

Problem F: UTF-8

問題作成: Masashi Tsubosaka
解法作成: Masashi Tsubosaka, Yusuke
Konishi
解説: Yuta Kitamura

問題

- ところどころ欠けたUTF-8のビット列が与えられます
 - x の部分には 0 か 1 が入ります
- UTF-8として正しい文字列は何通りできるか？
 - xxxxxxxx → 1バイト文字は 0 で始まるので 128通り

UTF-8 の規則

バイト長	ビットパターン
1	0xxxxxxx
2	110yyyyx 10xxxxxx
3	1110yyyy 10yxxxxx 10xxxxxx
4	11110yyy 10yyxxxx 10xxxxxx 10xxxxxx

- x/y の部分には任意のビットが入る
- ただし y のうちどれかは必ず 1

解法

- 解法: **DP**
- $E[M]$ = 与えられた文字列の N バイト目でちょうど終了するような正しいエンコーディングの数
- $E[M] =$
 - $E[N - 1] * (N \text{ バイト目が1バイト文字になる場合の数})$
 - $+ E[N - 2] * (N - 1 \text{ バイト目から2バイト文字が始まる場合の数})$
 - $+ E[N - 3] * (N - 2 \text{ バイト目から3バイト文字が……})$
 - $+ E[N - 4] * (N - 3 \text{ バイト目から4バイト文字が……})$

注意点

- **y**さんのせいで若干めんどくさいです
 - 場合の数を丁寧に計算しておきましょう
- **オーバーフローに注意**
 - 1000000^2 は32bitからはみ出ます
 - こまめに剰余をとりましょう

結果

- 総提出数： 58
- 提出者数： 34
- 正解者数： 32
- 最初の正解者： Kazuhiro Hosakaさん (73分)