



Progetto “Codifica di Huffman” – Parte II

20 Maggio 2014

1. Codici di Huffman dei caratteri

Scrivi un metodo statico in Java che, dati i nomi (`String`) di due file di testo, uno di input e l'altro di output, analizza il file di input per determinarne l'istogramma delle frequenze dei caratteri contenuti, e crea il file di output che riporta la tabella dei codici di Huffman associati ai caratteri e delle relative frequenze. I caratteri ammessi sono quelli con codici ASCII standard nell'intervallo 0–127. Il file generato deve prevedere una riga per ciascun carattere effettivamente utilizzato, riga nella quale saranno riportati il simbolo del carattere, il corrispondente codice di Huffman e il numero di occorrenze nel file di input. Nel caso fosse utile farvi riferimento, i caratteri speciali *nuova-linea*, *capo-linea* e *tabulazione* possono essere rappresentati in Java dai letterali `'\n'`, `'\r'`, `'\t'`.

Per produrre l'istogramma, l'albero e la tabella dei codici di Huffman puoi applicare i metodi statici sviluppati a lezione. Le classi del package `huffman_toolkit` semplificano l'accesso e le operazioni sui file di testo (codici ASCII limitati all'intervallo 0–127); la relativa documentazione è disponibile attraverso le pagine del corso.

2. Analisi di file di testo

Utilizza i programmi realizzati in relazione alla parte I e al punto precedente per generare la tabella dei codici di Huffman dei caratteri contenuti in un sorgente Java e in un file di testo di analoghe dimensioni (in termini di numero complessivo di caratteri) generato casualmente. Confronta, quindi, le caratteristiche dei codici di Huffman risultanti.

3. Dimensioni dei file “compressi”

In base ai dati riportati nella tabella, in particolare codice di Huffman e numero di occorrenze di ciascun carattere, determina la dimensione in *byte* del testo compresso. Tieni conto che, ai fini dell'esempio discusso a lezione, un byte si compone di 7 bit utili, anziché 8, per via di alcune semplificazioni. Calcola, inoltre, il numero di caratteri che compongono l'intestazione e quindi la dimensione complessiva del file “compressso”. Esegui questi calcoli, che possono essere portati a termine agevolmente tramite uno spreadsheet, per entrambi i file di testo analizzati nel punto precedente.

4. Verifica

Applica, infine, il programma di compressione sviluppato a lezione e verifica i risultati ottenuti sopra confrontandoli con le dimensioni effettive dei file “compressi” prodotti.