

# String Similarity

According to [http://en.wikipedia.org/wiki/Jaccard\\_index](http://en.wikipedia.org/wiki/Jaccard_index)

The **Jaccard index**, also known as the **Jaccard similarity coefficient** (originally coined *coefficient de communauté* by Paul Jaccard), is a statistic used for comparing the similarity and diversity of sample sets. The Jaccard coefficient measures similarity between finite sample sets, and is defined as the size of the intersection divided by the size of the union of the sample sets:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

(If  $A$  and  $B$  are both empty, we define  $J(A, B) = 1$ .) Clearly,

$$0 \leq J(A, B) \leq 1$$

In this problem you are to complete three methods in the `StringSimilarity` class. The three methods are `stringUnion`, `stringIntersection`, and the `getJaccardIndex` method. These methods are case sensitive. That is, "A" and "a" are different.

The `stringUnion` method has two `String` parameters and returns a list of `List<String>`. Each element of the `List` has length 1. All elements of list are contained in either parameter (or both). Duplicate elements may exist in the list if either parameter contains duplicate elements.

The following table shows the results of two calls of the `stringUnion` method.

The method	Returns a List containing the following elements
<code>stringUnion("ABC", "CDE")</code>	"A", "B", "C", "D", "E"
<code>stringUnion("ABBCDD", "BCDDDE")</code>	"A", "B", "B", "C", "D", "D", "D", "E"
<code>stringUnion("AB", "ab")</code>	"A", "B", "a", "b"

The `stringIntersection` method has two `String` parameters and returns a `List<String>`. Each element of the List has length 1 and all elements of list are contained in both parameters. Duplicate elements may exist in the List if both parameters contain the same duplicate elements.

The following table shows the results of two calls of the `stringIntersection` method.

The method	Returns a List containing the following elements
<code>stringIntersection("ABC", "CDE")</code>	<code>"C"</code>
<code>stringIntersection("ABBCDD", "BCDDDE")</code>	<code>"B", "C", "D", "D"</code>
<code>stringIntersection("ABC", "abc");</code>	<code>"c"</code>

The `getJaccardIndex` method has two `String` parameters and returns the number of elements in the String Intersection divided by the number of elements in the String Union. The return value is a `double` and should demonstrate division use by primitives of the type `double`.

The following table shows the results of two calls of the `getJaccardIndex` method.

The method	Returns the following double
<code>getJaccardIndex("ABC", "CDE")</code>	$0.2 = 1.0 / 5.0$
<code>getJaccardIndex("ABBCDD", "BCDDDE")</code>	$0.5 = 4. / 8. = 4. / (1+2+1+3+1)$