Reti di calcolatori

Homework 1

Usando il linguaggio di programmazione MATLAB oppure C, si creino gli script:

UTF8_encoding

UTF8_decoding

che eseguano rispettivamente la codifica e la decodifica secondo lo standard UTF-8.

Lo script UTF8_encoding legge dati da un file "input.data" in parole da 4 byte. Ogni parola costituisce un codepoint dello Universal Character Set UCS-4. In base al valore assunto dalla parola, lo script genera un codice UTF-8, secondo lo schema allegato (si veda la lezione L02-Digitalization.pdf), esteso per codificare codepoints che richiedono fino a 32 bit. Le codeword così generate vanno salvate nel file "UTF8.data".

Lo script UTF8_decoding compie l'operazione inversa, leggendo il file "UTF8.data" e riconvertendo le codeword in parole di 4 byte corrispondenti ai codepoint originari. L'output va salvato sul file "output.data" che, ovviamente, deve risultare identico a "input/data".