

## C: Compact Terms

Source file name: compact.c, compact.cpp, compact.java, or compact.py

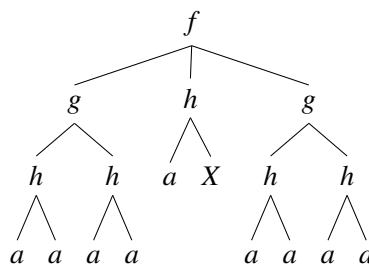
Author: Camilo Rocha

A term is a tree-like structure used in mathematics, computer science, and philosophy to identify entities in a domain of discourse. What is remarkable about terms is that they offer a direct encoding of such entities that a computer can process.

For a particular domain of discourse, the *set of terms* is parametric on a *set of variables*  $\mathcal{V}$  and a *set of function symbols*  $\mathcal{F}$ ; it is inductively defined by the following rules:

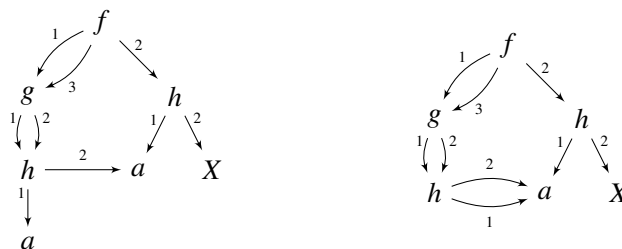
- Any variable is a term.
- Any function symbol  $f$  with  $\text{ar}(f) = 0$  is a term.
- Any expression  $f(t_1, \dots, t_n)$ , with  $\text{ar}(f) = n \geq 1$  and each argument  $t_i$  a term ( $1 \leq i \leq n$ ), is a term.

Here  $\text{ar}$  is a function from the set of function symbols  $\mathcal{F}$  to the natural numbers. For each  $f \in \mathcal{F}$ , the expression  $\text{ar}(f)$  denotes the number of arguments of  $f$ . For example, for  $\mathcal{F} = \{a, f, g, h\}$  with  $\text{ar}(a) = 0$ ,  $\text{ar}(g) = \text{ar}(h) = 2$ ,  $\text{ar}(f) = 3$ , and  $X$  a variable, let  $t$  be the term  $f(g(h(a, a), h(a, a)), h(a, X), g(h(a, a), h(a, a)))$ . The term  $t$  can be graphically represented as a labeled tree with 18 vertices as follows:



A string representation of a term, in general, can be unnecessarily lengthy and thus inefficient to compute with. An important observation to cope with this limitation is that subterms can be shared and then a tree can be represented by a *directed acyclic multigraph* (or *dam*). In a dam, vertices are labeled with function symbols from  $\mathcal{F}$  and variables from  $\mathcal{V}$ , and edges are labeled with the argument position of the corresponding subterm in the term structure. Note that there can be more than one edge between any pair of vertices of a dam.

The following dams are representations of the term  $t$  abovementioned. In these representations, for example, the topmost function symbol  $f$  is applied over three terms of which the first and third are shared: this is indicated by labels 1 and 3 in the edges with source  $f$ .



Although these dams represent the same term, they differ in the number of vertices: the one on the left has 7 vertices and the one on the right has 6 vertices. There can be many other dams representing the term  $t$ , but any of them requires at least 6 vertices.

In this problem, the interest is in computing the size (with respect to the number of vertices) of a dam representing a term which shares the most number of subterms. More precisely, the goal is to compute the minimum number of vertices required to represent a term as a dam.

### Input

The input consists of several test cases, each being a line containing a string representation of a term  $s$  ( $1 \leq |s| \leq 10^5$ ). A variable in  $\mathcal{V}$  is represented in  $s$  by a string  $v$  ( $1 \leq |v| \leq 20$ ) made of lowercase and uppercase letters of the English alphabet starting with an uppercase letter. Similarly, a function symbol in  $\mathcal{F}$  is represented in  $s$  by a string  $f$  ( $1 \leq |f| \leq 20$ ) made of lowercase and uppercase letters of the English alphabet starting with a lowercase letter. You can assume that  $s$  contains at most 2 000 variables and function symbols (not necessarily different).

*The input must be read from standard input.*

### Output

For each test case output the minimum number of vertices required to represent  $s$  as a dam.

*The output must be written to standard output.*

Sample Input	Sample Output
$f(g(h(a, a), h(a, a)), h(a, X), g(h(a, a), h(a, a)))$	6
$h(a, b)$	3
$h(a, a)$	2
$h(A, a)$	3
CCPL	1
$f(a, a, a, a, a, a, a, a, a, a, a, a, a, a, a, a, a, a, a, a)$	2
$f(g(a, X, a), g(a, a, X))$	5
$f(g(a, X, a), g(a, X, a))$	4