


Intervallgrenzen l_x und r_y

	1	2	3	4	5	6	7	8	9	10	11	12
$S_2 =$	(1	<u>3</u>	<u>5</u>	<u>2</u>	<u>4</u>	<u>5</u>	<u>2</u>	<u>4</u>	<u>3</u>	<u>6</u>	<u>2</u>	5)
	l_2					l_1	ρ			r_1		r_2



- markierte Buchstaben in S_1 sind $C' = \{2,3,4\}$
- Buchstabe $c = 2$ wurde zuletzt markiert
- wir betrachten Position $\rho = 7$
- für $\delta = 1$ müssen wir die Intervalle [2,9], [2,11], [7,9] und [7,11] in S_2 überprüfen

Intervallgrenzen l_x und r_y

	1	2	3	4	5	6	7	8	9	10	11	12	
$S_2 =$	(1	3	5	2	4	5	2	4	3	6	2	5)	
	l_2						l_1	p				r_1	r_2

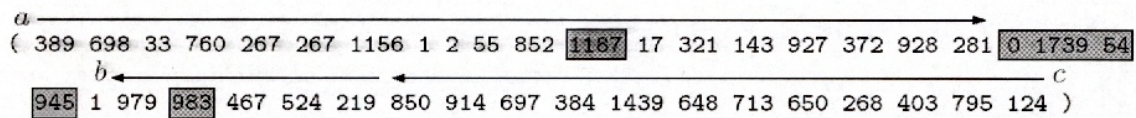
- markierte Buchstaben in S_1 sind $C' = \{2,3,4\}$
- Buchstabe $c = 2$ wurde zuletzt markiert
- wir betrachten Position $p = 7$
- für $\delta = 1$ müssen wir die Intervalle $[2,9]$, $[2,11]$, $[7,9]$ und $[7,11]$ in S_2 überprüfen

Connecting Intervals with Errors

```
1: build data structures POS and NUM for  $S_2$ 
2: resultSet  $\leftarrow \emptyset$ 
3: for  $i = 1, \dots, |S_1|$  do
4:   for each  $c \in \Sigma$  let  $OCC[c] \leftarrow 0$ 
5:    $|OCC| \leftarrow 0$ 
6:   minDist  $\leftarrow 0$ 
7:    $j = i$ 
8:   while  $j \leq |S_1|$  and  $S_1(i, j)$  is left-maximal do
9:      $c \leftarrow S_1[j]$ 
10:     $OCC[c] \leftarrow 1$ 
11:     $|OCC| \leftarrow |OCC| + 1$ 
12:    while  $S_1[i, j]$  is not right-maximal do
13:       $j \leftarrow j + 1$ 
14:    end while
15:    minDist  $\leftarrow \text{minDist} + 1$ 
16:    for each position  $p$  in  $POS[c]$  do
17:      mark position  $p$  in  $S_2$ 
18:      find positions  $l_1, \dots, l_{\delta+1}$  and  $r_1, \dots, r_{\delta+1}$ 
19:      for each pair  $(l_x, r_y)$  with  $1 \leq x, y \leq \delta + 1$  do
20:         $z \leftarrow$  the number of different unmarked characters in  $S_2[l_x + 1, r_y - 1]$ 
21:         $dist \leftarrow |OCC| - NUM[l_x + 1, r_y - 1] + 2z$ 
22:        if  $dist < \text{minDist}$  then
23:          minDist  $\leftarrow dist$ 
24:        end if
25:      end for
26:    end for
27:    if minDist  $\leq d$  then
28:      resultSet  $\leftarrow \text{resultSet} \cup (i, j)$ 
29:    end if
30:     $j \leftarrow j + 1$ 
31:  end while
32: end for
```

Gencluster mit Fehlern

Cyanobacterium glutamicum



Mycobacterium tuberculosis

