Shortest Common Super Sequence

Muhammad Zawad Mahmud
ID:1931401642
CSE373.5

Major in Computer Science & Engineering

Department of Electrical & Computer Engineering

North South University

What is Shortest Common Super Sequence?

To understand Shortest Common Super Sequence (SCS), first we need to know about sequence and longest common subsequence.

What is a sequence?

- A sequence is an enumerated collection of objects in which repetitions are allowed and order matters.
- For example: We have a string.
  Str-1: bcd
- Here a sequence of this string can be:
  cd or bc
- But bd is not a sequence of this string as the order is not maintained.

What is a longest common subsequence?

- The longest common subsequence (LCS) is defined as the longest subsequence that is common to all the given sequences, provided that the elements of the subsequence are not required to occupy consecutive positions within the original sequences.
- For example, We have two string.
- Str-1: AB
- Str-2: ABC
- Here common subsequence are: {AB}, {B}, {A},
- But the longest one is {AB}
  So LCS here is AB

Shortest Common Super Sequence (SCS)

- The shortest common super sequence of two sequences X and Y is the shortest sequence which has X and Y as subsequences This is a problem closely related to the longest common subsequence problem.
- For example,
- X= AGGTAB
- Y= GXTXAYB
- SCS of this two string is AGXGTXYAB
- Here the string has both AGGTAB and GXTXAYB as subsequence.
- Maybe the order is not maintained but that thing is not important here.

Things need to be covered

- To get the shortest common super sequence we need to get two things. Those are:
- Length of the shortest common super sequence.
- Find the string shortest common super sequence.

Shortest Common Super Sequence (SCS)

We have two string.

Str-1: A G G T A B

Str-2: G X T X A Y B

In worst case, the SCS is :

Str-3: A G G T A B G X T X A Y B

Str-1 length: 6

Str-1 length: 7

Str-3 length: (6+7)=13 (In worst case)

Here we can see many letters are common. To get rid of those LCS comes into play. Here the LCS of this two string is G T A B which's length is 4

So the length of SCS will be (Str-1 length+Str-2 length)-LCS length = 9

Shortest Common Super Sequence (SCS)

In the previous side, we got the length of the SCS. Now we will see how can we get the string of SCS.

Str-1: A G G T A B

Str-2: G X T X A Y B

LCS: G T A B

Str-3:

Shortest Common Super Sequence (SCS)

Str-1: A G G T A B

Str-2: G X T X A Y B

LCS: G T A B

Str-3: A

Shortest Common Super Sequence (SCS)

Str-1: A G G T A B

Str-2: G X T X A Y B

LCS: G T A B

Str-3: A G

Shortest Common Super Sequence (SCS)

Str-1: A G G T A B

Str-2: G X T X A Y B

LCS: G T A B

Str-3: A G G

Shortest Common Super Sequence (SCS)

Str-1: A G G T A B

Str-2: G X T X A Y B

LCS: G T A B

Str-3: A G G X

Shortest Common Super Sequence (SCS)

Str-1: A G G T A B

Str-2: G X T X A Y B

LCS: G T A B

Str-3: A G G X T

Shortest Common Super Sequence (SCS)

Str-1: A G G T A B

Str-2: G X T X A Y B

LCS: G T A B

Str-3: A G G X T X

Shortest Common Super Sequence (SCS)

Str-1: A G G T A B

Str-2: G X T X A Y B

LCS: G T A B

Str-3: A G G X T X A

Shortest Common Super Sequence (SCS)

Str-1: A G G T A B

Str-2: G X T X A Y B

LCS: G T A B

Str-3: A G G X T X A Y

Shortest Common Super Sequence (SCS)

Str-1: A G G T A B

Str-2: G X T X A Y B

LCS: G T A B

Str-3: A G G X T X A Y B

Shortest Common Super Sequence (SCS)

Str-1: A G G T A B

Str-2: G X T X A Y B

LCS: G T A B

Str-3: A G G X T X A Y B

So SCS is A G G X T X A Y B

Shortest Common Super Sequence (SCS) Runtime

Time Complexity of the SCS will be:

$$O(L1*l2)$$

Where L1 is the length of str-1 and L2 is the length of str-2.

Thank you