

Explanation: Define  $\text{overlap}(s, t)$  = the length of <sup>the longest</sup> substring which is suffix of  $s$  and prefix of  $t$ . Then every iteration, greedily merge the two strings which have largest overlap. Use a binary <sup>max</sup> heap to keep track of all overlaps left. Before merging, remove all strings which are substring of others.

Pseudocode: algorithm short-reads-reconstruct( $s\_list$ ):

$i\_set = \text{set}(\text{range}(0, \text{len}(s\_list)))$

Remove all strings which are others' substring ~~to~~ from  $i\_set$

for  $j$  in  $i\_set$ :

for  $k$  in  $i\_set$ ,  $k \neq j$ :

push ( $\text{overlap}(s\_list[j], s\_list[k])$ ,  $j, k$ )  
to heap

while  ~~$i\_set$  not empty~~ ( $\text{len}(i\_set) > 1$ ):

~~$(\text{max\_common}, \text{max\_index1}, \text{max\_index2})$~~

= pop (from heap) the largest element

Merge  $s\_list[\text{max\_index1}]$  and  $s\_list[\text{max\_index2}]$   
and store in  $s\_list[\text{max\_index1}]$

Remove  $\text{max\_index2}$  from  $i\_set$

for  $j$  in  $i\_set$ :

if  $j \neq \text{max\_index1}$ :

change overlap (~~in heap~~  $s\_list[\text{max\_index1}]$ ,  $s\_list[j]$ )  
in heap to  $\text{overlap}(s\_list[\text{max\_index2}], s\_list[j])$

Delete  $\text{overlap}(s\_list[j], s\_list[\text{max\_index2}])$

Delete  $\text{overlap}(s\_list[\text{max\_index2}], s\_list[j])$

Return ~~the~~  $s\_list$  [only element in  $i\_set$ ]

Running time:  $O(n^2 \log n + n^2 \log n)$

Calculate all overlap values and push to heap takes  $O(n^2 k^2 + n^2 \log n)$

Delete all strings which are substrings of others takes  $O(n^2)$

There are  $O(n)$  iterations, each of which takes  $O(\log n)$  to pop,  $O(k)$  to merge,  $O(1)$  to remove from  $i\_set$ , and

$n \cdot O(\log n) = O(n \log n)$  to delete + change overlap

Thus, the algorithm takes  $O(n^2k^2 + n^2 \log n)$  time.