σταθμός μεταξύ προγραμμάτων "έξυπνης αναζήτησης" και εμπείρων συστημάτων
  - ❑ Απ' ευθείας καταγραφή ειδικής γνώσης
  - ❑ Η αναζήτηση λύσεων περιορίζεται με πληροφορίες του χρήστη
- ❖ **Αντικείμενο:** Ο καθορισμός της μοριακής δομής (στερεοχημικό τύπος) αγνώστων οργανικών ουσιών
  - ❑ Από την ανάλυση των αποτελεσμάτων του φασματογράφου μάζας
  - ❑ Με τη χρήση τροποποιημένης μεθόδου δημιουργίας και ελέγχου (generate-and-test) εναλλακτικών λύσεων.
- ❖ **Στερεοχημικός τύπος:** δομή των ατόμων που αποτελούν το μόριο στο χώρο, καθώς και το είδος των μεταξύ τους χημικών δεσμών
  - ❑ Δεδομένος ο μοριακός τύπος (π.χ.  $C_6H_{13}NO_2$ )
  - ❑ Τα ισομερή που αντιστοιχούν σε ένα μοριακό τύπο είναι συνήθως πάρα πολλές (π.χ. 10000)

# Το Σύστημα DENDRAL

## Βάση γνώσης

- ❖ Περιέχει **περιορισμούς** που πρέπει να ικανοποιεί η δομή της ένωσης, για να περιορίσει τον μεγάλο αριθμό ισομερών της ένωσης
- ❖ Βασίζονται στην παρουσία ή απουσία κάποιων τμημάτων από το φάσμα.
- ❖ Χρησιμοποιούνται για να περιορίσουν το μεγάλο αριθμό εναλλακτικών δομών.
  - ❑ **Απαιτούμενοι:** Οι υποψήφιες ενώσεις πρέπει να ικανοποιούν τα στοιχεία που παρατηρήθηκαν.
  - ❑ **Απαγορευτικοί:** Οι πιθανές ενώσεις πρέπει να είναι χημικά σταθερές.

# Το Σύστημα DENDRAL

## Λειτουργία

### ❖ Το πρόγραμμα:

- Συσχετίζει τις κορυφές εντάσεων του φάσματος με τμήματα της διασπασμένης ένωσης
- Συνδυάζει τα τμήματα ώστε να ταιριάζουν στο δομικό σκελετό
- Δίνει μια λίστα ημιτελών υποθέσεων σχετικά με τη δομή της άγνωστης ένωσης

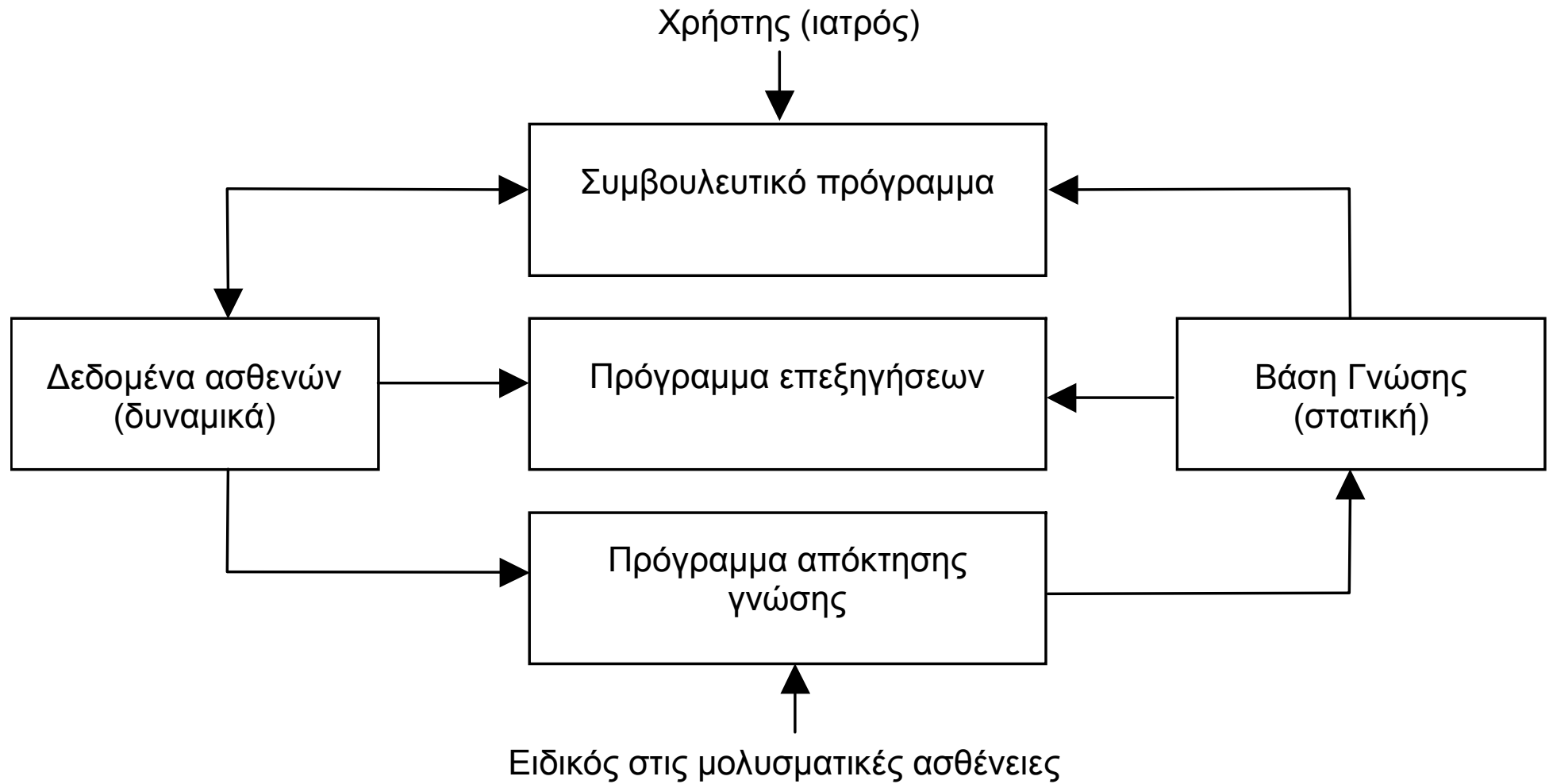
### ❖ Έλεγχος εκτέλεσης:

- Δημιουργία και έλεγχος υποθέσεων (**hypothesize-and-test**) (μέθοδος K1).
- Μεγάλο αρχικό σύνολο υποψήφιων λύσεων (υποθέσεις)
- Κάθε υπόθεση μπορεί να ελεγχθεί με την ύπαρξη ή την απουσία σχετικών δεδομένων
- Μπορεί να γίνει περισσότερο συγκεκριμένη ή να αποκλειστεί
- Η διαδικασία επαναλαμβάνεται, προσθέτοντας περισσότερους περιορισμούς

# Μελέτη Περίπτωσης - Το Σύστημα MYCIN

- ❖ Αντιμετώπιση μολύνσεων του αίματος από βακτήρια-μικρόβια.
  - ❑ Έγκαιρη διάγνωση των πιθανών μικροοργανισμών που προκαλούν τη μόλυνση.
  - ❑ Πρόταση ενός ή περισσότερων αντιβιοτικών για την αντιμετώπισή της.
- ❖ Λειτουργεί ως απόδειξη θεωρημάτων
  - ❑ Επίτευξη ενός στόχου ως επίτευξη μιας σειράς υποστόχων.
  - ❑ Ερευνά το μεγαλύτερο μέρος των εναλλακτικών διαδρομών προς την επίλυση ενός προβλήματος
  - ❑ Αξιολογεί τις εναλλακτικές διαδρομές βάσει κριτηρίων
- ❖ Βήματα:
  - ❑ Λήψη απόφασης για το αν ο ασθενής έχει κάποια σοβαρή μόλυνση.
  - ❑ Καθορισμός των πιθανών μικροοργανισμών (μικροβίων) που εμπλέκονται.
  - ❑ Επιλογή του συνόλου των κατάλληλων φαρμάκων.
  - ❑ Επιλογή του καταλληλότερου φαρμάκου ή συνδυασμού φαρμάκων.

# Βασική Δομή του MYCIN



# Βάση Γνώσης του MYCIN

## ❖ Κανόνες:

**IF** condition<sub>1</sub> AND ... AND condition<sub>n</sub>  
**THEN** conclusion<sub>1</sub> AND ... AND conclusion<sub>n</sub>

### ❑ Παράδειγμα κανόνα που καθορίζει την κατηγορία ενός μικροοργανισμού:

- Ο αριθμός 0.8 είναι η **βεβαιότητα** του κανόνα και καθορίζει πόσο σίγουρο είναι το συμπέρασμά του, με την προϋπόθεση ότι ικανοποιούνται οι συνθήκες του.

**IF** The stain of the organism is Gram negative, and  
The morphology of the organism is rod, and  
The aerobicity of the organism is aerobic  
**THEN** There is strongly suggestive evidence (0.8) that  
the class of the organism is Enterobacteriaceae

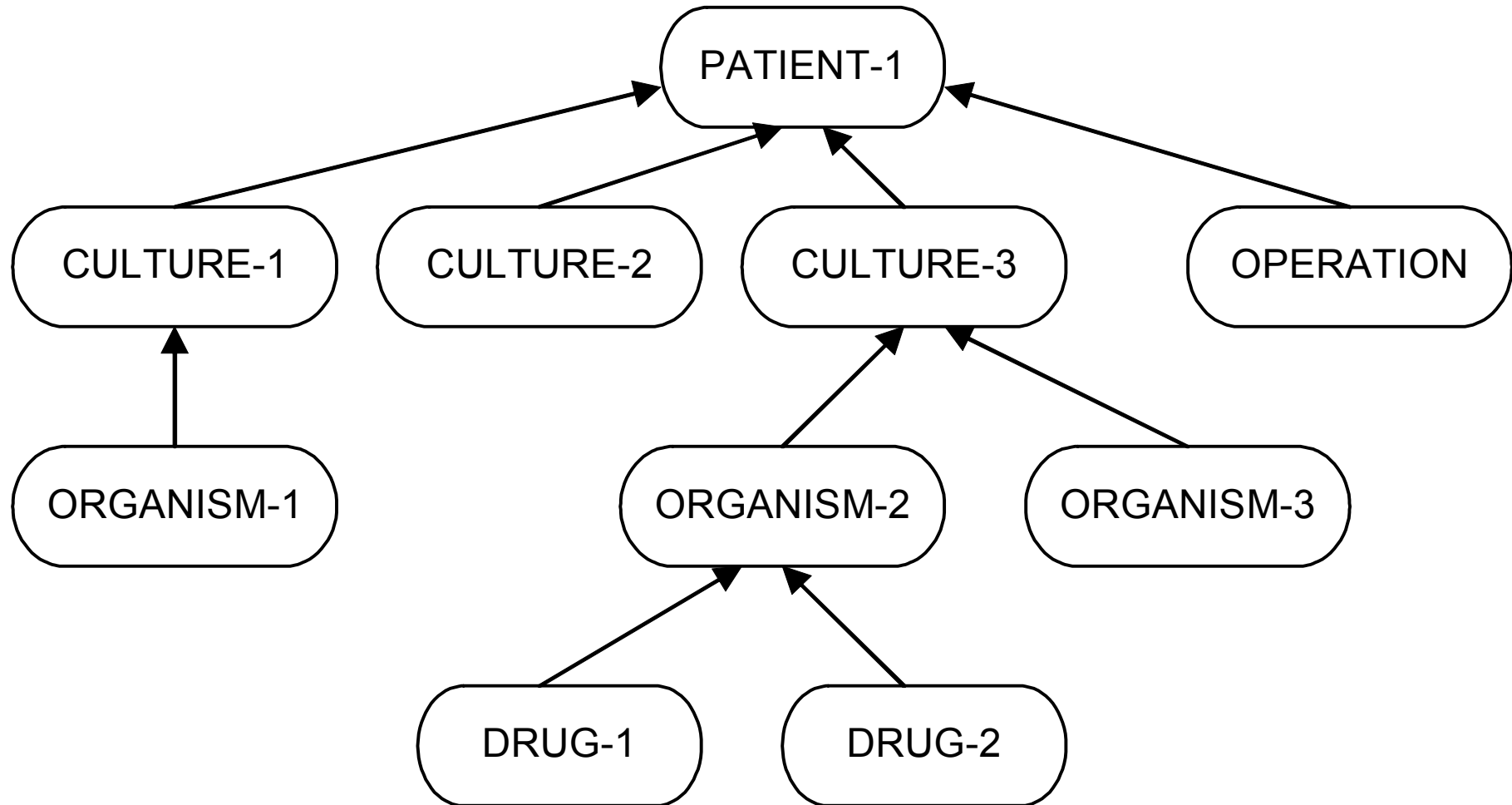
## ❖ Η βάση γνώσης περιέχει επίσης:

- ❑ **Απλές λίστες**, π.χ. η λίστα όλων των μικροοργανισμών που γνωρίζει το σύστημα.
- ❑ **Πίνακες γνώσης**: Εγγραφές κλινικών παραμέτρων και τιμές που παίρνουν
- ❑ Σύστημα **ταξινόμησης κλινικών παραμέτρων** ανάλογα με την κατηγορία στην οποία ανήκουν, π.χ. αν είναι χαρακτηριστικά ασθενών ή μικροοργανισμών.

## ❖ Οι πληροφορίες για τον ασθενή αποθηκεύονται στο **context tree** (δένδρο περιβάλλοντος).



# Παράδειγμα Δένδρου Περιβάλλοντος Ασθενή



# Έλεγχος Εκτέλεσης στο MYCIN

- ❖ Οι κανόνες εκτελούνται ανάστροφα
- ❖ Ο αρχικός στόχος είναι η σύσταση κατάλληλης θεραπείας
  - ❑ Σταδιακά αναλύεται σε απλούστερους υποστόχους
  - ❑ Οι υποστόχοι περιλαμβάνουν τον καθορισμό των εμπλεκομένων μικροοργανισμών και τη διαπίστωση της σοβαρότητάς τους ως παράγοντες της μολυσματικής ασθένειας.
  - ❑ Οι περισσότεροι υποστόχοι έχουν δικούς τους υποστόχους, όπως π.χ. τον καθορισμό των ιδιοτήτων της χρώσης και της μορφολογίας του μικροοργανισμού.
- ❖ Οι απλούστεροι στόχοι είναι η ανάκληση γεγονότων από τη βάση δεδομένων ή το χρήστη.
  - ❑ Εργαστηριακά δεδομένα που δεν μπορούν να εξαχθούν με λογικούς συμπερασμούς.
- ❖ Βασικός στόχος:
  - IF υπάρχει κάποιος μικροοργανισμός που χρειάζεται αντιμετώπιση ΚΑΙ όλοι οι υπόλοιποι μικροοργανισμοί έχουν αντιμετωπισθεί
  - THEN φτιάξε μία λίστα με πιθανές θεραπείες-φάρμακα και εξακρίβωσε την καλύτερη από αυτές.

# Ανάστροφη Ακολουθία Εκτέλεσης στο MYCIN

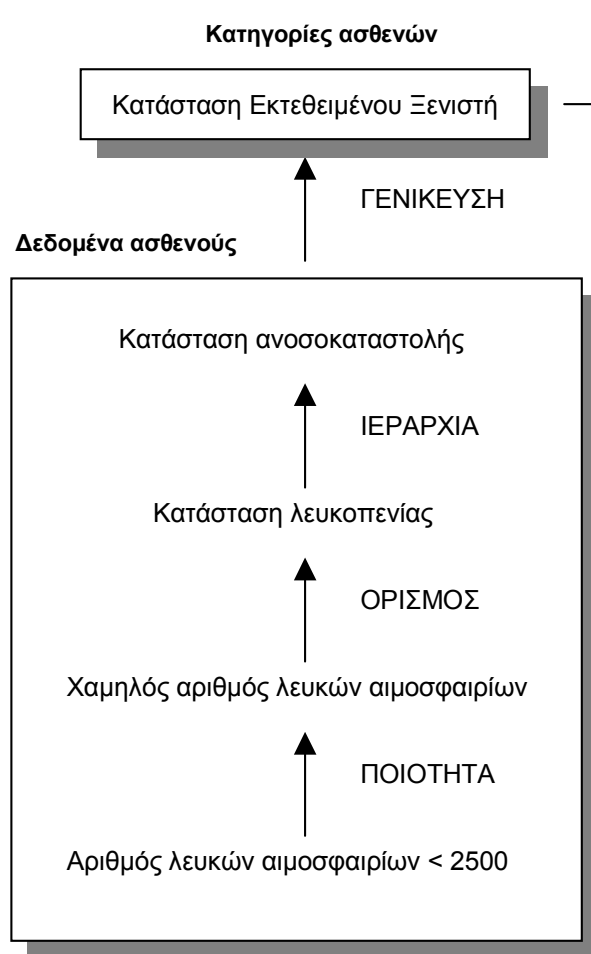
- ❖ Οι κανόνες που περιέχουν τις κύριες παραμέτρους εκτελούνται πριν από τους υπόλοιπους.
- ❖ Πολλοί από τους συμπερασμούς στο MYCIN είναι αβέβαιοι.
  - ❑ Συλλέγονται πληροφορίες από όλους τους σχετικούς κανόνες.
  - ❑ Η βεβαιότητά τους συνδυάζεται για να εξαχθεί η τελική βεβαιότητα κάποιου συμπεράσματος.
  - ❑ Αν κάποιος κανόνας έχει βεβαιότητα 1.0 τότε χρησιμοποιείται μόνο αυτός για την εξαγωγή συμπεράσματος.
  - ❑ Αν κάποιο συμπέρασμα εξαχθεί με βεβαιότητα μεταξύ  $-0.2$  και  $+0.2$ , τότε θεωρείται ότι δεν εξήχθη καθόλου (αυθαίρετη παραδοχή).
  - ❑ Αν κάποια από τις συνθήκες ενός κανόνα είναι από την αρχή σίγουρα ψευδής (βεβαιότητα  $-1.0$ ), τότε ο κανόνας αυτός δεν εξετάζεται καθόλου.
  - ❑ Αν κάποια παράμετρος εξαχθεί με απόλυτη βεβαιότητα (1.0), τότε προηγούνται οι κανόνες που έχουν αυτήν την παράμετρο στη συνθήκη τους.

# Κατηγοριοποίηση στο MYCIN

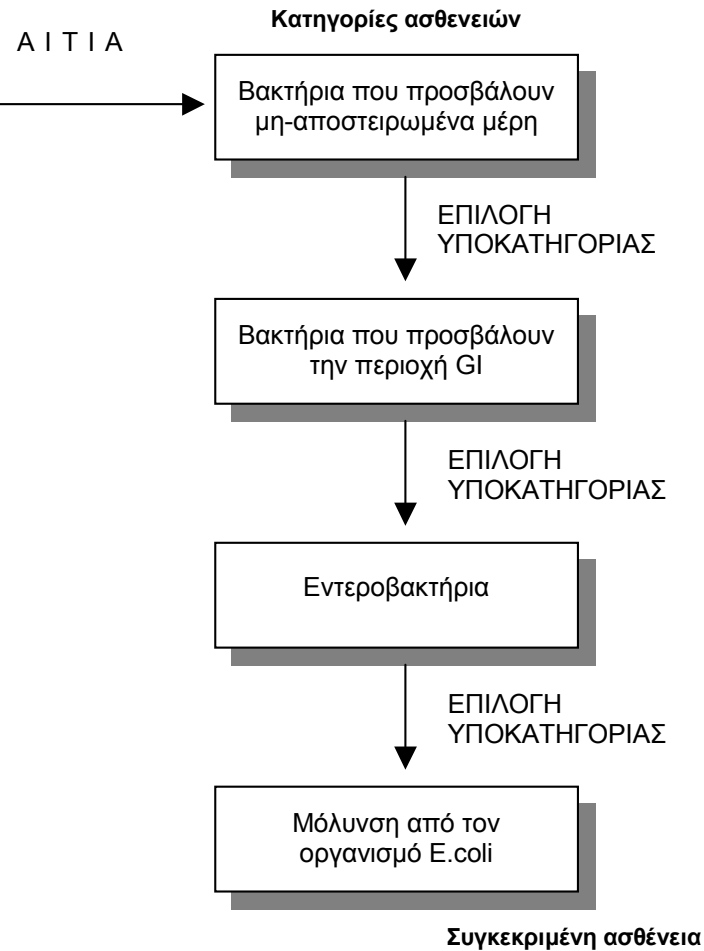
- ❖ Χρησιμοποιεί μέθοδο κατηγοριοποίησης που είναι πιο κοντά στην K3.
    - ❑ Το σύστημα οδηγείται από τις λύσεις προς τα δεδομένα.
    - ❑ Χρησιμοποιεί δένδροειδή ιεράρχηση των υπο-στόχων.
    - ❑ Το MYCIN χρησιμοποιεί πρώτα σε βάθος αναζήτηση (ενώ η K3 πρώτα σε πλάτος).
  - ❖ Παράδειγμα κανόνα:
    - IF      A complete blood count is available AND  
          The white blood count is less than 2500/cm<sup>3</sup>
    - THEN   The following bacteria might be causing infection:  
          E.coli (0.75), Pseudomonas (0.5), Pneumonia (0.5)
  - ❑ Η πρώτη σειρά του κανόνα αποτρέπει τη χρήση του στην περίπτωση που δεν είναι διαθέσιμα τα αποτελέσματα της εξέτασης του αίματος.
  - ❑ Αν δεν υπήρχε, το δεύτερο μέρος θα ζητούσε αυτά τα δεδομένα από το χρήστη.
  - ❑ Το τρίτο μέρος προσδιορίζει πιθανές υποψήφιες λύσεις.
- ❖ Βασικά στάδια κατηγοριοποίησης:
  - ❑ Χρησιμοποιούνται οι σχέσεις **γενίκευσης**, για να βρεθεί μία γενική κατηγορία ασθενούς για την οποία μπορεί να βρει μία πιθανή **αιτία** (ασθένεια).
  - ❑ Η γενική κατηγορία γίνεται ολοένα και πιο συγκεκριμένη, **επιλέγοντας υποκατηγορίες** της αρχικής ασθένειας από την ιεραρχία των μικροβίων.

# Βασικά Στάδια Κατηγοριοποίησης στο ΜΥCIN

Χώρος Δεδομένων



Χώρος Κατηγοριών-Λύσεων



# Μελέτη Περίπτωσης - Το Σύστημα PROSPECTOR

- ❖ Έμπειρο σύστημα για την αξιολόγηση γεωλογικών δεδομένων
  - ❑ Καθορισμός πιθανότητα ύπαρξης αξιόλογων ορυκτών κοιτασμάτων σε μια περιοχή
- ❖ Χρησιμοποιήθηκε πειραματικά μόνο 1 φορά,
  - ❑ Προέβλεψε την ύπαρξη ορυκτού κοιτάσματος μολυβδαίνιου
- ❖ Λειτουργία
  - ❑ Δέχεται παρατηρήσεις (δεδομένα) και κάνει ερωτήσεις για την πιθανή ύπαρξη χρήσιμου αποθέματος
  - ❑ Αν η πιθανότητα είναι καλή, εμφανίζει γραφική απεικόνιση των πιθανών σημείων γεώτρησης
- ❖ Η γνώση αναπαριστάται με:
  - ❑ **Σημασιολογικά δίκτυα:** Αναπαριστούν γεωλογικές γνώσεις
  - ❑ **Δίκτυα Συμπερασμού:** Αναπαριστούν τους κανόνες

# Σημασιολογικά Δίκτυα

❖ Αναπαριστούν γεωλογικές γνώσεις.

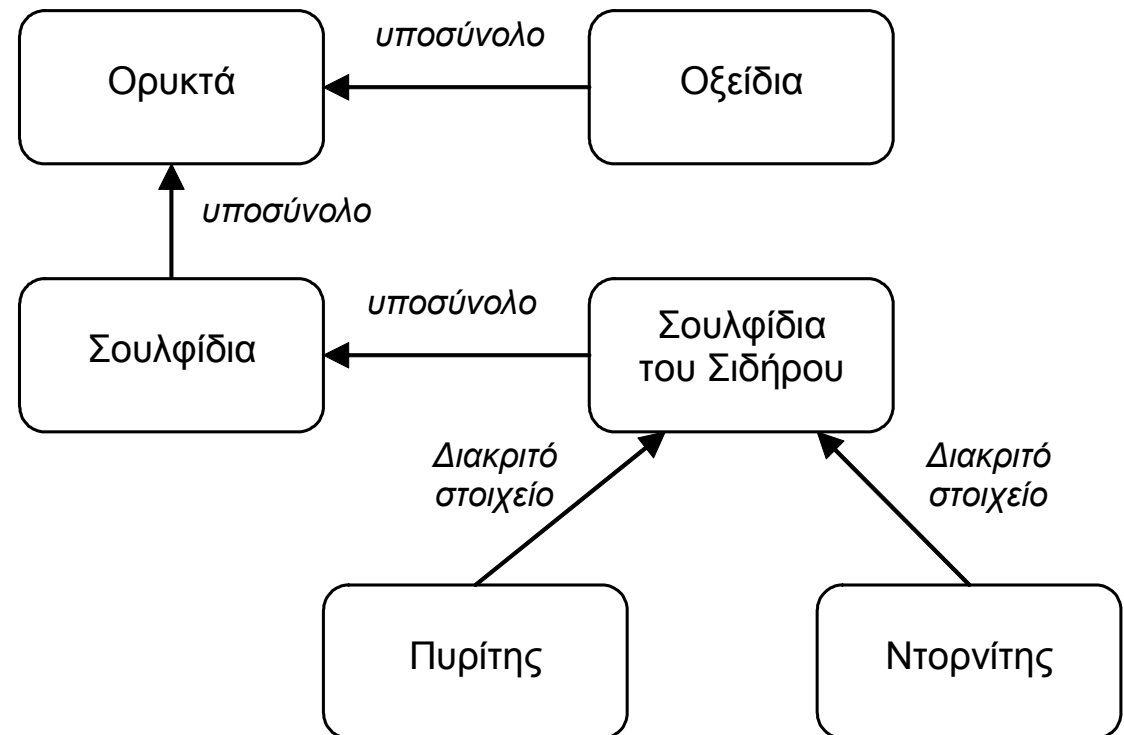
❖ Χρήσεις:

□ Εξαγωγή συμπερασμάτων

- Π.χ., αν υπάρχουν πυρίτες στην περιοχή, τότε υπάρχει θειούχος σίδηρος και θειούχες ενώσεις

□ Έλεγχος πληροφοριών που δίνει ο χρήστης

- Π.χ., αν δεν υπάρχουν θειούχα κοιτάσματα στην περιοχή δεν μπορούν να υπάρχουν και πυρίτες.



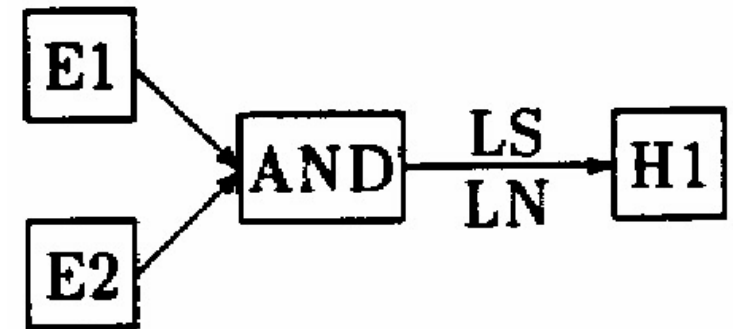
# Δίκτυα Συμπερασμού

## ❖ Αναπαριστούν τους κανόνες.

- ❑ Υπάρχει ένα δίκτυο για κάθε ορυκτό.
- ❑ Τα τόξα αναπαριστούν τη συνεπαγωγή ενώ οι κόμβοι τις λογικές πράξεις AND/OR
- ❑ Κάθε κόμβος έχει μια προϋπάρχουσα πιθανότητα
- ❑ Με τη **λογική επάρκεια (LS)** και **αναγκαιότητα (LN)** η πιθανότητα κάθε κόμβου μεταβάλλεται, λόγω της παρουσίας άλλων πληροφοριών

## ❖ Οι πιθανότητες και τα μεγέθη LS, LN παρέχονται από τους ειδικούς-γεωλόγους

- ❑ Υποκειμενικότητα - Πιθανότητα σφάλματος.
- ❑ Ο τρόπος δόμησης της γνώσης και της αβεβαιότητας είναι κατάλληλος μόνο για μικρό αριθμό δικτύων (15 δίκτυα, 150 κανόνες)





# Λογική Επάρκεια και Λογική Αναγκαιότητα

## ❖ Λογική Επάρκεια (Logical Sufficiency – LS)

- ❑ Πόσο πιθανότερο είναι να συνδεθεί ένα γεγονός  $E$  με την αλήθεια ενός υποθετικού συμπεράσματος  $H$ , παρά με την άρνηση του  $H$  (συμβολίζεται  $\neg H$ )

$$LS = \frac{P(E | H)}{P(E | \neg H)}$$

## ❖ Λογική Αναγκαιότητα (Logical Necessity – LN)

- ❑ Πόσο πιθανότερο είναι να συνδεθεί η απουσία ενός γεγονότος  $E$  με την αλήθεια του υποθετικού συμπεράσματος  $H$  παρά με την άρνηση του  $H$

$$LN = \frac{P(\neg E | H)}{P(\neg E | \neg H)}$$

# Έλεγχος Εκτέλεσης στο PROSPECTOR

- ❖ **Εισαγωγή Δεδομένων:** Ο χρήστης παραθέτει ένα σύνολο παρατηρήσεων.
  - ❑ Κάθε πληροφορία συνοδεύεται από μία τιμή από  $-5$  έως  $5$
  - ❑ Υπολογίζεται η πιθανότητα ύπαρξης της παρατήρησης  $E$ , βάσει των παρατηρήσεων  $E'$  του χρήστη
  - ❑  $-5 \rightarrow P(E|E')=0$ : το  $E$  δεν υπάρχει βάσει των παρατηρήσεων του χρήστη.
  - ❑  $0 \rightarrow P(E|E')=P(E)$ : Οι παρατηρήσεις δε μεταβάλλουν τις προϋπάρχουσες πιθανότητες.
  - ❑  $+5 \rightarrow P(E|E')=1$ : Το  $E$  υπάρχει βάσει των παρατηρήσεων.
- ❖ **Προώθηση Πιθανοτήτων:** Το σύστημα προωθεί στο δίκτυο τις μεταβολές των πιθανοτήτων με **ορθή ακολουθία εκτέλεσης**
  - ❑ Συνεχίζεται έως ότου προκύψει η ύπαρξη ή όχι κάποιου ορυκτού στην περιοχή
  - ❑ Το πιο πιθανό δίκτυο επιλέγεται ως υποψήφιο για την επόμενη φάση
- ❖ **Επιβεβαίωση Υπόθεσης:** Το σύστημα κινείται από το συμπέρασμα προς τις παρατηρήσεις
  - ❑ Η διαδικασία συνεχίζεται έως ότου το σύστημα φθάσει σε τερματικούς κόμβους.

# Κατηγοριοποίηση στο PROSPECTOR

- ❖ Χρησιμοποιεί τη μέθοδο κατηγοριοποίησης K4.
  - ❑ Αρχικά οδηγείται από τα δεδομένα του χρήστη σε πιθανές λύσεις.
  - ❑ Μετά προσπαθεί να φτάσει από τις υποψήφιες λύσεις σε δεδομένα που τις στηρίζουν
- ❖ Χρησιμοποιεί ενδιάμεσες υποθέσεις στην κατηγοριοποίηση.
  - ❑ Περιοχή που βρίσκεται κάποιο πέτρωμα, ηλικία και τρόπος σχηματισμού του, κλπ.

