

Μια απλοποιημένη έκδοση του αλγορίθμου Huffman αποτελείται από τα εξής στάδια:

- Βήμα 1** Μετράμε τη συχνότητα του κάθε χαρακτήρα στην ακολουθία
- Βήμα 2:** Ταξινομούμε τις συχνότητες εμφάνισης σε μια λίστα κατά φθίνουσα τάξη
- Βήμα 3:** Κατασκευάζουμε ένα «δένδρο» για την κωδικοποίηση ξεκινώντας με τους συχνότερα εμφανιζόμενους χαρακτήρες
- Βήμα 4:** Αντιστοιχίζουμε τα δυαδικά '0' και '1' σε κάθε κόμβο του δέντρου: Αρχίζοντας από την ρίζα του δέντρου, προσθέτουμε '0' για κάθε αριστερό παιδί και '1' για κάθε δεξί. Οι χαρακτήρες που κωδικοποιούνται είναι τα φύλλα στην βάση του δέντρου. Αρχίζοντας από την κορυφή (ρίζα) του δέντρου, διατρέχοντας το μοναδικό μονοπάτι προς κάθε φύλλο, συλλέγουμε 0 ή 1 και ορίζουμε τον κώδικα για το χαρακτήρα που αντιστοιχεί στο φύλλο αυτό.

Έστω ότι θέλουμε να κωδικοποιήσουμε τη λέξη «ΑΛΛΟΣ».

Βήμα 1,2: Δημιουργούμε μια λίστα με τους χαρακτήρες ταξινομώντας τους με σειρά εμφάνισης. Έτσι έχουμε τη λίστα {«Λ», «Α», «Ο», «Σ»} αφού το «Λ» εμφανίζεται δύο φορές και οι υπόλοιποι χαρακτήρες από μία (τους οποίους και τοποθετούμε στη λίστα με τη σειρά εμφάνισης τους).

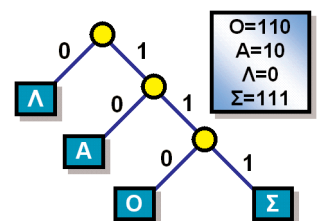
Βήμα 3: Ξεκινώντας από τη ρίζα, δημιουργούμε δύο παιδιά: το «Λ» (το γράμμα που εμφανίζεται πρώτο στη λίστα) και έναν εσωτερικό κόμβο. Στη συνέχεια από τον κόμβο αυτό δημιουργούμε πάλι δύο παιδιά, ένα με τον επόμενο χαρακτήρα στη λίστα, δηλαδή το «Α» και έναν εσωτερικό κόμβο. Διαγράφουμε το «Λ» από τη λίστα. Επαναλαμβάνουμε την διαδικασία μέχρι να τελειώσουν όλοι οι χαρακτήρες της λίστας.

Τελικά όλοι οι χαρακτήρες «Λ», «Α», «Ο», και «Σ» αποθηκεύονται σε φύλλα του δέντρου.

Βήμα 4: Αντιστοιχίζουμε τα δυαδικά 0 και 1, διατρέχοντας το μοναδικό μονοπάτι προς κάθε φύλλο-χαρακτήρα.

Τελικά, η ακολουθία «ΑΛΛΟΣ» που είχε αρχικό μήκος $5 \times 8 = 40$ bits, συμπιεσμένη γίνεται: **1000110111**, συνολικού μήκους 11 bits, άρα περίπου 4 φορές μικρότερη. Βέβαια, στην πράξη, στο τέλος κάθε ακολουθίας προστίθεται και το σχήμα της κωδικοποίησης, ώστε να μπορεί να γίνει η αποσυμπίεση.

Η ακολουθία κωδικοποίησης που παράγεται με τον προηγούμενο αλγόριθμο είναι μοναδική, ώστε να μπορεί να γίνεται η αποσυμπίεση. Έτσι, αν διαθέτουμε το σχήμα της κωδικοποίησης (πίνακας στο πλάι) και την κωδικοποιημένη ακολουθία, τότε μπορούμε να κάνουμε εύκολα την αποσυμπίεση της κωδικοποιημένης ακολουθίας **1000110111**. Διαβάζεται από αριστερά προς τα δεξιά το πρώτο 1. Δεν υπάρχει χαρακτήρας που να αντιστοιχεί σε αυτό. Διαβάζεται και το 0. Ο χαρακτήρας Α αντιστοιχεί στο 10, άρα αντικαθίσταται το 10 με αυτόν. Στη συνέχεια διαβάζουμε το 0, και ο μοναδικός χαρακτήρας που αρχίζει με 0 είναι ο Λ. Όμοια και για τα υπόλοιπα γράμματα.



Συμπίεση LZW

Πολλά αρχεία, ιδιαίτερα αρχεία χαρακτήρων, έχουν συμβολοσειρές που επαναλαμβάνονται συχνά, για παράδειγμα το άρθρο «τον». Για την