The Jaccard index

Cost: 15 | Solved: 101

Memory limit: 256 MBs

Time limit: 1 s

Input: input.txt

Output: output.txt

Task:

You are given two strings S_1 and S_2 that can contain any ASCII symbols, except for the line feed character.

Normalize the strings: delete all symbols that are not latin letters, digits or a whitespace character, and decapitalize every letter (make it lowercase). You should then delete all words which length is less or equal to 3 and all extra whitespace characters (so that there's no more than one whitespace character between two words).

For the resulting strings count the following value: $c / (|S_1| + |S_2| - c)$, where c is the number of matching symbols between the strings S_1 and S_2 and $|S_n|$ is the length of the n-th string.

Input:

Two strings S_1 and S_2 separated with a line feed character.

Output:

The calculated value with at least 10^{-6} accuracy.

Example:

Input	Output
□L8Q4EJv"J,6696 7 3AK.&+&'42a o □S M8KBK5□43V 1□J□ 6K2Z	0.030303030303030303871

Report a bug (/en/webform-feedback/nojs?submittedfrom=tasks/task/47443)