

課題 データ圧縮 (ランレングス符号化)

ファイルの容量を圧縮するアルゴリズムとして、ランレングス符号化 (run-length encoding) に基づくデータ圧縮と解凍プログラムを作成し、次ページのデータをファイルに保存し、圧縮前と圧縮後でのファイルサイズを比較せよ。また、圧縮されたデータから元のデータを復元できることを確認せよ。データはアルファベットのみのテキストデータとする。

ランレングス符号化

例えば、文字列「aaaaabbbbbbbcccddeeee」が与えられたとき、同じ文字が続く塊「aaaaa」、「bbbbbbb」、「ccc」、「dd」、「eeee」を「連」という。ランレングス符号化では各連を連の長さとその文字で表す。例えば「aaaaa」は「5a」のように表す。こうすることで元の文字数5を文字数2に圧縮することができる。ただし、連の長さが1または2のときは圧縮の効果が得られないのでそのまま表示するものとする。したがって「aaaaabbbbbbbcccddeeee」は「5a7b3cdd4e」と圧縮される。こうした符号化は画像データで同じ色が続くような場合には特に有効である。

(実行例)

2 ページに元データ, 3 ページに圧縮されたデータを示す。
この場合圧縮によってファイルサイズは $2388/8190 = 29.2\%$ に縮小されている。

