Machakos University

ISO 9001:2008 Certified

SCS 474 Compression Techniques Assignment

QUESTION ONE

(a) Define each of the following as used in compression
   i. Prefix code
   ii. Entropy
   iii. Code efficiency

(b) Given the following string as input
COUSCOUSOU
with the initial dictionary below, encode the string with the LZW algorithm, showing the intermediate steps:

INDEX	ENTRY
1	C
2	O
3	S
4	U

(c) A certain source emits symbols $\{A, B, C, D, F\}$ with probabilities $P(C) = \frac{10}{61}$, $P(D) = \frac{8}{61}$ $P(E) = \frac{9}{61}$ $P(F) = \frac{3}{61}$ $P(A) = \frac{25}{61}$, $P(B) = \frac{6}{61}$.

   i. Estimate the entropy for this source
   ii. Derive the Shanon-Fano code for each symbol
   iii. Derive the Huffman code for each symbol
   iv. Estimate the average code word length for each of the codes in (i) and (ii)

QUESTION TWO

(a) Distinguish between lossless and lossy data compression
(b) Encode the following stream of characters using decimal arithmetic encoding technique
MEDIA
You may assume that characters occur with probabilities of M = 0.1, E = 0.3, D = 0.3, I = 0.2 and A = 0.1.
(c) Show how your solution to (b) would be decoded