

文字列の圧縮・置換

文字列の圧縮

- ① 同一の文字が3文字以上連続する場合、
次の3文字の様式に置き換える。

圧縮マーク＋連続する同じ文字の個数
＋連続する同じ文字の内容

- ② 圧縮後の文字列を同じ配列上に設定する場合と
異なる配列上に設定する場合がある。

- ③ 同じ配列上に設定する場合は、
圧縮後の文字列の詰め合わせを
実行しながら圧縮処理を行う必要がある。

④ 具体例

圧縮前の文字列 AAAAABCCCDDEEEE 15文字

圧縮後の文字列 *5AB*3CDD*4E 12文字

*は圧縮記号を示す。

文字列の圧縮の手順

- ① 文字列の先頭から隣り合う2つの配列の要素を比較して、同一文字が連続する間、連続する同じ文字の個数を求める。
- ② 同じ文字の個数と文字の内容を認識する。
- ③ 配列の隣り合う文字列が異なる場合に、文字の転送または文字列の圧縮を行う。
- ④ 隣り合わせの文字が異なると、転送と圧縮の処置を行う。

- ⑤ 連続する同じ文字数が
2以下の場合は圧縮せずにそのまま転送する。
- ⑥ 個数分に相当する文字を転送する。
- ⑦ 連続する同じ文字数が3以上の場合、
圧縮記号＋同じ文字数
＋同じ文字の内容の3文字で表す。
- ⑧ ①で認識した文字の個数、
文字の内容の情報を使用する。
- ⑨ 文字列の最後になると、終端記号を付加して終了する。

同じ配列を使用した圧縮

- ① 2つのポインタ変数を使用して、圧縮する。
- ② 添字Iは圧縮前の文字の位置を表す。
- ③ 添字Jは同じ配列上の圧縮後の文字の位置を表す。

文字列の置換

① 置換

- ① テキスト文字列において
パターン文字列に一致する文字列を探索する。
- ② 一致文字列を置換文字列に置き換える処理である。

② 文字列の探索

- ① テキスト文字列上の部分文字列と
パターン文字列の対応する要素を比較する。
- ② パターン文字列の大きさと
同じ連続個数一致すれば探索成功となる。

- ③ 処理結果を同じ配列上に設定する場合は、
パターン文字列の大きさと
置換文字列の大小関係によって
テキスト文字列の挿入・削除の処理が必要になる。
- ④ パターン文字列の大きさが置換文字列の大きさよりも
大きい場合は詰め合わせが必要となり、
削除の考え方を利用する。
- ⑤ 置換文字列の大きさがパターン文字列の大きさよりも
大きい場合は空白が必要となり、
挿入の考え方を利用する。

文字列置換に必要な機能

- ① 置換対象文字列(長さL)を検索する機能
- ② 配列の詰め合わせ機能($L < M$ の場合)
- ③ 配列に空白を挿入する機能($L > M$ の場合)
- ④ 置換文字列(長さM)を処理する機能
- ⑤ 配列文字長を修正する機能