

## TEST 19

1.

여러 언론사에서 쏟아지는 뉴스, 특히 속보성 뉴스를 보면 비슷비슷한 제목의 기사가 많아 정작 필요한 기사를 찾기가 어렵다. Daum 뉴스의 개발 업무를 맡게 된 신입사원 남영이는 사용자들이 편리하게 다양한 뉴스를 찾아볼 수 있도록 문제점을 개선하는 업무를 맡게 되었다.

개발의 방향을 잡기 위해 남영이는 우선 최근 화제가 되고 있는 "카카오 신입 개발자 공채" 관련 기사를 검색해보았다.

- 카카오 첫 공채.. '블라인드' 방식 채용
- 카카오, 합병 후 첫 공채.. 블라인드 전형으로 개발자 채용
- 카카오, 블라인드 전형으로 신입 개발자 공채
- 카카오 공채, 신입 개발자 코딩 능력만 본다
- 카카오, 신입 공채.. "코딩 실력만 본다"

카카오 "코딩 능력만으로 2018 신입 개발자 뽑는다"기사의 제목을 기준으로 "블라인드 전형"에 주목하는 기사와 "코딩 테스트"에 주목하는 기사로 나뉘는 걸 발견했다. 남영이는 이들을 각각 묶어서 보여주면 카카오 공채 관련 기사를 찾아보는 사용자에게 유용할 듯싶었다.

유사한 기사를 묶는 기준을 정하기 위해서 논문과 자료를 조사하던 남영이는 "자카드 유사도"라는 방법을 찾아냈다.

## TEST 19

자카드 유사도는 집합 간의 유사도를 검사하는 여러 방법 중의 하나로 알려져 있다. 두 집합 A, B 사이의 자카드 유사도  $J(A, B)$ 는 두 집합의 교집합 크기를 두 집합의 합집합 크기로 나눈 값으로 정의된다.

예를 들어 집합  $A = \{1, 2, 3\}$ , 집합  $B = \{2, 3, 4\}$ 라고 할 때, 교집합  $A \cap B = \{2, 3\}$ , 합집합  $A \cup B = \{1, 2, 3, 4\}$ 이 되므로, 집합 A, B 사이의 자카드 유사도  $J(A, B) = 2/4 = 0.5$ 가 된다. 집합 A와 집합 B가 모두 공집합일 경우에는 나눗셈이 정의되지 않으니 따로  $J(A, B) = 1$ 로 정의한다.

자카드 유사도는 원소의 중복을 허용하는 다중집합에 대해서 확장할 수 있다.

다중집합 A는 원소 "1"을 3개 가지고 있고, 다중집합 B는 원소 "1"을 5개 가지고 있다고 하자.

이 다중집합의 교집합  $A \cap B$ 는 원소 "1"을  $\min(3, 5)$ 인 3개, 합집합  $A \cup B$ 는 원소 "1"을  $\max(3, 5)$ 인 5개 가지게 된다.

다중집합  $A = \{1, 1, 2, 2, 3\}$ , 다중집합  $B = \{1, 2, 2, 4, 5\}$ 라고 하면, 교집합  $A \cap B = \{1, 2, 2\}$ , 합집합  $A \cup B = \{1, 1, 2, 2, 3, 4, 5\}$ 가 되므로, 자카드 유사도  $J(A, B) = 3/7$ , 약 0.42가 된다.

이를 이용하여 문자열 사이의 유사도를 계산하는데 이용할 수 있다.

문자열 "FRANCE"와 "FRENCH"가 주어졌을 때, 이를 두 글자씩 끊어서 다중집합을 만들 수 있다. 각각  $\{FR, RA, AN, NC, CE\}$ ,  $\{FR, RE, EN, NC, CH\}$ 가 되며, 교집합은  $\{FR, NC\}$ , 합집합은  $\{FR, RA, AN, NC, CE, RE, EN, CH\}$ 가 되므로, 두 문자열 사이의 자카드 유사도  $J(\text{"FRANCE"}, \text{"FRENCH"}) = 2/8 = 0.25$ 가 된다.

## TEST 19

### 제한사항

- ▶ 입력으로는 `str1`과 `str2`의 두 문자열이 들어온다. 각 문자열의 길이는 2 이상, 1,000 이하이다.
- ▶ 입력으로 들어온 문자열은 두 글자씩 끊어서 다중집합의 원소로 만든다. 이때 영문자로 된 글자 쌍만 유효하고, 기타 공백이나 숫자, 특수 문자가 들어있는 경우는 그 글자 쌍을 버린다. 예를 들어 "ab+"가 입력으로 들어오면, "ab"만 다중집합의 원소로 삼고, "b+"는 버린다.
- ▶ 다중집합 원소 사이를 비교할 때, 대문자와 소문자의 차이는 무시한다. "AB"와 "Ab", "ab"는 같은 원소로 취급한다.
- ▶ 입력으로 들어온 두 문자열의 자카드 유사도를 출력한다. 유사도 값은 0에서 1 사이의 실수이므로, 이를 다루기 쉽도록 65536을 곱한 후에 소수점 아래를 버리고 정수부만 출력한다.

### 입출력 예

str1	str2	answer
FRANCE	French	16384
handshake	shake hands	65536
aa1+aa2	AAAA12	43690
E=M*C^2	e=m*c^2	65536

## TEST 19

2.

셀수있는 수량의 순서있는 열거 또는 어떤 순서를 따르는 요소들의 모음을 튜플(tuple)이라고 합니다.  $n$ 개의 요소를 가진 튜플을  $n$ -튜플( $n$ -tuple)이라고 하며, 다음과 같이 표현할 수 있습니다.

- $(a_1, a_2, a_3, \dots, a_n)$

튜플은 다음과 같은 성질을 가지고 있습니다.

1. 중복된 원소가 있을 수 있습니다. ex :  $(2, 3, 1, 2)$
2. 원소에 정해진 순서가 있으며, 원소의 순서가 다르면 서로 다른 튜플입니다. ex :  $(1, 2, 3) \neq (1, 3, 2)$
3. 튜플의 원소 개수는 유한합니다.

원소의 개수가  $n$ 개이고, 중복되는 원소가 없는 튜플  $(a_1, a_2, a_3, \dots, a_n)$ 이 주어질 때(단,  $a_1, a_2, \dots, a_n$ 은 자연수), 이는 다음과 같이 집합기호 '{', '}'를 이용해 표현할 수 있습니다.

- $\{\{a_1\}, \{a_1, a_2\}, \{a_1, a_2, a_3\}, \{a_1, a_2, a_3, a_4\}, \dots, \{a_1, a_2, a_3, a_4, \dots, a_n\}\}$

예를 들어 튜플이  $(2, 1, 3, 4)$ 인 경우 이는

- $\{\{2\}, \{2, 1\}, \{2, 1, 3\}, \{2, 1, 3, 4\}\}$

와 같이 표현할 수 있습니다. 이때, 집합은 원소의 순서가 바뀌어도 상관없으므로

- $\{\{2\}, \{2, 1\}, \{2, 1, 3\}, \{2, 1, 3, 4\}\}$
- $\{\{2, 1, 3, 4\}, \{2\}, \{2, 1, 3\}, \{2, 1\}\}$
- $\{\{1, 2, 3\}, \{2, 1\}, \{1, 2, 4, 3\}, \{2\}\}$

는 모두 같은 튜플  $(2, 1, 3, 4)$ 를 나타냅니다.

특정 튜플을 표현하는 집합이 담긴 문자열  $s$ 가 매개변수로 주어질 때,  $s$ 가 표현하는 튜플을 배열에 담아 return 하도록 solution 함수를 완성해주세요.

## TEST 19

### 제한사항

- ▶ s의 길이는 5 이상 1,000,000 이하입니다.
- ▶ s는 숫자와 '{', '}', ',', ' ' 로만 이루어져 있습니다.
- ▶ 숫자가 0으로 시작하는 경우는 없습니다.
- ▶ s는 항상 중복되는 원소가 없는 튜플을 올바르게 표현하고 있습니다.
- ▶ s가 표현하는 튜플의 원소는 1 이상 100,000 이하인 자연수입니다.
- ▶ return 하는 배열의 길이가 1 이상 500 이하인 경우만 입력으로 주어집니다.

### 입출력 예

s	Result
"{{2}, {2,1},{2,1,3},{2,1,3,4}}"	[2,1,3,4]
"{{1,2,3},{2,1},{1,2,4,3},{2,}}"	[2,1,3,4]
"{{20,111},{111}}"	[111,20]
"{{123}}"	[123]
"{{4,2,3},{3},{2,3,4,1},{2,3}}"	[3,2,4,1]

### #입출력 예 3

(111, 20)을 집합 기호를 이용해 표현하면 {{111}, {111,20}}이 되며, 이는 {{20,111},{111}}과 같습니다.

### #입출력 예#4

(123)을 집합 기호를 이용해 표현하면 {{123}} 입니다.

### #입출력 예 5

(3, 2, 4, 1)을 집합 기호를 이용해 표현하면 {{3},{3,2},{3,2,4},{3,2,4,1}}이 되며, 이는 {{4,2,3},{3},{2,3,4,1},{2,3}}과 같습니다.