

Εισαγωγή στην Πληροφορική

Ενότητα 9 : Συμπύεση δεδομένων



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης





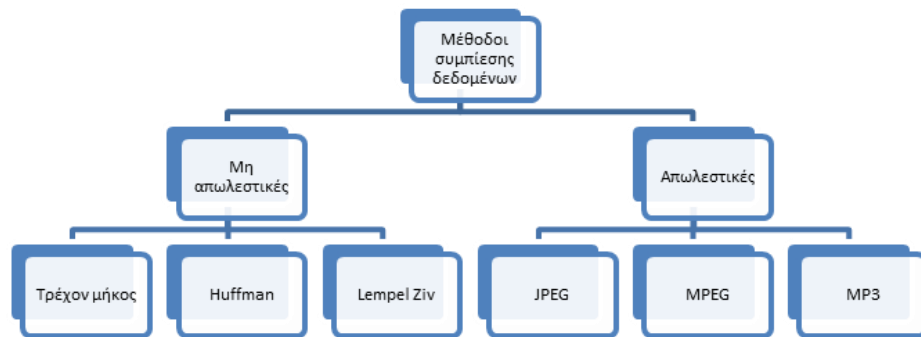
Άδειες Χρήσης

- 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.





Συμπίεση δεδομένων



- Συμπίεση δεδομένων είναι η αποθήκευση ή η μετάδοση της πληροφορίας με μικρότερο αριθμό bits σε σχέση με την αρχική μορφή της πληροφορίας. Αυτό επιτυγχάνεται με έξυπνες τεχνικές οι οποίες αντιστοιχούν μεγαλύτερες σε μήκος ακολουθίες bits σε μικρότερες



Μή απωλεστική συμπίεση

- Στην μη απωλεστική συμπίεση (lossless compression) διατηρείται η ακεραιότητα των δεδομένων. Αυτό σημαίνει ότι από τα συμπιεσμένα δεδομένα μπορούμε να ανακτήσουμε στην πλήρη μορφή της την αρχική πληροφορία.
- Η απωλεστική συμπίεση χρησιμοποιείται όταν είναι απαραίτητη να μην χαθεί ούτε ένα μπιτ δεδομένων κατά την συμπίεση.
- Τρεις διαδεδομένοι αλγόριθμοι συμπίεσης είναι:
 - Κωδικοποίηση τρέχοντος μήκους (RLE).
 - Κωδικοποίηση Huffman.
 - Αλγόριθμος Lempel Ziv.

Πολλές φορές τα αποθηκευμένα δεδομένα εμπεριέχουν **πλεονασμό**. Η συμπίεση περιορίζει τον πλεονασμό οργανώνοντας αποδοτικότερα την αποθήκευση.



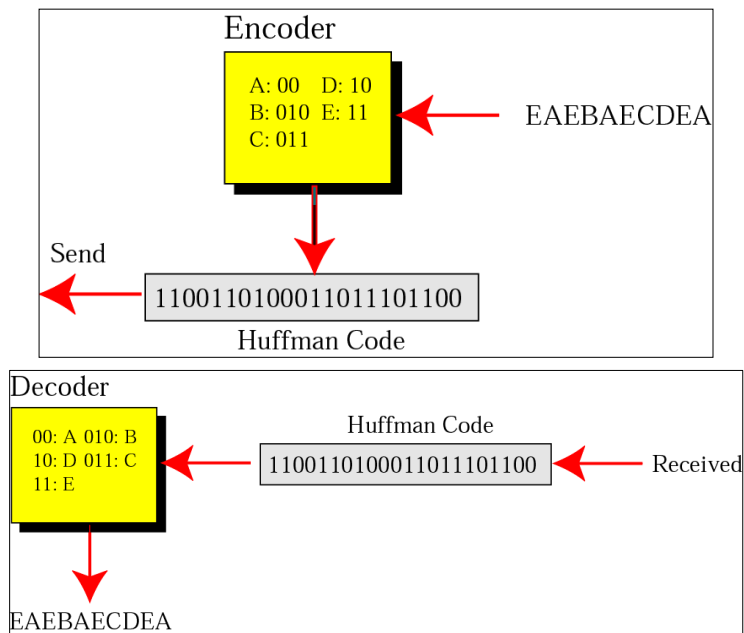
Κωδικοποίηση τρέχοντος μήκους (RLE)

- **Βασική αρχή:** Αντικατάσταση των συνεχόμενων επαναλαμβανόμενων παρουσιών ενός συμβόλου από μόνο μια παρουσία του συμβόλου που συνοδεύεται από τον αριθμό των παρουσιών
- π.χ. Η σειρά **ΑΑΑΑΑΒΒΒΒΑΑΔΔΔΔΔΔΔΔΔ** που περιέχει 20 σύμβολα μπορεί να γραφεί ως **5Α4Β2Α9Δ** που περιέχει 8 σύμβολα. Συνεπώς το μέγεθος της σειράς έχει μεταβληθεί κατά $(8-20)/20 = -60\%$ σε σχέση με το αρχικό μέγεθος.
- Η μέθοδος αυτή είναι αποδοτική στο δυαδικό σύστημα. Στην παραλλαγή αυτή αποθηκεύεται το πλήθος των μηδενικών bits ανάμεσα σε δύο άσσους.
- π.χ. Η σειρά 0000000001000010000000 που περιέχει 20 bits μπορεί να γραφεί ισοδύναμα ως [9][4][7]. Με κωδικοποίηση 4 bits για κάθε αριθμό το μήκος της συμπιεσμένης ακολουθίας θα είναι 12 bits και η προκύπτουσα ακολουθία θα είναι η: **1001 0100 0111**. Συνεπώς το μέγεθος της σειράς έχει μεταβληθεί κατά $(12-20)/20 = -40\%$ σε σχέση με το αρχικό μέγεθος. ⁵



Κωδικοποίηση Huffman

- Κατά την κωδικοποίηση Huffman αντιστοιχίζονται συντομότεροι κωδικοί σε σύμβολα που συναντώνται συχνότερα και μεγαλύτεροι σε αυτά που συναντώνται σπανιότερα
- Η κωδικοποίηση λαμβάνει μια σειρά χαρακτήρων και την μετατρέπει σε μια σειρά από δυαδικά ψηφία με βάση τον πίνακα κωδικοποίησης για κάθε χαρακτήρα.
- Η αποκωδικοποίηση με βάση τον πίνακα κωδικοποίησης είναι σε θέση να αντιστοιχίσει σε κάθε ακολουθία δυαδικών ψηφίων ένα χαρακτήρα.





Παράδειγμα κωδικοποίησης Huffman

- Έστω η παρακάτω σειρά 36 χαρακτήρων ASCII:
ADBACCAEEDDCCBABBBAACBACABDDDB
CCADADACAE

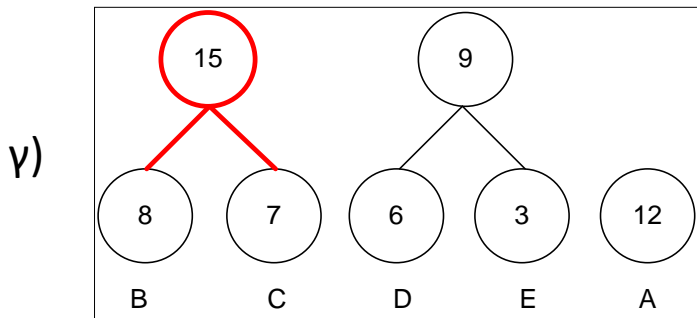
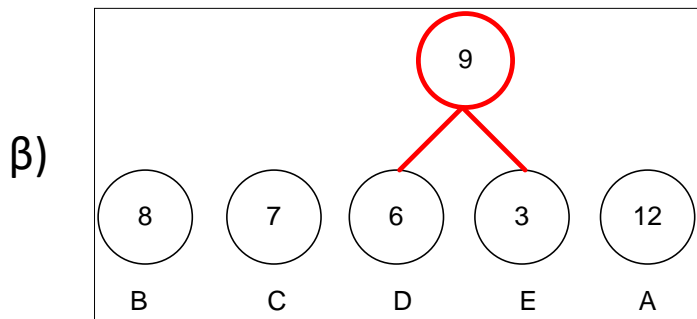
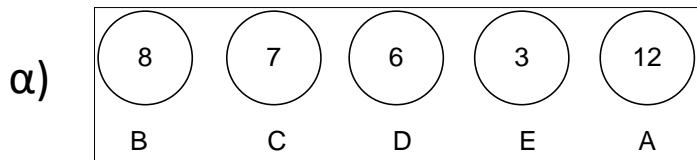
A	→	12
B	→	8
C	→	7
D	→	6
E	→	3

=====

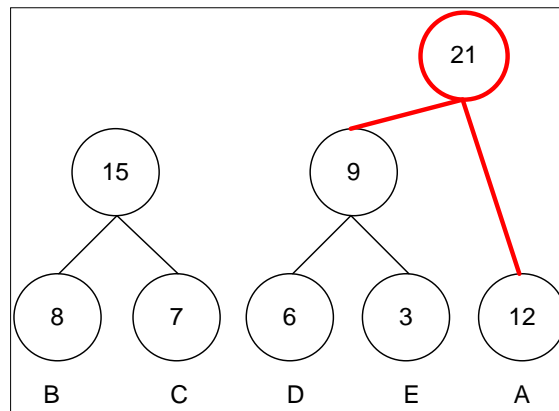
ΣΥΝΟΛΟ		36
--------	--	----
- Όλοι οι κόμβοι τοποθετούνται σε μια σειρά και ο καθένας έχει ως βάρος τον αριθμό των εμφανίσεών του
- Συνενώνονται οι 2 κόμβοι με τα μικρότερα βάρη και δημιουργείται προς τα επάνω ένας νέος κόμβος με βάρος ίσο με το άθροισμα των κόμβων που συνενώθηκαν
- Καθώς απαιτούνται 8 bits για κάθε χαρακτήρα στην κωδικοποίηση ASCII **36*8=288 bits** είναι η απαιτούμενη χωρητικότητα για την ακολουθία στο σύνολό της



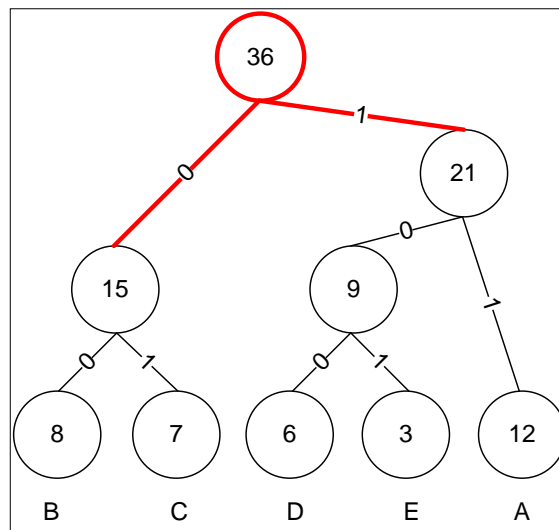
Παράδειγμα κωδικοποίησης Huffman



δ)



ε)



A → 11

B → 00

C → 01

D → 100

E → 101

Στην κωδικοποίηση
τώρα απαιτούνται

$$12 \cdot 2 + 8 \cdot 2 + 7 \cdot 2 + 6 \cdot 3 + 3 \cdot 3 = 24$$

$$+ 16 + 14 + 18 + 9 = 81$$

Άρα επιτεύχθηκε
συμπίεση στο

$$(81 - 288) / 288 =$$

-71,875% του αρχικού

(χωρίς να ληφθεί
υπόψη η μετάδοση
του πίνακα
κωδικοποίησης)



Απωλεστικές μέθοδοι συμπίεσης

- Απωλεστικές μέθοδοι συμπίεσης (lossy data compression): Κατά την συμπίεση χάνονται κάποια δεδομένα.
- Οι απωλεστικές μέθοδοι είναι οικονομικότερες και απαιτούν λιγότερο χρόνο και χώρο.
- Μορφές δεδομένων στις οποίες είναι αποδεκτή η απώλεια δεδομένων κατά την συμπίεση: Ήχος, εικόνα, βίντεο.
- Διαδεδομένες απωλεστικές τεχνικές συμπίεσης είναι η μέθοδος JPEG για εικόνες και η μέθοδος MPEG για βίντεο.



Συμπίεση εικόνας JPEG

- Η αποθήκευση εικόνων με την μορφοποίηση JPEG είναι ιδιαίτερα δημοφιλής καθώς επιτυγχάνει την συμπίεση του τελικού αρχείου διατηρώντας παράλληλα την ποιότητα της εικόνας σε ικανοποιητικά επίπεδα.
- Κατά την κωδικοποίηση η εικόνα χωρίζεται σε τμήματα 8X8 pixels (pixel=εικονοστοιχείο). Σε κάθε τμήμα οι διαφορές στα χρώματα εξομαλύνονται έτσι ώστε να καταλαμβάνει μικρότερη χωρητικότητα.
- Μπορεί να αναπαραστήσει μέχρι 2^{24} χρώματα.



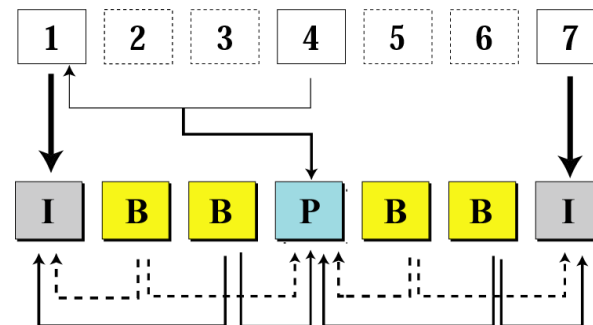
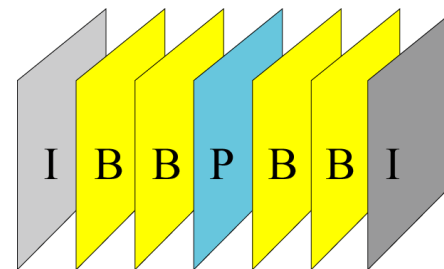
Συμπίεση βίντεο

- Για 72 λεπτά βίντεο σε ένα CD είτε θα πρέπει να μειωθεί η ποιότητα είτε να συμπιεστούν τα περιεχόμενα με τέτοιο τρόπο που να μπορούν να αναπαραχθούν σε πραγματικό χρόνο. Ο συντελεστής συμπίεσης θα πρέπει να είναι της τάξης του 200:1
- Η βασική ιδέα της συμπίεσης με την μέθοδο MPEG είναι η περιγραφή ορισμένων καρτέ με βάση κάποια άλλα και η περιγραφή μόνο των διαφορών τους



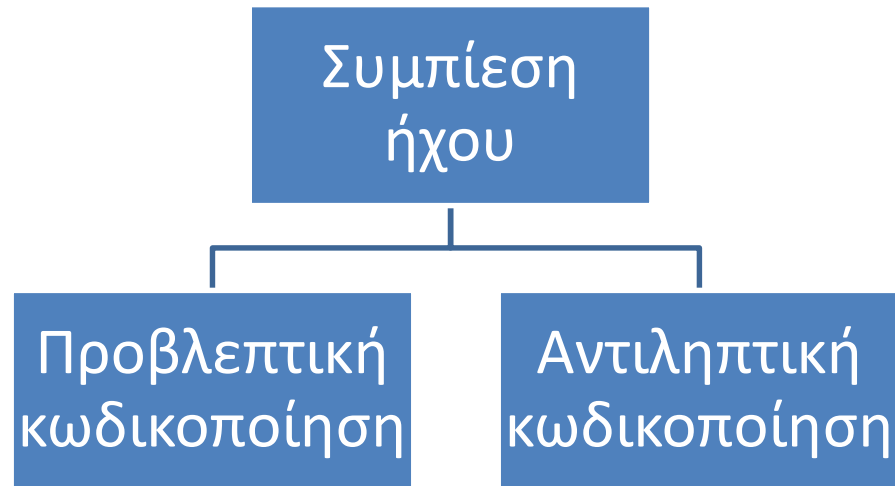
MPEG = Moving Pictures Experts Group

- **Χωρική συμπίεση.** Για κάθε καρέ η συμπίεση γίνεται με την μέθοδο JPEG ή με κάποια παραλλαγή της.
- **Χρονική συμπίεση.** Κατηγοριοποίηση καρέ σε:
 - Ενδοκωδικοποιημένα καρέ (intracoded frames=i-frames). Είναι ανεξάρτητα καρέ που δεν σχετίζονται με κανένα άλλο καρέ. Εμφανίζονται σε τακτά χρονικά διαστήματα (π.χ. κάθε ένατο καρέ)Τα i-Frames είναι ανεξάρτητα από άλλα καρέ και δεν μπορούν να παρασκευαστούν από αυτά.
 - Προβλεπόμενα καρέ (predicted frames=p-frames). Κάθε προβλεπόμενο καρέ περιέχει μόνο τις αλλαγές από το προηγούμενο i-frame ή p-frame.
 - Αμφίδρομα καρέ (bidirectional frames=b-frames). Σχετίζονται με το προηγούμενο και με το επόμενο i-frame ή p-frame.
- Εκδόσεις της συμπίεσης MPEG
 - MPEG 1 → CD-ROM με ταχύτητα μεταφοράς 1,5Mbps
 - MPEG-2 → DVD-ROM με ταχύτητα μεταφοράς 3-6Mbps
 - MPEG-3 ή MP3 → στάνταρντ για συμπίεση ήχου.
- Συνήθως υπάρχουν 2 με 3 b-frames μεταξύ των i-frames ή p-frames και 3 με 5 p-frames μεταξύ των i-frames.





Συμπίεση ήχου





Προβλεπτική κωδικοποίηση

- Predictive encoding: Κωδικοποιούνται οι διαφορές μεταξύ των δειγμάτων και όχι όλες οι τιμές δειγματοληψίας
- Πρότυπα: GSM, G.729 και άλλα



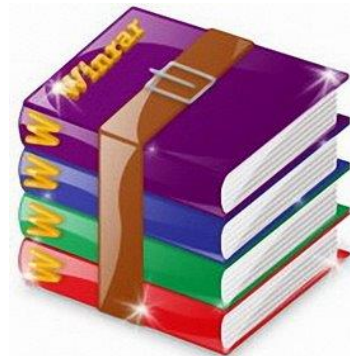
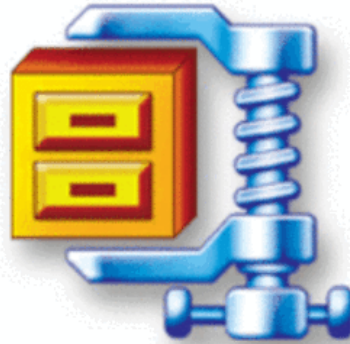
Αντιληπτική κωδικοποίηση MP3

- Perceptual encoding
- Βασίζεται στην μελέτη του τρόπου με τον οποίο αντιλαμβάνεται τον ήχο ο άνθρωπος
- **Κάλυψη συχνοτήτων:** Ένας ήχος μπορεί να καλύπτει πλήρως ή μερικώς ένα χαμηλότερο ήχο σε άλλο εύρος συχνοτήτων
- **Χρονική κάλυψη:** Η ευασθησία της ακοής μπορεί να μειωθεί προσωρινά από ένα δυνατό ήχο ακόμα και όταν αυτός πάψει να μεταδίδεται
- Ρυθμοί μετάδοσης στο MP3: 96kbps, 128kbps, 160kbps, 320kbps



Διαδεδομένα μορφότυπα συμπίεσης

- Zip
- 7z
- Rar





Βιβλιογραφία

1. Forouzan B., Mosharaf F. Εισαγωγή στην επιστήμη των υπολογιστών. Εκδόσεις Κλειδάριθμος (2010)
2. Καρολίδης Δ., Ξαρχάκος Κ.. Εισαγωγή στην πληροφορική και στο διαδίκτυο. Εκδόσεις Άβακας (2008).
3. Σφακιανάκης Μ. Εισαγωγή στην πληροφορική σκέψη. Εκδόσεις Κλειδάριθμος (2003).
4. Τσιτμηδέλης Σ., Τικτοπούλου Ε. Εισαγωγή στην πληροφορική. Πανεπιστημιακές εκδόσεις Αράκυνθος (2009).
5. Γιαγλής Γ. Εισαγωγή στην πληροφορική. Γκιούρδας εκδοτική (2009).
6. Αβούρης Ν., Κουφοπαύλου Ο., Σερπάνος Δ. Εισαγωγή στους υπολογιστές. Εκδόσεις tyrograma (2004).
7. Biermann A. Σπουδαίες ιδέες στην επιστήμη των υπολογιστών. Πανεπιστημιακές εκδόσεις Κρήτης (2008).
8. Brookshear J.G. Η επιστήμη των υπολογιστών, μια ολοκληρωμένη παρουσίαση. Εκδόσεις Κλειδάριθμος (2009).
9. Ceruzzi P.E. Ιστορία της υπολογιστικής τεχνολογίας. Από τον ENIAC μέχρι το διαδίκτυο. Εκδόσεις Κάτοπτρο (2006).



Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά Δημιουργού-Μη Εμπορική Χρήση-Όχι Παράγωγα Έργα 4.0 Διεθνές [1] ή μεταγενέστερη. Εξαιρούνται τα αυτοτελή έργα τρίτων π.χ. φωτογραφίες, Διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



Ο δικαιούχος μπορεί να παρέχει στον αδειοδόχο ξεχωριστή άδεια να χρησιμοποιεί το έργο για εμπορική χρήση, εφόσον αυτό του ζητηθεί.

[1] <http://creativecommons.org/licenses/by-nc-nd/4.0/deed.el>



Τέλος Ενότητας

Επεξεργασία: Α. Αναγνωστάκης



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ



ΕΣΠΑ
2007-2013
Πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΙΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



Τέλος Ενότητας

Συμπίεση δεδομένων



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

