

HÀM PHẠT (Full)

Tên chương trình: GPSF.PAS

Xét 2 chuỗi X và Y trong đó các ký tự của mỗi chuỗi thuộc tập $\Sigma = \{A, G, C, T\}$. Các chuỗi X và Y sẽ được cân bằng bằng cách biến đổi thành các chuỗi có cùng độ dài: $X^* = x_1x_2\dots x_n$ và $Y^* = y_1y_2\dots y_n$ bằng việc chèn vào X hoặc Y một số ký tự '-' sao cho x_i và y_i không cùng là ký tự '-' ($i = 1, 2, \dots, n$).

Ví dụ: 2 chuỗi $X = \text{GATCCGA}$ và $Y = \text{GAAAGCAGA}$ có thể cân bằng thành:

$$X^* = \text{G-A--TCCGA}$$

$$Y^* = \text{GAAAG-CAGA}$$

hoặc: $X^* = \text{GA---TCCGA}$

$$Y^* = \text{GAAAG-CAGA}$$

Số đo $\text{GPS}(X^*, Y^*)$ của cân bằng được tính như sau: $\text{GPS}(X^*, Y^*) = S + Z$ với

- $S = \sigma(x_1, y_1) + \dots + \sigma(x_n, y_n)$ trong đó:

$$\sigma(x_i, y_i) = \begin{cases} 2, & \text{nếu } x_i = y_i \\ -1, & \text{nếu } x_i \neq y_i \end{cases} \quad (i = 1, 2, \dots, n)$$

- $Z = -4 * m$ với m là số đoạn thuộc X^* và Y^* mà mỗi đoạn gồm các ký tự '-' liên tiếp.

Ở hai ví dụ trên, GPS tương ứng sẽ là:

$$\begin{aligned} & 2+(-1)+2+(-1)+(-1)+(-1)+2+(-1)+2+2+(-4*3) = -7 \\ \text{Và} & 2+2+(-1)+(-1)+(-1)+(-1)+2+(-1)+2+2+(-4*2) = -3 \end{aligned}$$

Một trong các cân bằng có số đo lớn nhất sẽ là:

$$X^* = \text{G--ATCCGA}$$

$$Y^* = \text{GAAAGCAGA}$$

$$\text{Với số đo tương ứng } \text{GPS}(X^*, Y^*) = 2+(-1)+(-1)+2+(-1)+2+(-1)+2+2+(-4*1) = 2$$

Yêu cầu: Cho 2 chuỗi X và Y , độ dài mỗi chuỗi không quá 500. Hãy xác định cách cân bằng để có số đo GPS lớn nhất.

Dữ liệu: Vào từ file văn bản GPSF.INP:

- Dòng đầu tiên chứa số nguyên K - số lượng tests ($K \leq 50$),
- K nhóm 2 dòng tiếp theo: chứa X và Y , mỗi chuỗi trên một dòng.

Kết quả: Đưa ra file văn bản GPSF.OUT trên K nhóm 3 dòng, mỗi nhóm có dòng đầu chứa một số nguyên - là GPS lớn nhất tìm được ứng với mỗi cặp X, Y ; hai dòng tiếp theo chứa lần lượt X^* và Y^* . Nếu có nhiều cặp X^*, Y^* thì chỉ cần đưa ra một trong chúng.

Ví dụ:

GPSF.INP
5
AAA
GGAAGAG
GATCCGA
GAAAGCAGA
ACGGCTTAGATCCGAGAGTTAGTAGTCCTAAGCTTGCA
AGCTTAGAAAGCAGACACTTGATCCTGACGGCTTGAA
TTGAGTAGTGTTTTAGTCCTACACGACACATCAAATTCGGACAAGGCCTAGCT
TTCAAGTCCTACAATGTGTGTCAAATTCGCTTGGCCGAAAGCC
TTTGGGAACGTGTGTAGACTTCCCCATGCGATGG
AACACACACGGACTTCATGCTGG

GPSF.OUT

-8
----AAA
GGAAGAG
2
G--ATCCGA
GAAAGCAGA
18
ACGGCTTAGATCCGAGAGTTAGTAG-TCCTAA--GCTTGCA
A--GCTTAGAAAGCAGA--CACTTGATCCTGACGGCTTGAA
20
TTGAGTAGTGTTTTAGTCCTACACGACACATCAAATTCGGACAAGGCC--TAGCT
TTCA-----AGTCCTACAATGTGTGTCAAATTCG--CTTGCCGAAAGCC
2
TTTGGGAACGTGTGTAGACTTCCCCATGCGATGG
-----AACACACACGGAATT---CATGC--TGG