

Egyszerű tömörítő

A feladat.be fájl tartalmát tömörítve kell beleírni a feladat.ki fájlba, mégpedig a következő képpen:

Keress meg az öt leggyakoribb legalább 3 betűs szót a fájlban, és ezek helyére csak egy rövid kódot rakj. A kód két karakteres szó, a leggyakoribb szó kódja !1 , a második leggyakoribbé !2 ,... míg az ötödik leggyakoribbé pedig !5 .

A többi szót hagyd meg ahogy volt. A kódolt fájl elején fel kell sorolni a kódok jelentését.

A programnak dekódolnia is kell tudni. Így ha a fájl a ! karakterrel kezdődik, akkor be kell olvasni a kódokat és a hozzájuk tartozó szavakat, majd a szöveget visszaállítani az eredeti állapotába.

A normál szövegben nem szerepel a ! karakter. Mindig van leglább 5 fajta minimum hárombetűs szó a szövegben, így más esetre nem kell felkészülni. A szavak maximum 22 karakteresek, és csak az angol ábécé kisbetűit tartalmazzák. A szöveg maximum 2000 szavas. A dekódolás teszteseteknél egyáltalán nem biztos hogy a gyakorisági sorrend stimmel, de ez lényegtelen, mert dekódolásnál nem kell gyakoriságokat számolni.

Bemenet

A feladat.be fájl maximum 2000 db szót tartalmaz csupa szóközzel elválasztva egymástól, és az utolsó szó után nincs space hanem egy újsor karakter. Egyik szóban sem szerepel a ! karakter.

Figyelem lehet, hogy a feladat.be fájl formája nem az előbb leírtaknak megfelelő, hanem éppen a Kimenet-nél leírt formájú. Ekkor nyilván a feladat.ki formája felel meg a Bemenet-nek.

Kimenet

A kimenet.be fájl első öt sora tartalmazza a kódolási információkat. Mindegyikben az első szó a kód (pl.: !1), a második pedig az eredeti szó amit helyettesít. Közöttük egy space van és a második szó után újsor karakter.

Ezután pedig jön a kódolt szöveg, melyben nem kódolt szavak és kódok vegyesen követik egymást space-szel elválasztva, az utolsó szó után nincs space, csak egy újsor.

Figyelem lehet, hogy a feladat.ki fájl formája nem a most leírtaknak megfelelő, hanem éppen a Bemenet-nél leírt formájú. Ekkor nyilván a feladat.be formája felel meg a Kimenet-nek.

Példa1

bemenet

```
d acppva d pbsekhn pbsekhn lnm aqigu d lnm pbsekhn acppva pbsekhn fbse acppva  
acppva ilbnjc fbse ilbnjc lnm acppva
```

kimenet

!1 acppva !2 pbsekhn !3 lnm !4 fbse !5 ilbnjc d !1 d !2 !2 !3 aqigu d !3 !2 !1 !2 !4 !1
!1 !5 !4 !5 !3 !1

Példa2

bemenet

!1 lsuagac !2 aphqk !3 fqojh !4 hpqsl !5 lbs fff !4 !1 fff boedm !1 !3 !5 !1 !4 !5 fff boedm
!1 boedm !4 !2 !2 boedm !5

kimenet

fff hpqsl lsuagac fff boedm lsuagac fqojh lbs lsuagac hpqsl lbs fff boedm lsuagac
boedm hpqsl aphqk aphqk boedm lbs

Példa3

bemenet

nterse ulctndktlmusjiorq erevsmtpcatpi nterse ibqpcsmvdhbnovpopk ulctndktlmusji-
orq ii ulctndktlmusjiorq dbvd ulctndktlmusjiorq nterse ibqpcsmvdhbnovpopk qgpbtmstnftnsgrs
seqmdclsdkkuhvkkp qgpbtmstnftnsgrs seqmdclsdkkuhvkkp ulctndktlmusjiorq ajjvj-
vevuodq nterse qgpbtmstnftnsgrs vdtlrtsugmkbs dbvd dbvd hn erevsmtpcatpi ajjvj-
vevuodq teilolf ulctndktlmusjiorq ii ii ulctndktlmusjiorq ulctndktlmusjiorq nterse
ibqpcsmvdhbnovpopk teilolf ajjvjvevuodq ibqpcsmvdhbnovpopk dbvd ii ajjvjvevu-
odq nterse dbvd dhhgfkmbpa ibqpcsmvdhbnovpopk ulctndktlmusjiorq dbvd ajjvjve-
vuodq ii seqmdclsdkkuhvkkp nterse qgpbtmstnftnsgrs teilolf dbvd ibqpcsmvdhbnov-
popk erevsmtpcatpi seqmdclsdkkuhvkkp dhhgfkmbpa ibqpcsmvdhbnovpopk ulctndktl-
musjiorq teilolf teilolf hn dhhgfkmbpa teilolf erevsmtpcatpi ulctndktlmusjiorq vdtlr-
sugmkbs erevsmtpcatpi ii ii qgpbtmstnftnsgrs ibqpcsmvdhbnovpopk dhhgfkmbpa
vdtlrtsugmkbs teilolf teilolf ibqpcsmvdhbnovpopk teilolf vdtlrtsugmkbs qgpbtmstnftnsgrs
ajjvjvevuodq dbvd hn seqmdclsdkkuhvkkp qgpbtmstnftnsgrs dhhgfkmbpa ulctndktl-
musjiorq ajjvjvevuodq teilolf teilolf qgpbtmstnftnsgrs qgpbtmstnftnsgrs dbvd vdtlr-
sugmkbs dbvd vdtlrtsugmkbs dhhgfkmbpa nterse dhhgfkmbpa dhhgfkmbpa qgpbtmstnftnsgrs

kimenet

!1 ulctndktlmsujiorq !2 teilolf !3 dbvd !4 qgpbtmstnftnsgrs !5 ibqpcsmvdhbnovpopk
ntersc !1 erevsmtpcatpi ntersc !5 !1 ii !1 !3 !1 ntersc !5 !4 seqmdclsdkkuhvkkp !4
seqmdclsdkkuhvkkp !1 ajjvjvevuodq ntersc !4 vdtlrtsugmkbs !3 !3 hn erevsmtpcatpi
ajjvjvevuodq !2 !1 ii ii !1 !1 ntersc !5 !2 ajjvjvevuodq !5 !3 ii ajjvjvevuodq
ntersc !3 dhhgfkmbpa !5 !1 !3 ajjvjvevuodq ii seqmdclsdkkuhvkkp ntersc !4 !2 !3
!5 erevsmtpcatpi seqmdclsdkkuhvkkp dhhgfkmbpa !5 !1 !2 !2 hn dhhgfkmbpa !2
erevsmtpcatpi !1 vdtlrtsugmkbs erevsmtpcatpi ii ii !4 !5 dhhgfkmbpa vdtlrtsugmkbs
!2 !2 !5 !2 vdtlrtsugmkbs !4 ajjvjvevuodq !3 hn seqmdclsdkkuhvkkp !4 dhhgfkmbpa
!1 ajjvjvevuodq !2 !2 !4 !4 !3 vdtlrtsugmkbs !3 vdtlrtsugmkbs dhhgfkmbpa ntersc
dhhgfkmbpa dhhgfkmbpa !4

Hivatkozások

[1] Kövesi-Nagy Dániel: *h254153@stud.u-szeged.hu*