

Bài tập Xử lý chuỗi – Tuần 2

Bài 1 – Từ ghép

Bạn phải tìm tất cả các từ ghép (hai từ đơn) trong một từ điển tiếng Anh. Một từ ghép là một từ trong từ điển được nối lại bởi đúng hai từ khác trong từ điển.

Dữ liệu: Vào từ file TUGHEP.INP. Dữ liệu chuẩn bao gồm một số từ viết thường, mỗi từ một dòng, theo thứ tự chữ cái. Số lượng từ < 120000.

Kết quả: Xuất ra file TUGHEP.OUT tất cả từ ghép, mỗi từ một dòng theo thứ tự chữ cái.

Ví dụ:

TUGHEP.INP	TUGHEP.OUT
a	alien
alien	newborn
born	
less	
lien	
never	
nevertheless	
new	
newborn	
the	
zebra	

Bài 2 – Chuỗi số

Cho một số, chúng ta có thể tạo một chuỗi số bằng cách

1. sắp xếp các chữ số của nó theo thứ tự giảm dần
2. sắp xếp các chữ số của nó theo thứ tự tăng dần
3. trừ số nhận được từ (1) cho số nhận được từ (2) để tạo ra một số mới
4. và lặp lại các bước này nếu số mới chưa xuất hiện trong chuỗi số

Lưu ý rằng 0 là một chữ số hợp lệ. Số lượng các số khác nhau trong chuỗi là *độ dài* của chuỗi. Bạn hãy viết chương trình đọc các số và xuất ra chuỗi số và độ dài chuỗi số cho mỗi số đọc được.

Dữ liệu: Vào từ file CHUOISO.INP. Dữ liệu bao gồm một dãy các số dương, tất cả đều nhỏ

hơn 10^9 , mỗi số nằm trên một dòng và kết thúc bằng số 0. File dữ liệu chứa tối đa 5000 số.

Kết quả: Xuất ra file CHUOISO.OUT. Kết quả bao gồm các chuỗi số phát sinh bởi các số dữ liệu, kế tiếp là các độ dài của dãy số theo định dạng như bên trong ví dụ mẫu. Sau mỗi chuỗi số và độ dài chuỗi là một dòng trống, bao gồm cả kết quả cuối. Không có chuỗi nào có nhiều hơn 1000 số khác nhau.

Ví dụ:

CHUOISO.INP	CHUOISO.OUT
123456789	So ban dau 123456789
1234	987654321 - 123456789 = 864197532
444	987654321 - 123456789 = 864197532
0	Chieu dai chuoi 2
	So ban dau 1234
	4321 - 1234 = 3087
	8730 - 378 = 8352
	8532 - 2358 = 6174
	7641 - 1467 = 6174
	Chieu dai chuoi 4
	So ban dau 444
	444 - 444 = 0
	0 - 0 = 0
	Chieu dai chuoi 2

Bài 3 – Bẻ khoá Văn bản

Một trong những hệ thống mã hoá sớm nhất thuộc về Hoàng đế Caesar: nếu một kí tự được mã hoá là kí tự thứ m trong bảng chữ cái, thay kí tự này bằng kí tự thứ $(m+k)$ trong đó k là một số nguyên cố định. Caesar dùng $k = 3$, hay kí tự d, làm mã khoá. Nghĩa là, “a” sẽ được mã hoá thành “d”, “b” thành “e”, và tương tự đến “z” thành “c”. Ví dụ với mã hoá $a \rightarrow k$, câu “this is a test” sẽ được mã hoá thành “drsc sc k docd”.

Bây giờ bạn có một câu mã hoá và biết được một từ được giải mã trong câu đó. Nhiệm vụ là phải tìm ra tất cả khoá mã hoá có thể của câu trên.

Dữ liệu: Vào từ file BEKHOA.INP, bắt đầu với một số, n , sau đó là n cặp bao gồm một câu mã hoá và một từ được giải mã. Mỗi bộ có hai dòng, dòng đầu tiên chứa câu được mã hoá

trong đó chỉ chứa các từ được cách nhau bằng khoảng trắng không có dấu câu, dòng thứ hai chứa một từ là một trong những từ được giải mã từ câu trên. Tất cả các từ đều được viết thường. Một từ dài tối đa 16 kí tự và không có dòng nào dài hơn 70 từ.

Kết quả: xuất ra file BEKHOA.OUT. Kết quả có n dòng tương ứng với n cặp câu mã hoá/từ giải mã. Mỗi dòng chứa các kí tự viết thường các hoá có thể. Nghĩa là nếu kí tự “a” được mã hoá thành “x” thì “x” sẽ được xuất trên dòng. Mỗi dòng kí tự phải được sắp xếp theo bảng chữ cái.

Ví dụ:

BEKHOA.INP	BEKHOA.OUT
2	k
drsc sc k docd	hl
test	
dl ruvd doha hp pz	
we	

Bài 4 – Tìm chuỗi trong Bảng kí tự

Cho một bảng $m \times n$ kí tự ($1 \leq m, n \leq 50$) và một danh sách của các từ, hãy tìm vị trí trong bảng mà từ được tìm thấy.

Một từ khớp theo đường thẳng các kí tự liên tục trong bảng. Một từ có thể khớp các kí tự trong lưới không phân biệt hoa thường. Việc so khớp có thể thực hiện theo bất kỳ một trong 8 hướng ngang, dọc hay chéo trong bảng.

Dữ liệu: Vào từ file BANGCHU.INP. Dữ liệu bắt đầu bởi một cặp số nguyên m và n , $1 \leq m, n \leq 50$, trên một dòng. Mỗi dòng trong số m dòng tiếp theo chứa n kí tự; đây là bảng kí tự cần tìm danh sách từ. Kí tự trong lưới có thể viết hoa hay thường. Theo sau bảng kí tự là một số nguyên k trên một dòng ($1 \leq k \leq 20$). k dòng tiếp theo chứa danh sách các từ cần tìm, mỗi từ một dòng. Các từ này chỉ chứa các kí tự viết hoa hay thường (không khoảng trắng, dấu cách hay các kí tự không phải là chữ cái khác).

Kết quả: Với mỗi từ trong danh sách, xuất ra một cặp số nguyên biểu diễn vị trí của từ tương ứng trong bảng. Các số nguyên được cách nhau bởi một khoảng trắng. Số nguyên đầu tiên là số dòng trong bảng mà kí tự đầu tiên của từ được tìm thấy (1 biểu diễn dòng trên cùng của bảng và m dòng dưới cùng). Số nguyên thứ hai là số cột trong bảng mà kí tự đầu tiên được tìm thấy (1 biểu diễn cột ngoài cùng bên trái và n cột ngoài cùng bên phải). Nếu một từ có thể được tìm thấy nhiều lần thì kết quả sẽ ứng với vị trí xuất hiện cao nhất của từ (dòng gần với dòng trên cùng nhất). Nếu hai hay nhiều từ cùng xuất hiện ở dòng cao nhất thì xuất ra vị trí bên trái nhất. Tất cả các từ đều có thể tìm thấy ít nhất một lần trong bảng.

Ví dụ:

BANGCHU.INP	BANGCHU.OUT
8 11	2 5
abcDEFGhigg	2 3
hEbkWalDork	1 2
FtyAwaldORm	7 8
FtsimrLqsrc	
byoArBeDeyv	
Klcbqwikomk	
strEBGadhrb	
yUiqlxcnBjf	
4	
Waldorf	
Bambi	
Betty	
Dagbert	