

Pattern Compression

СЕЗОН 10 – ФИНАЛ



Ето че матурата по български език и литература свърши, оценките на учениците са ясни и г-жа Лалка Денчева си отдъхна, че няма да пише (много) обяснения заради определени субекти, които са се провалили с гръм и трясък. След като направи подробен анализ на резултатите обаче, тя установи, че създаването на собствен аргументативен текст е затруднило голяма част от нейните ученици. Досега тя тайно подозираше, че не е възможно целият випуск, на който преподава, да сътворява такива шедьоври на интерпретативния жанр, каквито тя цяла година четеше денем и нощем. Сега нейните подозрения се потвърдиха и тя реши, че не може повече да допуска учениците ѝ да я заблуждават.

Г-жа Лалка Денчева копира едно от съчиненията, получени по време на електронното обучение, постави текста в интернет търсачката и ... получи грешка 414. След многобройни опити да разреши този проблем, сред които беше и закупуване на нов лаптоп, учителката се допита до един компютърно грамотен ученик от нейния клас, на когото може да се има доверие. Той с удоволствие ѝ обясни, че интернет търсачката не може да обработва толкова дълги заявки за търсене и предложи на учителката си да използва за шаблон само част от съчинението. Това обаче не се хареса на г-жа Денчева и ученикът се зае търси друго решение.

За целта на тази задача можем да приемем, че текстът на съчинението се състои от N малки латински букви. Необходимо е да се намери подходящ шаблон, който да се използва за търсенето на целия текст в търсачката. За щастие търсачката предоставя няколко опции за съкратено представяне на шаблона:

- 1) Валиден шаблон е всяка последователност от малки латински букви
 - ❖ Например: "aab" е шаблон за търсене на текста "aab"
- 2) Ако P е валиден шаблон за текста T , то $P?(l_1, l_2, \dots, l_k)$ е валиден шаблон за търсене на резултатите от конкатенацията на T и всяка от буквите l_1, l_2, \dots, l_k
 - ❖ Например: "aab?(c, d)" е шаблон за търсене на текстовете "aabc" и "aabd"
- 3) Ако P е валиден шаблон за множеството от текстове S , то $[P, k]$ е валиден шаблон за търсене на резултатите от конкатенацията на незадължително различните текстове T_1, T_2, \dots, T_k , всеки от които принадлежи на S , където k е по-голямо от 1
 - ❖ Например: "[aab?(c, d), 2]" е шаблон за търсене на текстовете "aabcaabc", "aabcaabd" и "aabdaabd"
- 4) Ако P е валиден шаблон за търсене на текста T , то такъв е и шаблонът получен от конкатенацията на: '#', '=', Q , '.' и P' , където Q е последователност от малки латински букви, а P' е получен от P чрез заместване на (част от) срещанията на Q със символа '#'. Този вид представяне може да бъде използван само в началото.
 - ❖ Например: "#=aab.#c#d" е шаблон за търсене на текста "aabcaabd"

Разбира се, съществуват множество различни шаблони, които могат да бъдат използвани за търсене на определен текст чрез търсачката. Ясно е, че шаблонът не трябва да е прекалено дълъг, но освен това той трябва да е сравнително точен, за да може търсеният текст да бъде открит.

Нека дефинираме следните две функции:

- ❖ $length(P)$ – броят на всички символи, от които се състои шаблонът P
- ❖ $uncertainty(T, P)$ – за всяка малка латинска буква от текста T , която е представена в шаблона P чрез '?', се сумира броят на различните букви, които може да замества

Вашата задача е по даден текст T да създадете валиден шаблон P , така че сумата от двете функции $length(P) + uncertainty(T, P)$ да е колкото се може по-малка.

Вход

От първия ред на входния файл `pattern.in` се въвежда символен низ T , съставен от N малки латински букви.

Изход

На единствения ред от изходния файл `pattern.out` изведете един низ, представляващ валиден шаблон за търсене на зададения текст. Той трябва да е съставен единствено от малки латински букви, цифри и символите '[', ']', '(', ')', '.', ',', '#', '=' и '?'. Дължината на шаблона не трябва да надвишава N .

Ограничения

$$1 \leq N \leq 10^5$$

Примери:

| Вход | Изход |
|------------------------------|------------------------|
| aaaaaaaaaab | [a,10]b |
| ababxababyababzababz | [[ab,2]?(x,y,z),4] |
| dingdangdongdangdingdongdang | [d?(i,a,o)ng,7] |
| dingdangdongdangdingdongdang | #=ngd.di#a#o#a#i#o#ang |

Обяснение:

В първия тестов пример $length(P) = 7$, а $uncertainty(T, P) = 0$, защото в шаблона няма въпросителни знаци.

Във втория тестов пример $length(P) = 18$, а $uncertainty(T, P) = 4 * 3 = 12$, защото 4 от буквите в текста са представени чрез '?', която може да замества 3 различни букви.

В третия тестов пример $length(P) = 15$, а $uncertainty(T, P) = 7 * 3 = 21$.

В четвъртия тестов пример $length(P) = 22$, а $uncertainty(T, P) = 0$.

Оценяване:

За всеки тест нека $minScore$ е най-малкият резултат измежду резултатите на всички участници, а $yourScore$ е вашият резултат, получен като сума от стойностите на функциите $length(P) + uncertainty(T, P)$. Ще получите $\frac{minScore + 1}{yourScore + 1} \times 100\%$ от точките, предвидени за този тест.

Подзадачи

| # | Брой тестове | Ограничение за N | Допълнителни ограничения |
|---|--------------|--------------------|---|
| 1 | 10% | $N \leq 100$ | Няма |
| 2 | 20% | $N \leq 1000$ | T е конкатенация на символни низове с равна дължина, не по-голяма от 4. Измежду тези низове има най-много 4 различни. |
| 3 | 30% | $N \leq 10000$ | T се състои от най-много 4 различни букви |
| 4 | 40% | $N \leq 100000$ | Няма |