



Progetto “Codifica di Huffman” – Parte I

13 Maggio 2014

1. Tabella delle frequenze dei caratteri in un file di testo

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 di tali frequenze. I caratteri ammessi sono quelli con codici ASCII standard nell'intervallo 0–127. Il file generato deve prevedere una riga di testo per ciascun carattere, riga nella quale saranno riportati il codice ASCII del carattere, il simbolo corrispondente 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 generare l'istogramma puoi applicare il metodo statico sviluppato in classe. Per l'accesso e le operazioni sui file di testo (codici ASCII limitati all'intervallo 0–127) puoi avvalerti delle classi del package `huffman_toolkit` la cui documentazione è disponibile attraverso le pagine del corso.

2. Generazione di un file di testo

Scrivi un metodo statico in Java che, dato il nome (`String`), genera un file di testo composto da caratteri i cui codici ASCII sono prodotti in modo casuale, con distribuzione uniforme, nell'intervallo 0–127.

Per l'accesso e le operazioni sui file di testo (codici ASCII limitati all'intervallo 0–127) puoi avvalerti delle classi del package `huffman_toolkit` la cui documentazione è disponibile attraverso le pagine del corso.

3. Analisi di file di testo

Utilizza i programmi realizzati in relazione ai due punti precedenti per generare la tabella delle frequenze dei caratteri 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 delle tabelle ottenute.