**3. Longest Substring Without Repeating Characters**

**Question**

Given a string, find the length of the **longest substring** without repeating characters.

**Examples:**

Given "abcabcbb", the answer is "abc", which the length is 3.

Given "bbbbb", the answer is "b", with the length of 1.

Given "pwwkew", the answer is "wke", with the length of 3. Note that the answer must be a **substring**, "pwke" is a *subsequence* and not a substring.

**Quick Navigation**

* [Solution](https://leetcode.com/articles/longest-substring-without-repeating-characters/#solution)
  + [Approach #1 Brute Force [Time Limit Exceeded]](https://leetcode.com/articles/longest-substring-without-repeating-characters/#approach-1-brute-force-time-limit-exceeded)
  + [Approach #2 Sliding Window [Accepted]](https://leetcode.com/articles/longest-substring-without-repeating-characters/#approach-2-sliding-window-accepted)
  + [Approach #3 Sliding Window Optimized [Accepted]](https://leetcode.com/articles/longest-substring-without-repeating-characters/#approach-3-sliding-window-optimized-accepted)

**Solution**

**Approach #1 Brute Force [Time Limit Exceeded]**

**Intuition**

Check all the substring one by one to see if it has no duplicate character.

**Algorithm**

Suppose we have a function boolean allUnique(String substring) which will return true if the characters in the substring are all unique, otherwise false. We can iterate through all the possible substrings of the given string s and call the function allUnique. If it turns out to be true, then we update our answer of the maximum length of substring without duplicate characters.

Now let's fill the missing parts:

1. To enumerate all substrings of a given string, we enumerate the start and end indices of them. Suppose the start and end indices are iii and jjj, respectively. Then we have 0≤i<j≤n0≤i<j≤n (here end index jjj is exclusive by convention). Thus, using two nested loops with iii from 0 to n−1n - 1n−1 and jjj from i+1i+1i+1 to nnn, we can enumerate all the substrings of s.
2. To check if one string has duplicate characters, we can use a set. We iterate through all the characters in the string and put them into the set one by one. Before putting one character, we check if the set already contains it. If so, we return false. After the loop, we return true.

**Java**

**public** **class** **Solution** **{**

**public** **int** **lengthOfLongestSubstring(**String s**)** **{**

**int** n **=** s**.**length**();**

**int** ans **=** 0**;**

**for** **(int** i **=** 0**;** i **<** n**;** i**++)**

**for** **(int** j **=** i **+** 1**;** j **<=** n**;** j**++)**

**if** **(**allUnique**(**s**,** i**,** j**))** ans **=** Math**.**max**(**ans**,** j **-** i**);**

**return** ans**;**

**}**

**public** **boolean** **allUnique(**String s**,** **int** start**,** **int** end**)** **{**

Set**<**Character**>** set **=** **new** HashSet**<>();**

**for** **(int** i **=** start**;** i **<** end**;** i**++)** **{**

Character ch **=** s**.**charAt**(**i**);**

**if** **(**set**.**contains**(**ch**))** **return** **false;**

set**.**add**(**ch**);**

**}**

**return** **true;**

**}**

**}**

**Complexity Analysis**

* Time complexity : O(n3)O(n^3)O(n​3​​).

To verify if characters within index range [i,j)[i, j)[i,j) are all unique, we need to scan all of them. Thus, it costs O(j−i)O(j - i)O(j−i) time.

For a given i, the sum of time costed by each j∈[i+1,n]j \in [i+1, n]j∈[i+1,n] is

∑i+1nO(j−i) \sum\_{i+1}^{n}O(j - i) ​i+1​∑​n​​O(j−i)

Thus, the sum of all the time consumption is:

O(∑i=0n−1(∑j=i+1n(j−i)))=O(∑i=0n−1(1+n−i)(n−i)2)=O(n3) O\left(\sum\_{i = 0}^{n - 1}\left(\sum\_{j = i + 1}^{n}(j - i)\right)\right) = O\left(\sum\_{i = 0}^{n - 1}\frac{(1 + n - i)(n - i)}{2}\right) = O(n^3) O(​i=0​∑​n−1​​(​j=i+1​∑​n​​(j−i)))=O(​i=0​∑​n−1​​​2​​(1+n−i)(n−i)​​)=O(n​3​​)

* Space complexity : O(min(n,m))O(min(n, m))O(min(n,m)). We need O(k)O(k)O(k) space for checking a substring has no duplicate characters, where kkk is the size of the Set. The size of the Set is upper bounded by the size of the string nnn and the size of the charset/alphabet mmm.

**Approach #2 Sliding Window [Accepted]**

**Algorithm**

The naive approach is very straightforward. But it is too slow. So how can we optimize it?

In the naive approaches, we repeatedly check a substring to see if it has duplicate character. But it is unnecessary. If a substring sijs\_{ij}s​ij​​ from index iii to j−1j - 1j−1 is already checked to have no duplicate characters. We only need to check if s[j]s[j]s[j] is already in the substring sijs\_{ij}s​ij​​.

To check if a character is already in the substring, we can scan the substring, which leads to an O(n2)O(n^2)O(n​2​​) algorithm. But we can do better.

By using HashSet as a sliding window, checking if a character in the current can be done in O(1)O(1)O(1).

A sliding window is an abstract concept commonly used in array/string problems. A window is a range of elements in the array/string which usually defined by the start and end indices, i.e. [i,j)[i, j)[i,j) (left-closed, right-open). A sliding window is a window "slides" its two boundaries to the certain direction. For example, if we slide [i,j)[i, j)[i,j) to the right by 111 element, then it becomes [i+1,j+1)[i+1, j+1)[i+1,j+1) (left-closed, right-open).

Back to our problem. We use HashSet to store the characters in current window [i,j)[i, j)[i,j) (j=ij = ij=i initially). Then we slide the index jjj to the right. If it is not in the HashSet, we slide jjj further. Doing so until s[j] is already in the HashSet. At this point, we found the maximum size of substrings without duplicate characters start with index iii. If we do this for all iii, we get our answer.

**Java**

**public** **class** **Solution** **{**

**public** **int** **lengthOfLongestSubstring(**String s**)** **{**

**int** n **=** s**.**length**();**

Set**<**Character**>** set **=** **new** HashSet**<>();**

**int** ans **=** 0**,** i **=** 0**,** j **=** 0**;**

**while** **(**i **<** n **&&** j **<** n**)** **{**

*// try to extend the range [i, j]*

**if** **(!**set**.**contains**(**s**.**charAt**(**j**))){**

set**.**add**(**s**.**charAt**(**j**++));**

ans **=** Math**.**max**(**ans**,** j **-** i**);**

**}**

**else** **{**

set**.**remove**(**s**.**charAt**(**i**++));**

**}**

**}**

**return** ans**;**

**}**

**}**

**Complexity Analysis**

* Time complexity : O(2n)=O(n)O(2n) = O(n)O(2n)=O(n). In the worst case each character will be visited twice by iii and jjj.
* Space complexity : O(min(m,n))O(min(m, n))O(min(m,n)). Same as the previous approach. We need O(k)O(k)O(k) space for the sliding window, where kkk is the size of the Set. The size of the Set is upper bounded by the size of the string nnn and the size of the charset/alphabet mmm.

**Approach #3 Sliding Window Optimized [Accepted]**

The above solution requires at most 2n steps. In fact, it could be optimized to require only n steps. Instead of using a set to tell if a character exists or not, we could define a mapping of the characters to its index. Then we can skip the characters immediately when we found a repeated character.

The reason is that if s[j]s[j]s[j] have a duplicate in the range [i,j)[i, j)[i,j) with index j′j'j​′​​, we don't need to increase iii little by little. We can skip all the elements in the range [i,j′][i, j'][i,j​′​​] and let iii to be j′+1j' + 1j​′​​+1 directly.

**Java (Using HashMap)**

**public** **class** **Solution** **{**

**public** **int** **lengthOfLongestSubstring(**String s**)** **{**

**int** n **=** s**.**length**(),** ans **=** 0**;**

Map**<**Character**,** Integer**>** map **=** **new** HashMap**<>();** *// current index of character*

*// try to extend the range [i, j]*

**for** **(int** j **=** 0**,** i **=** 0**;** j **<** n**;** j**++)** **{**

**if** **(**map**.**containsKey**(**s**.**charAt**(**j**)))** **{**

i **=** Math**.**max**(**map**.**get**(**s**.**charAt**(**j**)),** i**);**

**}**

ans **=** Math**.**max**(**ans**,** j **-** i **+** 1**);**

map**.**put**(**s**.**charAt**(**j**),** j **+** 1**);**

**}**

**return** ans**;**

**}**

**}**

**Java (Assuming ASCII 128)**

The previous implements all have no assumption on the charset of the string s.

If we know that the charset is rather small, we can replace the Map with an integer array as direct access table.

Commonly used tables are:

* int[26] for Letters 'a' - 'z' or 'A' - 'Z'
* int[128] for ASCII
* int[256] for Extended ASCII

**public** **class** **Solution** **{**

**public** **int** **lengthOfLongestSubstring(**String s**)** **{**

**int** n **=** s**.**length**(),** ans **=** 0**;**

**int[]** index **=** **new** **int[**128**];** *// current index of character*

*// try to extend the range [i, j]*

**for** **(int** j **=** 0**,** i **=** 0**;** j **<** n**;** j**++)** **{**

i **=** Math**.**max**(**index**[**s**.**charAt**(**j**)],** i**);**

ans **=** Math**.**max**(**ans**,** j **-** i **+** 1**);**

index**[**s**.**charAt**(**j**)]** **=** j **+** 1**;**

**}**

**return** ans**;**

**}**

**}**

**Complexity Analysis**

* Time complexity : O(n)O(n)O(n). Index jjj will iterate nnn times.
* Space complexity (HashMap) : O(min(m,n))O(min(m, n))O(min(m,n)). Same as the previous approach.
* Space complexity (Table): O(m)O(m)O(m). mmm is the size of the charset.