ΑΝΟΙΚΤΑ ΑΚΑΔΗΜΑΪΚΑ ΜΑΘΗΜΑΤΑ

# Θεωρία της Πληροφορίας (Θ)

## Ενότητα 3: Κωδικοποίηση Πηγής

ΔΙΔΑΣΚΩΝ: Δρ. Αναστάσιος Πολίτης

ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΦΑΡΜΟΓΩΝ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΤΕ

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο

ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

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 TEI 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.

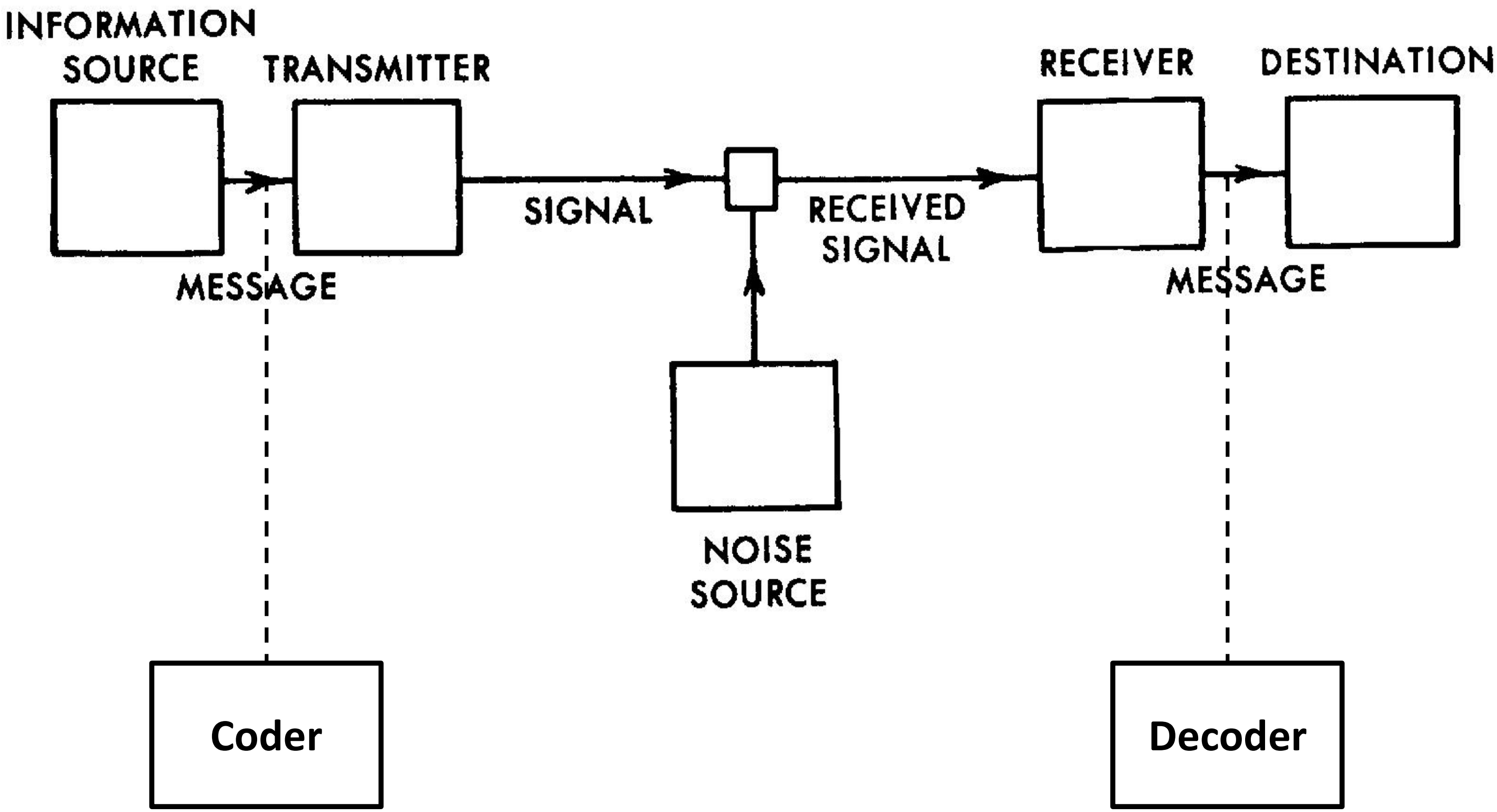
Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

# Ενότητα 3

## Κωδικοποίηση Πηγής

**ΔΙΔΑΣΚΩΝ: Δρ. Αναστάσιος Πολίτης**

# Κωδικοποίηση Πηγής




- Μεταξύ πομπού και καναλιού παρεμβάλλεται ο κωδικοποιητής (coder).
- Έργο του:
  - η αντικατάσταση των συμβόλων πληροφορίας της πηγής με εναλλακτικά σύμβολα ή λέξεις.

# Κωδικοποίηση Πηγής

- Λόγοι για την κωδικοποίηση:
  - ***Συμπίεση***
    - στόχος η μείωση του συνολικού μήκους του μηνύματος
    - εξοικονόμηση χώρου αποθήκευσης.
    - μείωση του χρόνου αποστολής.
    - περιορίζει την απαιτούμενη χωρητικότητα διαύλου για την μεταφορά πληροφοριών.
  - ***Ασφάλεια***
    - ο κωδικοποιητής αντιστοιχεί τα σύμβολα πηγής με σύνολο λέξεων που είναι γνωστά μόνο στον πομπό και δέκτη.
    - αποφυγή υποκλοπής και κατανόησης.
  - ***Ανίχνευση και διόρθωση σφαλμάτων***
    - αντικατάσταση συμβόλων πηγής με λέξεις μεγαλύτερου μεγέθους, που περιλαμβάνουν πληροφορίες για το αρχικό μήνυμα.
    - εισαγωγή πλεονάζουσας πληροφορίας.
    - αντίθετη αρχή λειτουργίας με την συμπίεση.

- **Θα μας απασχολήσει η κωδικοποίηση με στόχο την συμπίεση!**

# Ορισμοί

- **Ορισμός 1:**
  - ***Κωδικοποίηση*** *είναι η αντικατάσταση συμβόλων ή ομάδων συμβόλων πληροφορίας με νέα εναλλακτικά σύμβολα ή ομάδες συμβόλων.*
- **Όρισμός 2:**
  - ***Κώδικας*** *είναι ο αλγόριθμος με τον οποίο πραγματοποιείται η κωδικοποίηση, δηλαδή, η αντικατάσταση συμβόλων από νέα σύμβολα.*
- **Ορισμός 3:**
  - *Τα νέα διακεκριμένα σύμβολα $w_1$, $w_2$, …, $w_M$ που προκύπτουν κατά την κωδικοποίηση ονομάζονται* ***κωδικά σύμβολα.*** *Το πλήθος τους Μ μπορεί να είναι πεπερασμένο ή άπειρο.*
- **Ορισμός 4:**
  - *Το σύνολο $W=\{w_1, w_2, \ldots, w_M\}$ όλων των κωδικών συμβόλων λέγεται* ***αλφάβητο κωδικοποίησης****.*
- **Ορισμός 5:**
  - *Ο κώδικας που έχει Μ κωδικά σύμβολα ονομάζεται* ***Μ-αδικός κώδικας.***

# Παραδείγματα

• Κωδικοποίηση **Morse**

| | | | | | |
|---|---|---|---|---|---|
| *A* | .– | *J* | .––– | *S* | ... |
| *B* | –... | *K* | –.– | *T* | – |
| *C* | –.–. | *L* | .–.. | *U* | ..– |
| *D* | –.. | *M* | –– | *V* | ...– |
| *E* | . | *N* | –. | *W* | .–– |
| *F* | ..–. | *O* | ––– | *X* | –..– |
| *G* | ––. | *P* | .––. | *Y* | –.–– |
| *H* | .... | *Q* | ––.– | *Z* | ––.. |
| *I* | .. | *R* | .–. | | |

• Απεικόνιση γραμμάτων αλφαβήτου σε κωδικές λέξεις που σχηματίζονται από τελείες, παύλες και κενά διαστήματα.

• Κωδικά σύμβολα: **. – κενό**

• Πλήθος κωδικών συμβόλων M=**3**

• Αλφάβητο Κωδικοποίησης W={**., -, κενό**}

• Ο κώδικας Morse είναι ένας 3-αδικός κώδικας.

# Παραδείγματα

- Κωδικοποίηση **ASCII-7** (American Standard Code for Information Interchange)

ASCII Code Chart

| Bits 3-0 | Bits 6-4 | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 000 | 001 | 010 | 011 | 100 | 101 | 110 | 111 |
| 0000 | NULL | DLE | | 0 | @ | P | ` | p |
| 0001 | SOH | DC1 | ! | 1 | A | Q | a | q |
| 0010 | STX | DC2 | " | 2 | B | R | b | r |
| 0011 | ETX | DC3 | # | 3 | C | S | c | s |
| 0100 | EOT | DC4 | $ | 4 | D | T | d | t |
| 0101 | ENQ | NAK | % | 5 | E | U | e | u |
| 0110 | ACK | SYN | & | 6 | F | V | f | v |
| 0111 | BEL | ETB | ' | 7 | G | W | g | w |
| 1000 | BS | CAN | ( | 8 | H | X | h | x |
| 1001 | HT | EM | ) | 9 | I | Y | i | y |
| 1010 | LF | SUB | * | : | J | Z | j | z |
| 1011 | VT | ESC | + | ; | K | [ | k | { |
| 1100 | FF | FS | , | < | L | \ | l | \| |
| 1101 | CR | GS | - | = | M | ] | m | } |
| 1110 | SO | RS | . | > | N | ^ | n | ~ |
| 1111 | SI | US | / | ? | O | _ | o | DEL |

- Κωδικά σύμβολα: **0** και **1**
- Πλήθος κωδικών συμβόλων **M=2**
- Αλφάβητο Κωδικοποίησης W={**0,1**}
- Ο κώδικας ASCII-7 είναι ένας 2-αδικός κώδικας

# Ταξινόμηση Κωδίκων

- Ταξινόμηση με κριτήριο τις απώλειες
  - ***Κώδικες χωρίς απώλειες***
    - Η απεικόνιση της πηγής είναι ένα-προς-ένα (δηλαδή ένα σύμβολο αλφαβήτου πηγής αντιστοιχεί σε ένα σύμβολο αλφαβήτου κωδικοποίησης ή μια κωδική λέξη).
    - Η αποκωδικοποίηση του κωδικού μηνύματος θα μας δώσει το αρχικό μήνυμα (καμία απώλεια).
    - π.χ. Κώδικας Morse.
  - ***Κώδικες με απώλειες***
    - Η απεικόνιση συμβόλων πηγής σε κωδικά σύμβολα είναι πολλά-σε-ένα.
    - η αποκωδικοποίηση ενδέχεται να μας δώσει διαφορετικό μήνυμα από αυτό που μεταδόθηκε.
    - π.χ. Κωδικοποίηση έγχρωμης εικόνας σε αντίστοιχη με διαβαθμίσεις του γκρί.
- **Θα μας απασχολήσουν ΜΟΝΟ κώδικες χωρίς απώλειες.**

# Ταξινόμηση Κωδίκων

- Ταξινόμηση με κριτήριο το μήκος των κωδικών λέξεων
  - ***Σταθερού μήκους***
    - το μήκος των κωδικών λέξεων είναι σταθερό για κάθε σύμβολο της πηγής Χ.
    - π.χ. **ASCII-7** (κάθε χαρακτήρας πηγής κωδικοποιείται με έναν συνδυασμό από 7 δυαδικών ψηφίων).
  - ***Μεταβλητού μήκους***
    - χρησιμοποιούνται για την συμπίεση δεδομένων.
    - Τα σύμβολα πηγής με μεγαλύτερη πιθανότητα εμφάνισης, κωδικοποιούνται με μικρότερου μήκους κωδικές λέξεις.
    - π.χ. **Κώδικας Morse –** το γράμμα *E* που είναι το πιο σύνηθες στην αγγλική αντιστοιχίζεται με μικρότερη κωδική λέξη από το γράμμα *Q* που είναι το πιο σπάνια εμφανιζόμενο.
- **Δεν μας απασχολούν διαδικασίες ανίχνευσης και διόρθωσης σφαλμάτων**
  - Το κανάλι είναι «αθόρυβο»
    - *είναι το κανάλι που δέχεται ως είσοδο ένα σύνολο από «κωδικούς χαρακτήρες» και αναπαράγει στην έξοδο του τους χαρακτήρες αυτούς χωρίς πιθανότητα λάθους.* ***(Βλ. Ιδανικός Δίαυλος)***

# “Αθόρυβη” (Noiseless )Κωδικοποίηση

- Στόχος είναι:
  - η ελαχιστοποίηση του **μέσου** μήκους της κωδικής λέξης.
- Συνοπτικά, τα «συστατικά» του προβλήματος της «αθόρυβης κωδικοποίησης» είναι:
  - **Η πηγή εισόδου** $(X, P_X)$ στον ιδανικό δίαυλο με αλφάβητο πηγής $X=\{x_1, x_2, \ldots, x_N\}$ και κατανομή πιθανοτήτων $P_X=\{p_1, p_2, \ldots, p_N\}$.
  - **Το αλφάβητο κωδικοποίησης** $W=\{w_1, w_2, \ldots, w_M\}$. Σε κάθε σύμβολο $x_i$ πρόκειται να ανατεθεί μία πεπερασμένη ακολουθία από κωδικά σύμβολα η οποία αποτελεί την **κωδική λέξη.**
    - Π.χ. το σύμβολο $x_1$ μπορεί να αντιστοιχηθεί στην κωδική λέξη $w_1w_2$ και το $x_2$ στην λέξη $w_3w_7w_3w_8$.

# Σχεδιαστικά Ζητήματα Κωδίκων

- Έστω ο παρακάτω δυαδικός κώδικας:

| | |
|---|---|
| $\mathbf{x_1}$ | **10** |
| $\mathbf{x_2}$ | **010** |
| $\mathbf{x_3}$ | **0** |
| $\mathbf{x_4}$ | **10** |

- Ποιό είναι το πρόβλημα του κώδικα αυτού;
  - Τα σύμβολα της πηγής $\mathbf{x_1}$ και $\mathbf{x_4}$ αντιστοιχίζονται στην ίδια κωδική λέξη.
  - Ο δέκτης θα παρουσιάζει δυσκολία αποκωδικοποίησης της λέξης «**10**».
- Κατά τον σχεδιασμό θα πρέπει να αποφεύγονται τέτοιες αντιστοιχίσεις που οδηγούν σε «σύγχιση» τον δέκτη.
- **Ορισμός:**
  - ένας κώδικας που αντιστοιχεί κάθε σύμβολο της πηγής σε διαφορετική κωδική λέξη, λέγεται ***ευκρινής.***
- Εξασφαλίζουμε την παραπάνω σχεδιαστική απαίτηση αποφεύγοντας να αναθέσουμε την ίδια κωδική λέξη σε παραπάνω από ένα σύμβολα πηγής.

# Σχεδιαστικά Ζητήματα Κωδίκων

- Έστω ο παρακάτω δυαδικός κώδικας:

| | |
|---|---|
| $x_1$ | **0** |
| $x_2$ | **010** |
| $x_3$ | **01** |
| $x_4$ | **10** |

- Ποιό είναι το πρόβλημα του κώδικα αυτού;
  - Στον δέκτη καταφθάνει η ακολουθία κωδικών συμβόλων «**010**». Ποιό σύμβολο ή ακολουθία συμβόλων θα προκύψει κατά την αποκωδικοποίηση; **Απ.**: $x_2$, $x_1x_4$, $x_3x_1$
  - Ο δέκτης δεν είναι σε θέση να αντιληφθεί με απόλυτη βεβαιότητα την αρχή και το τέλος μιας κωδικής λέξης.
- Είναι επιθυμητό να εξαλειφθούν καταστάσεις σύγχισης όπως η παραπάνω.
- **Ορισμός:**
  - ένας κώδικας για τον οποίο κάθε κωδική λέξη αναγνωρίζεται με βεβαιότητα μέσα σε μία μακρά ακολουθία κωδικών συμβόλων λέγεται ***μονοσήματος.***

# Σχεδιαστικά Ζητήματα Κωδίκων

- Ένας τρόπος για την εξασφάλιση της μονοσημαντικότητας ενός κώδικα, είναι καμία κωδική λέξη να είναι **πρόθεμα** μιας άλλης. Η ιδιότητα αυτή λέγεται **προθεματική ιδιότητα.**
- **Προσοχή!**
  - ένας κώδικας στον οποίο καμία κωδική λέξη δεν είναι πρόθεμα μιας άλλης, λέμε ότι **έχει την προθεματική ιδιότητα.**
- Ένας κώδικας για τον οποίο ισχύει η προθεματική ιδιότητα ονομάζεται **στιγμιαία αποκωδικοποιήσιμος. (instantaneous)**
- Πρακτικά, ένας στιγμιαία αποκωδικοποιήσιμος κώδικας επιτρέπει την αποκωδικοποίηση των κωδικών λέξεων που βρίσκονται σε μια ακολουθία χωρίς να χρειάζεται να εξεταστούν οι γειτονικές (στιγμιαία).

| | |
|---|---|
| $\mathbf{x_1}$ | **0** |
| $\mathbf{x_2}$ | **10** |
| $\mathbf{x_3}$ | **11** |

Λαμβανόμενη ακολουθία: 10000110
Αποκωδικοποίηση: $\mathbf{x_2\ x_1\ x_1\ x_1\ x_3\ x_1}$

# Σχεδιαστικά Ζητήματα Κωδίκων

- Κάθε στιγμιαία αποκωδικοποιήσιμος κώδικας είναι και μονοσήμαντος. Δεν ισχύει το αντίστροφο!
    - *Παράδειγμα*

| | |
|---|---|
| $x_1$ | **1** |
| $x_2$ | **10** |
| $x_3$ | **100** |
| $x_4$ | **1000** |

Δεν ισχύει η προθεματική ιδιότητα → μη στιγμιαία αποκωδικοποιήσιμος, αλλά ο κώδικας είναι μονοσήμαντος!

- Κάθε μονοσήμαντος κώδικας είναι και ευκρινής. Δεν ισχύει το αντίστροφο!
    - *Παράδειγμα*

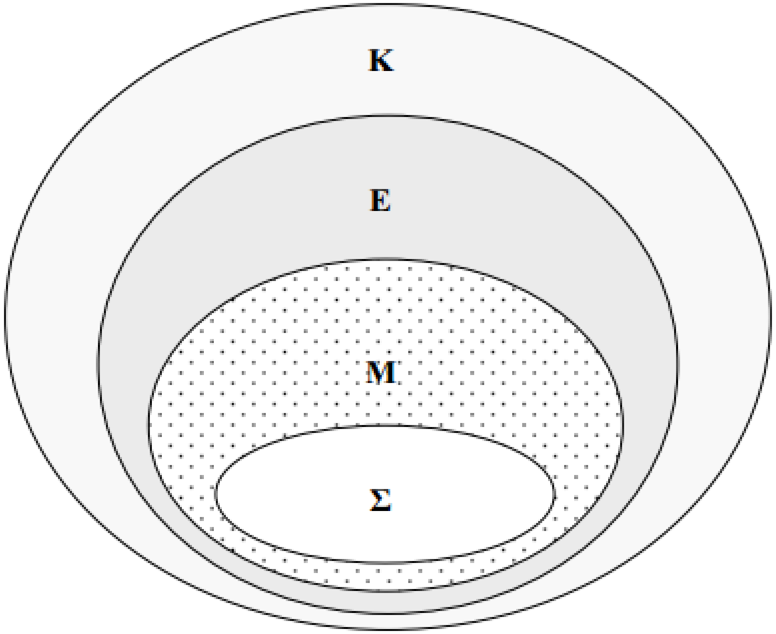
| | |
|---|---|
| $x_1$ | **0** |
| $x_2$ | **01** |
| $x_3$ | **10** |
| $x_4$ | **11** |

Ο κώδικας είναι ευκρινής αλλά η ακολουθία 0110 οδηγεί είτε σε $\mathbf{x_2x_3}$ είτε σε $\mathbf{x_1x_4x_1}$

- Κάθε ευκρινής κώδικας σταθερού μήκους είναι και στιγμιαία αποκωδικοποιήσημος.

# Σχεδιαστικά Ζητήματα Κωδίκων

- Γενικά ισχύει:




**K**: όλοι οι κώδικες
**E:** Ευκρινείς
**M:** Μονοσήμαντοι
**Σ:** Στιγμιαία Αποκωδικοποιήσιμοι

- Περιπτώσεις:
  - Ένας **Σ** κώδικας είναι αυτομάτως και **M** και **E**.
  - Ένας **M** κώδικας είναι αυτομάτως και **E** αλλά μπορεί να μην είναι **Σ**.
  - Ένας **E** κώδικας μπορεί να μην είναι **M** ή/και **Σ**. (Στην περίπτωση σταθερού μήκους είναι αυτομάτως και **Σ** και **M**)
  - Ένας κώδικας που δεν είναι **Σ** μπορεί να είναι **M**.
  - Ένας κώδικας που δεν είναι **M** μπορεί να είναι **E**.
  - Ένας κώδικας που δέν είναι **E** δεν είναι **Σ** και **M**.

# Εφαρμογές

- **Εφαρμογή Ι:** Να ταξινομηθούν οι παρακάτω κώδικες.

| **Σύμβολο** | **Κώδικας Α** | **Κώδικας Β** | **Κώδικας Γ** | **Κώδικας Δ** |
|---|---|---|---|---|
| $\mathbf{x_1}$ | **00** | **0** | **1** | **1** |
| $\mathbf{x_2}$ | **10** | **01** | **10** | **01** |
| $\mathbf{x_3}$ | **01** | **10** | **100** | **001** |
| $\mathbf{x_4}$ | **00** | **11** | **1000** | **0001** |

| **Ταξινόμιση** | **Κώδικας Α** | **Κώδικας Β** | **Κώδικας Γ** | **Κώδικας Δ** |
|---|---|---|---|---|
| Ευκρινής | **x** | √ | √ | √ |
| Μονοσήμαντος | **x** | **x** | √ | √ |
| Στιγμιαία αποκωδικοποιήσιμος | **x** | **x** | **x** | √ |

# Εφαρμογές

- **Εφαρμογή ΙΙ:** Να ταξινομηθούν οι παρακάτω κώδικες.

| Σύμβολο | Κώδικας Α | Κώδικας Β | Κώδικας Γ | Κώδικας Δ | Κώδικας Ε |
|---|---|---|---|---|---|
| $x_1$ | 0 | 00 | 110 | 000 | 01 |
| $x_2$ | 01 | 11 | 01 | 010 | 011 |
| $x_3$ | 11 | 10 | 00 | 101 | 0111 |
| $x_4$ | 10 | 01 | 10 | 111 | 0 |

| Ταξινόμιση | Κώδικας Α | Κώδικας Β | Κώδικας Γ | Κώδικας Δ | Κώδικας Ε |
|---|---|---|---|---|---|
| ευκρινής | √ | √ | √ | √ | √ |
| Μονοσήμαντος | **X** | √ | √ | √ | √ |
| Στιγμιαία αποκωδικοποιήσιμος | **X** | √ | √ | √ | **X** |

# Έλεγχος Μονοσημαντικότητας

- Ένας κώδικας που δεν είναι στιγμιαία αποκωδικοποιήσιμος μπορεί να είναι μονοσήμαντος.
- Είναι δυνατόν να ελέγξουμε πολύπλοκους κώδικες ως προς την μονοσημαντικότητα ακολουθώντας έναν έλεγχο που ονομάζεται **Έλεγχος Sardinas-Patterson** και παρουσιάζεται εδώ χωρίς απόδειξη.
- Η τεχνική βασίζεται στην διαδοχική κατασκευή συνόλων και έλεγχο των στοιχείων τους.
- Έστω ότι διαθέτουμε τον παρακάτω κώδικα:

| | **Κώδικας** |
|---|---|
| $x_1$ | **a** |
| $x_2$ | **c** |
| $x_3$ | **ad** |
| $x_4$ | **abb** |
| $x_5$ | **bad** |
| $x_6$ | **deb** |
| $x_7$ | **bbcde** |

- Παρατηρούμε οτι ο κώδικας δεν είναι στιγμιαία αποκωδικοποιήσιμος:
  - η κωδική λέξη (σύμβολο) «a» είναι πρόθεμα των «ad» και «abb».

# Έλεγχος Μονοσημαντικότητας

- Κατασκευάζουμε βήμα-βήμα μία ακολουθία από σύνολα $S_0$, $S_1$, $S_2$, …, $S_n$ όπως παρακάτω:
  - Το πρώτο σύνολο $S_0$ έχει ως στοιχεία τις αρχικές κωδικές λέξεις.
  - Για την κατασκευή του $S_1$ συγκρίνουμε όλες τις κωδικές λέξεις μέσα στο $S_0$. Εάν μια κωδική λέξη είναι πρόθεμα μιας άλλης, τότε το επίθεμα (αυτό που απομένει δηλαδή) της τελευταίας εισάγεται ως στοιχείο στο $S_1$.

**Κώδικας**

| | |
|---|---|
| $x_1$ | **a** |
| $x_2$ | **c** |
| $x_3$ | **ad** |
| $x_4$ | **abb** |
| $x_5$ | **bad** |
| $x_6$ | **deb** |
| $x_7$ | **bbcde** |

**Σύνολα**

| $S_0$ | $S_1$ |
|---|---|
| **a** | **d** |
| **c** | **bb** |
| **ad** | |
| **abb** | |
| **bad** | |
| **deb** | |
| **bbcde** | |

- Κατόπιν εκτελούμε τον ακόλουθο έλεγχο:
  - Υπάρχει κάποιο στοιχείο των δύο συνόλων που είναι ίδια;
    - **Ναι;** Τότε ο κώδικας δεν είναι μονοσήμαντος.
    - **Όχι;** Η διαδικασία συνεχίζεται όπως παρακάτω.

# Έλεγχος Μονοσημαντικότητας

- Για τα υπόλοιπα σύνολα $S_n$, $n>1$ συγκρίνουμε τα στοιχεία των $S_0$ και $S_{n-1}$ και ψαχνουμε για κωδικές λέξεις του ενός συνόλου που είναι πρόθεμα του άλλου. Το επίθεμα τοποθετείται στο νέο σύνολο.
- **Περιπτώσεις:**
  - Εάν δεν βρεθούν επιθέματα τότε το $S_n$ είναι το κενό σύνολο και η διαδικασία σταματά.
    - Ο κώδικας **είναι μονοσήμαντος.**
  - Εάν το προκύπτον σύνολο $S_n$ είναι ίσο με κάποιο από τα προηγούμενα σύνολα εκτός του $S_0$ η διαδικασία σταματά.
    - Ο κώδικας **είναι μονοσήμαντος**.
  - Εάν κάποιο από τα στοιχεία του $S_n$ είναι ίδιο με κάποιο από τα στοιχεία του $S_0$ η διαδικασία σταματά.
    - Ο κώδικας **δεν είναι μονοσήμαντος.**
  - Εάν δεν συμβεί τίποτε από τα παραπάνω η διαδικασία συνεχίζεται ομοίως.

# Έλεγχος Μονοσημαντικότητας

**Κώδικας**

| | |
|---|---|
| $x_1$ | **a** |
| $x_2$ | **c** |
| $x_3$ | **ad** |
| $x_4$ | **abb** |
| $x_5$ | **bad** |
| $x_6$ | **deb** |
| $x_7$ | **bbcde** |

**Σύνολα**

| $S_0$ | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | $S_7$ | $S_8$ |
|---|---|---|---|---|---|---|---|---|
| **a** | **d** | **eb** | **de** | **b** | **<u>ad</u>** | **d** | **eb** | **{empty}** |
| **c** | **bb** | **cde** | | | **bcde** | | | |
| **<u>ad</u>** | | | | | | | | |
| **abb** | | | | | | | | |
| **bad** | | | | | | | | |
| **deb** | | | | | | | | |
| **bbcde** | | | | | | | | |

- Ο παραπάνω κώδικας <u>δεν</u> είναι στιγμιαία αποκωδικοποιήσιμος διότι ένα στοιχείο του $S_5$ είναι ίδιο με ένα στοιχείο του $S_0$ («ad»).

# Εφαρμογές

- **Εφαρμογή Ι:** Να ελεγχθεί ως προς την μονοσημαντικότητα ο παρακάτω κώδικας:

| | |
|---|---|
| $x_1$ | 010 |
| $x_2$ | 0001 |
| $x_3$ | 0110 |
| $x_4$ | 1100 |
| $x_5$ | 00011 |
| $x_6$ | 00110 |
| $x_7$ | 11110 |
| $x_8$ | 101011 |

Απ:

| *S0* | *S1* | *S2* | *S3* | *S4* | *S5* | *S6* |
|---|---|---|---|---|---|---|
| 010 | 1 | 100 | 11 | 00 | 01 | 0 |
| 0001 | | 1110 | | 110 | 011 | 10 |
| 0110 | | 01011 | | | 110 | 001 |
| 1100 | | | | | 0 | 110 |
| 00011 | | | | | | 0011 |
| 00110 | | | | | | 0110 |
| 11110 | | | | | | |
| 101011 | | | | | | |

- **Εφαρμογή ΙΙ:** Να ελεγχθεί ως προς την μονοσημαντικότητα ο παρακάτω κώδικας:

| | |
|---|---|
| $x_1$ | abc |
| $x_2$ | abcd |
| $x_3$ | e |
| $x_4$ | dba |
| $x_5$ | bace |
| $x_6$ | ceac |
| $x_7$ | ceab |
| $x_8$ | eabd |

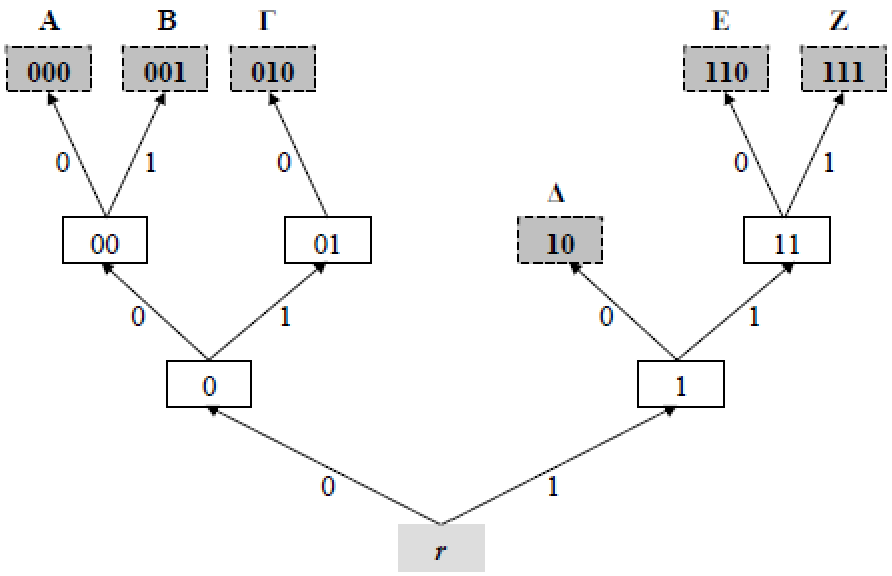
Απ:

| *S0* | *S1* | *S2* | *S3* | *S4* | *S5* | *S6* | *S7* | *S8* | *S9* | *S10* |
|---|---|---|---|---|---|---|---|---|---|---|
| abc | d | ba | ce | ac | c | eac | ac | c | eac | ac |
| abcd | abd | | | ab | cd | eab | ab | cd | eab | ab |
| e | | | | | | | d | ba | ce | d |
| dba | | | | | | | | | | |
| bace | | | | | | | | | | |
| ceac | | | | | | | | | | |
| ceab | | | | | | | | | | |
| eabd | | | | | | | | | | |

# Δενδροδιάγραμμα Απόφασης

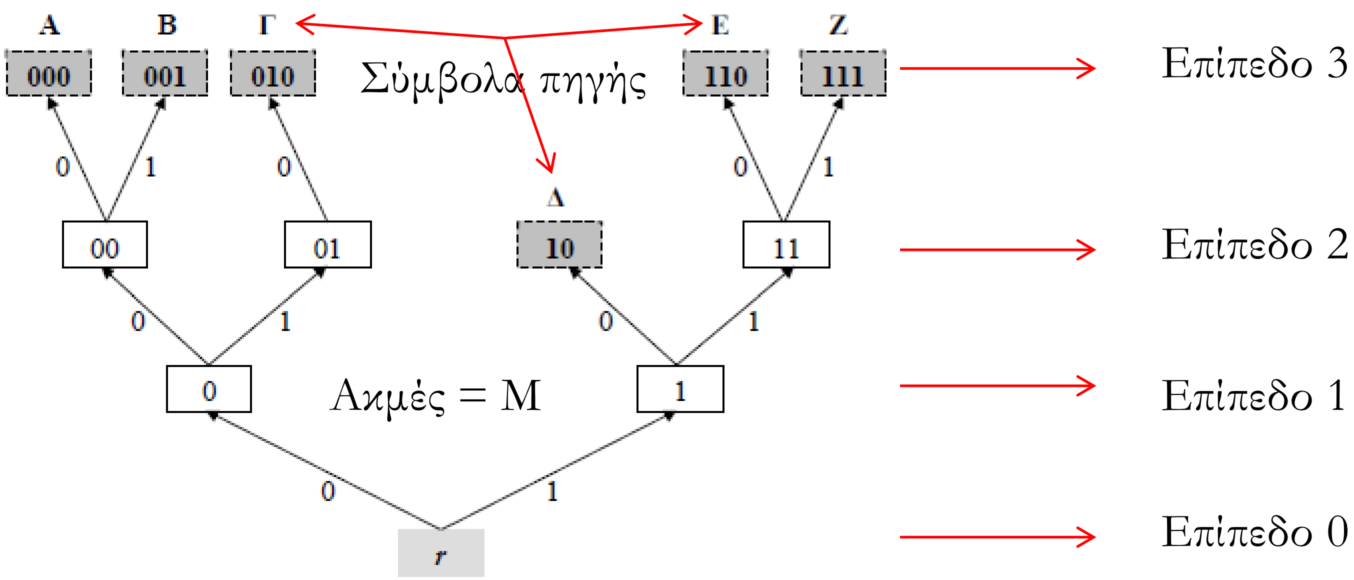
- Ένας στιγμιαία αποκωδικοποιήσιμος κώδικας μπορεί να αναπαρασταθεί με ένα **δενδροδιάγραμμα.**
- Δενδροδιάγραμμα:
  - γράφος με τα εξής στοιχεία:
    - ***ρίζα***
    - ***κόμβους ή φύλλα***
      - ένας κόμβος λέγεται φύλο όταν αναπαριστά ένα σύμβολο πηγής.
    - ***ακμές*** (κατευθυνόμενες).
- Παράδειγμα:

| Σύμβολο | Κωδική Λέξη |
|---|---|
| Α | 000 |
| Β | 001 |
| Γ | 010 |
| Δ | 10 |
| Ε | 110 |
| Ζ | 111 |



# Δενδροδιάγραμμα Απόφασης

- Χαρακτηριστικά δενδροδιαγράμματος:
  - επίπεδα
    - όσα και το μήκος της μεγαλύτερης κωδικής λέξης (μη συμπεριλαμβανομένου του επιπέδου της ρίζας) – *λέγεται και* ***ύψος του δένδρου***
  - φύλλα
    - όσα και τα σύμβολα της πηγής (μπορεί και να μην είναι στο υψηλότερο επίπεδο)
  - κατευθυνόμενες ακμές που εκκινούν από κάθε κόμβο
    - το πολύ όσες και τα κωδικά σύμβολα

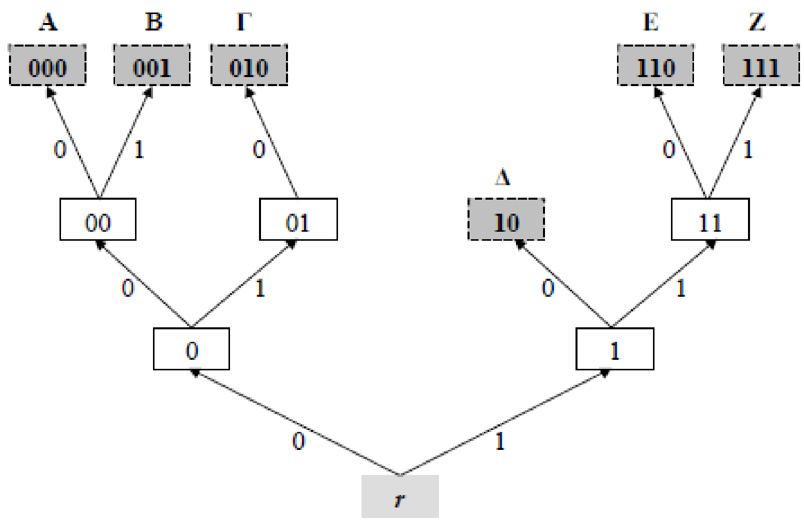


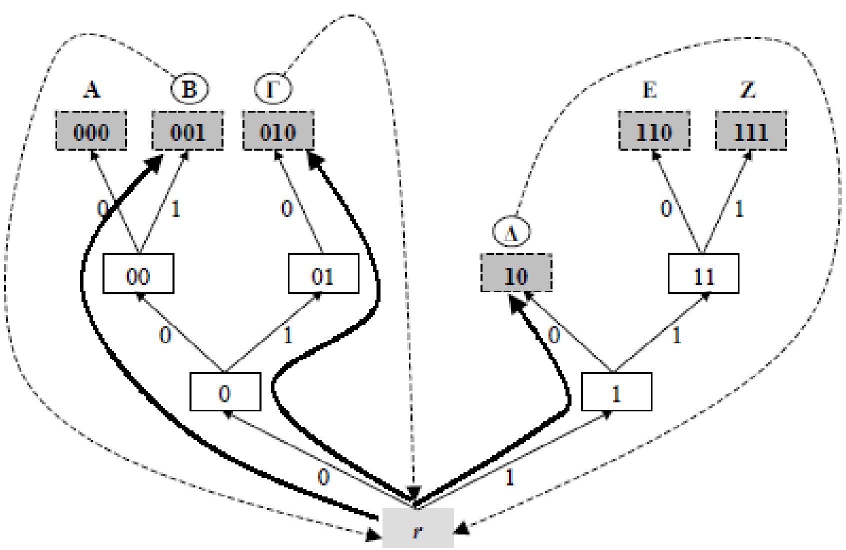
# Δενδροδιάγραμμα Απόφασης

- Το δενδροδιάγραμμα είναι σημαντικό εργαλείο για την γρήγορη αποκωδικοποίηση ενός κωδικού μηνύματος.
- Διαδικασία αποκωδικοποίησης ενός μηνύματος:
    1. Ξεκινάμε από την ρίζα
    2. Με την λήψη ενός κωδικού συμβόλου, μεταβαίνουμε στον κόμβο του αμέσως υψηλότερου επιπέδου μέσω της ακμής που αντιστοιχεί στο ληφθέν σύμβολο.
    3. Επαναλαμβάνουμε το 2ο βήμα μέχρι να φθάσουμε σε φύλλο και καταγράφουμε το σύμβολο της πηγής που αντιστοιχεί σε αυτό.
    4. Μεταβαίνουμε στην ρίζα του δένδρου και επαναλαμβάνουμε την διαδικασία

# Δενδροδιάγραμμα Απόφασης

- Παράδειγμα:
  - λαμβάνουμε την ακολουθία «00110010»




Ακολουθία Συμβόλων Πηγής:
**ΒΔΓ**

# Εφαρμογές

- **Εφαρμογή I:** Έστω ο στιγμιαία αποκωδικοποιήσιμος κώδικας {010,101,111,001} που κωδικοποιεί μία πηγή με αλφάβητο {A,B,C,D}.
  - Να φτιάξετε το δενδροδιάγραμμα.
  - Να αποκωδικοποιηθεί το μήνυμα 001010001010 με τη χρήση του δενδροδιαγράμματος. **Απ.:** DADA
- **Εφαρμογή II:** Έστω ο στιγμιαία αποκωδικοποιήσιμος κώδικας {a,c,d,bb,bad,eb,bcde} που κωδικοποιεί μία πηγή με αλφάβητο $\{x_1,x_2,x_3,x_4,x_5,x_6,x_7\}$.
  - Να φτιάξετε το δενδροδιάγραμμα.
  - Να αποκωδικοποιηθεί το μήνυμα cbbbcdebaddad με τη χρήση του δενδροδιαγράμματος. **Απ.:** $x_2x_4x_7x_5x_3x_1x_3$

# Οι Ταυτοανισότητες Kraft - McMillan

- Ταυτοανισότητα **McMillan**:
  - Τα μήκη των κωδικών λέξεων $l_1, l_2, \ldots, l_N$ ενός **μονοσήμαντου** Μ-αδικού κώδικα θα πρέπει να ικανοποιούν την ταυτοανισότητα:

$$\sum_{i=1}^{N} M^{-l_i} \leq 1$$

- Ταυτοανισότητα **Kraft**:
  - Υπάρχει **στιγμιαία αποκωδικοποιήσιμος** Μ-αδικός κώδικας με μήκη κωδικών λέξεων $l_1, l_2, \ldots, l_N$, αν και μόνο αν ισχύει η παρακάτω ταυτοανισότητα:

$$\sum_{i=1}^{N} M^{-l_i} \leq 1$$

- Και οι δύο ταυτοανισότητες αναφέρονται στην ύπαρξη ή όχι στιγμιαίου αποκωδικοποιήσιμου κώδικα (**Kraft**) ή μονοσήμαντου (**McMillan**) εφόσον γνωρίζουμε τα μήκη $l_i$ των κωδικών λέξεων:

# Παραδείγματα

- Έστω ο στιγμιαία αποκωδικοποιήσιμος κώδικας:

| | Κώδικας | Μήκη |
|---|---|---|
| A | 0 | 1 |
| B | 10 | 2 |
| C | 110 | 3 |
| D | 111 | 3 |

$$\sum_{i=1}^{N} M^{-l_i} = \frac{1}{2^1} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^3} = 1$$

- Έστω ο μονοσήμαντος κώδικας:

| | Κώδικας | Μήκη |
|---|---|---|
| A | 101 | 3 |
| B | 00 | 2 |
| C | 0001 | 4 |
| D | 1 | 1 |

$$\sum_{i=1}^{N} M^{-l_i} = \frac{1}{2^3} + \frac{1}{2^2} + \frac{1}{2^4} + \frac{1}{2^1} = \frac{15}{16} < 1$$

- Έστω ο κώδικας:

| | Κώδικας | Μήκη |
|---|---|---|
| A | 0 | 1 |
| B | 1 | 1 |
| C | 01 | 2 |
| D | 10 | 2 |

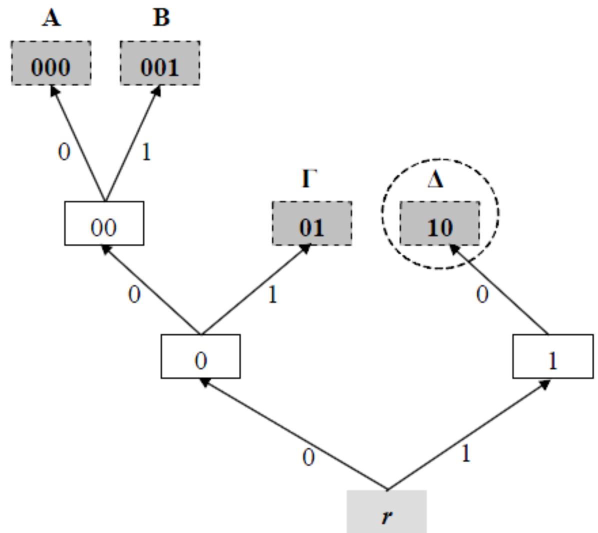
$$\sum_{i=1}^{N} M^{-l_i} = \frac{1}{2^1} + \frac{1}{2^1} + \frac{1}{2^2} + \frac{1}{2^2} = 1.5 > 1$$

# Οι Ταυτοανισότητες Kraft - McMillan

- **Παρατηρήσεις:**
    - Οι ταυτοανισότητες Kraft-McMillan **δεν λένε ότι:** ένας συγκεκριμένος Μ-αδικός κώδικας με συγκεκριμένα μήκη κωδικών λέξεων <u>είναι</u> στιγμιαία αποκωδικοποιήσιμος ή μονοσήμαντος.
    - Εάν ξέρουμε ότι ένας συγκεκριμένος Μ-αδικός κώδικας με δεδομένα μήκη κωδικών λέξεων είναι μονοσήμαντος τότε θα υπάρχει και στιγμιαία αποκωδικοποιήσιμος κώδικας με αυτά τα μήκη λέξεων.
    - Ένας μονοσήμαντος κώδικας δεν παρουσιάζει ενδιαφέρον διότι θα υπάρχει στιγμιαία αποκωδικοποιήσιμος κώδικας με τα ίδια μήκη λέξεων που θα αποκωδικοποιείται με μεγαλύτερη ευκολία.

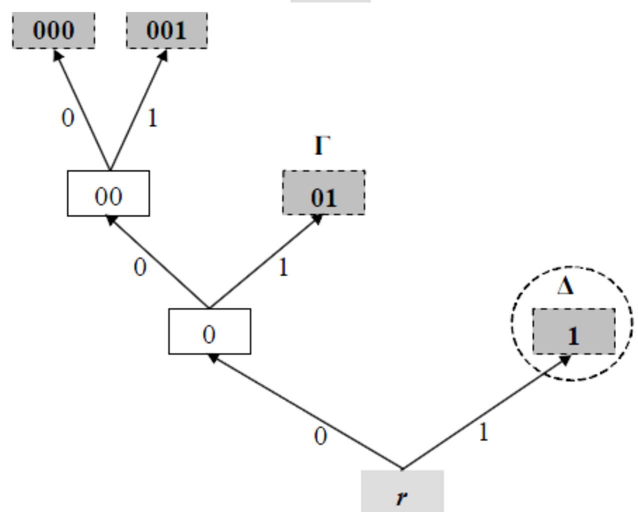
# Οι Ταυτοανισότητες Kraft - McMillan

- Εάν η ταυτοανισότητα Kraft ισχύει ως **ανισότητα** τότε ο κώδικας έχει **πλεόνασμα (redundancy)**
- Πρακτικά μπορούμε να προσθέσουμε και νέα κωδική λέξη.




$$\sum_{i=1}^{N} M^{-l_i} = \frac{3}{4} < 1$$

- Εάν η ταυτοανισότητα Kraft ισχύει ως **ισότητα** τότε ο κώδικας θεωρείται **ολόκληρος (complete)** – χωρίς πλεονασμό
- Πρακτικά δεν μπορούμε να προσθέσουμε και νέα κωδική λέξη.




$$\sum_{i=1}^{N} M^{-l_i} = 1$$

- Ποιόν από τους δύο κώδικες θα επιλέγατε; Θα μπορούσατε να προτείνετε έναν καλύτερο;

# Μέσο Μήκος Κώδικα

- Εάν θεωρήσουμε:
  - Πηγή πληροφορίας με **Αλφάβητο** $\boldsymbol{X=\{x_1,x_2,\ldots,x_N\}}$ και **Κατανομή** $\boldsymbol{P_X=\{p_1,p_2,\ldots,p_N\}}$
  - **Μ-αδικό Κώδικα με Αλφάβητο** $\boldsymbol{W=\{w_1,w_2,\ldots,w_M\}}$
  - Ότι κάθε σύμβολο της πηγής $\boldsymbol{x_i}$ αντιστοιχίζεται σε μία **κωδική λέξη** $\boldsymbol{u_i}$ με **μήκος** $\boldsymbol{l_i}$
  - Ορίζουμε **Μέσο Μήκος Κώδικα:**

$$\overline{L}=\sum_{i=1}^{N} p_i \cdot l_i$$

  - Εκφράζει το μέσο αριθμό των κωδικών συμβόλων ανά σύμβολο πηγής.
- Παράδειγμα
  - **ASCII-7**
    - μέσο μήκος: 7-bit
    - δεν υπολογίζονται οι πιθανότητες εμφάνισης των συμβόλων πηγής
    - σύμβολα με μικρή πιθανότητα εμφάνισης κωδικοποιούνται όπως και τα σύμβολα με μεγάλη πιθανότητα εμφάνισης
    - **Δεν είναι η βέλτιστη λύση στην «αθόρυβη κωδικοποίηση»**

# Βέλτιστος Κώδικας

- Είναι ο στιγμιαία αποκωδικοποιήσιμος κώδικας με το μικρότερο μέσο μήκος κώδικα:

$$\overline{L}^* = \min_{l_1, l_2, ..., l_N} \{\overline{L}\} = \min_{l_1, l_2, ..., l_N} \{\sum_{i=1}^{N} p_i \cdot l_i\}$$

  - Ο σχεδιασμός του βέλτιστου κώδικα απαιτεί τον προσδιορισμό των μηκών των κωδικών λέξεων που ελαχιστοποιούν το μέσο μήκος.

- Αποδεικνύεται ότι τα μήκη των κωδικών λέξεων του βέλτιστου στιγμιαία αποκωδικοποιήσιμου κώδικα είναι ίσα με:

$$l_i^* = -\log_M(p_i) \quad \gamma\iota\alpha \, 1 \le i \le N$$

- **Άρα, για δυαδικό κώδικα θα ισχύει:**

$$\overline{L}^* = H(X)$$

# Εφαρμογές

- **Εφαρμογή Ι:** Υπάρχει δυαδικός στιγμιαία αποκωδικοποιήσιμος κώδικας με κωδικές λέξεις μήκους {1,4,3,3,2,3}; **Απ.:** 19/16, όχι δεν υπάρχει.
- **Εφαρμογή ΙΙ:** Υπάρχει τριαδικός στιγμιαία αποκωδικοποιήσιμος κώδικας με κωδικές λέξεις μήκους {1,2,3,3,2,3}; **Απ.:** 18/27, υπάρχει.
- **Εφαρμογή ΙΙΙ:** Ποιοί από τους παρακάτω στιγμιαία αποκωδικοποιήσιμους κώδικες έχουν πλεονασμό και ποιοί όχι.

| **I** | **II** | **III** |
|---|---|---|
| 01 | 011 | 1 |
| 11 | 10 | 01 |
| 10 | 11 | 001 |
| 00 | 00 | 0001 |

**Απ.:** Οι ΙΙ και ΙΙΙ έχουν. Ο Ι δεν έχει.
Για τον ΙΙ: {010}
Για τον ΙΙΙ: {0000}

Αναφέρατε από μια κωδική λέξη που μπορούμε να προσθέσουμε στους κώδικες με πλεονασμό για να τους μετατρέψουμε σε ολόκληρους.

# Εφαρμογές

• **Εφαρμογή IV** (Εξεταστική 2007)**:** Παρακάτω δίνεται πίνακας με 4 κώδικες κωδικοποίησης τεσσάρων συμβόλων πηγής, καθώς και οι πιθανότητες εμφάνισης των συμβόλων.

a) Να ταξινομηθούν οι κώδικες.
b) Να υπολογιστεί το μέσο μήκος τους.
c) Εάν έπρεπε να επιλέξουμε έναν από τους παρακάτω κώδικες ποιος θα ήταν ο βέλτιστος και γιατί?

| $x_i$ | $p_i$ | I | II | III | IV |
|---|---|---|---|---|---|
| A | 0.15 | 01 | 011 | 0 | 010 |
| B | 0.25 | 11 | 10 | 01 | 101 |
| C | 0.35 | 10 | 11 | 011 | 111 |
| D | 0.25 | 01 | 00 | 0111 | 001 |

# Ασκήσεις Επανάληψης

• **Εφαρμογή Ι** (Εξεταστική 2010)**:** Δίνεται ο εξής κώδικας: Α:000, Β:001, Γ:01, Δ:10, Ε:11. Μπορούμε να προσθέσουμε ένα ακόμα σύμβολο Ζ (με την κατάλληλη κωδικοποίηση) έτσι ώστε ο νέος κώδικας που θα προκύψει να είναι στιγμιαία αποκωδικοποιήσιμος; **Απ:** Ναί

• **Εφαρμογή ΙΙ** (Εξεταστική 2009): α) Έστω ότι για έναν τριαδικό κώδικα γνωρίζουμε τα μήκη των κωδικών λέξεων που είναι τα εξής {1,2,3,1,2,3,1}. Ο κώδικας αυτός

i) Είναι σίγουρα στιγμιαία αποκωδικοποιήσιμος

ii) Σίγουρα δεν είναι στιγμιαία αποκωδικοποιήσιμος

iii) Μπορεί ναι, μπορεί και όχι (πρέπει να γνωρίζουμε τις κωδικές λέξεις για να αποφανθούμε)

**Απ:** α) το ii), β) το iii)

β) Επιλύστε το παραπάνω ερώτημα για έναν κώδικα με τα εξής μήκη {1,2,3,3,2,3}.

• **Εφαρμογή ΙΙΙ** (Εξεταστική 2009): Έστω Μ-αδικός ευκρινής κώδικας σταθερού μήκους (*k* κωδικών λέξεων με μήκος *l* η κάθε μία)

α) Τι πρέπει να προσέξουμε στην κατασκευή της νέας κωδικής λέξης έτσι ώστε ο νέος κώδικας να είναι στιγμιαία αποκωδικοποιήσιμος;

β) Σε ποιά περίπτωση δεν μπορεί να κατασκευαστεί στιγμιαία αποκωδικοποιήσιμος κώδικας με τον παραπάνω τρόπο;

# Ασκήσεις Επανάληψης

• **Εφαρμογή IV** (Εξεταστική 2009)**:** Έστω ότι δίνεται ένας κώδικας ο οποίος είναι μονοσήμαντος αλλά όχι στιγμιαία αποκωδικοποιήσιμος. Μας ζητείται να μετατρέψουμε τον κώδικα αυτό σε στιγμιαία αποκωδικοποιήσιμο με την παρακάτω διαδικασία:

- Σε μερικές (ή όλες) τις κωδικές λέξεις αλλάζουμε τα σύμβολα τους (όσα επιθυμούμε και όπως εμείς επιθυμούμε) χωρίς όμως να προσθέσουμε ή να αφαιρέσουμε σύμβολο. (π.χ. Εάν είναι δυαδικός κώδικας κάποια μηδενικά μπορώ να τα κάνω άσσους και το αντίστροφο)

Πότε μπορούμε με την παραπάνω διαδικασία να μετατρέψουμε έναν μονοσήμαντο αλλά μη στιγμιαία αποκωδικοποιήσιμο κώδικα σε στιγμιαία αποκωδικοποήσιμο:

- **Ποτέ** (κανένας μονοσήμαντος και μή στιγμιαία αποκωδικοποιήσιμος κώδικας δεν μπορεί να μετατραπεί σε στιγμιαία αποκωδικοποιήσιμο με την παραπάνω διαδικασία).
- **Μερικές φορές** (μπορεί να γίνει, αλλά για κάποιους μονοσήμαντους και μη στιγμαία αποκωδικοποιήσιμους κώδικες μόνο).
- **Πάντα** (μπορεί να γίνει για όλους τους μονοσήμαντους και μη στιγμιαία αποκωδικοποιήσιμους κώδικες).

**Απ:** Πάντα

# Τέλος Ενότητας