



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΛΟΠΟΝΝΗΣΟΥ,
ΤΜΗΜΑ ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΒΕΣ 04: ΣΥΜΠΙΕΣΗ ΚΑΙ ΜΕΤΑΔΟΣΗ ΠΟΛΥΜΕΣΩΝ

Ακαδημαϊκό Έτος 2005 – 2006, Χειμερινό Εξάμηνο

Θεωρητικές Ασκήσεις (# 2): Συμπίεση Δεδομένων (Κειμένου)

Η συγκεκριμένη εργασία δεν έχει κάποια συγκεκριμένη λύση. Ο αλγόριθμος συμπίεσης που θα αναπτυχθεί πρέπει να λαμβάνει υπόψη τρία βασικά χαρακτηριστικά του συγκεκριμένου κειμένου:

1. Το κείμενο περιέχει μεγάλες ακολουθίες κενών. Αυτό σημαίνει ότι οποιαδήποτε τεχνική αντικατάστασης προτύπων (όπως ο κλασικός αλγόριθμος LZW) είναι μια κατάλληλη λύση για συμπίεση. Εναλλακτικά μπορεί να χρησιμοποιηθούν τεχνικές κωδικοποίησης μήκους διαδρομής ειδικά για το κενό (όλοι οι υπόλοιποι χαρακτήρες μπορεί να κωδικοποιηθούν ως έχουν).
2. Το συγκεκριμένο κείμενο περιέχει περιορισμένο αριθμό χαρακτήρων σε σχέση με τον κώδικα ASCII (π.χ. περιέχει ελάχιστα κεφαλαία γράμματα, σημεία στίξης και αριθμητικά σύμβολα). Επομένως ακόμη και με κωδικοποίηση σταθερού μήκους κωδικής λέξης (< 8 bits) επιτυγχάνεται συμπίεση,
3. Ορισμένοι χαρακτήρες (π.χ. ο χαρακτήρας 'ε') εμφανίζονται πολύ πιο συχνά σε σχέση με άλλους (π.χ. τα κεφαλαία γράμματα, ο χαρακτήρας 'ρ', κλπ.). Η παρατήρηση αυτή υποδηλώνει ότι οποιαδήποτε μορφή στατιστικής κωδικοποίησης (Huffman, Αριθμητική) μπορεί να χρησιμοποιηθεί με σημαντικά αποτελέσματα ως προς τη συμπίεση.