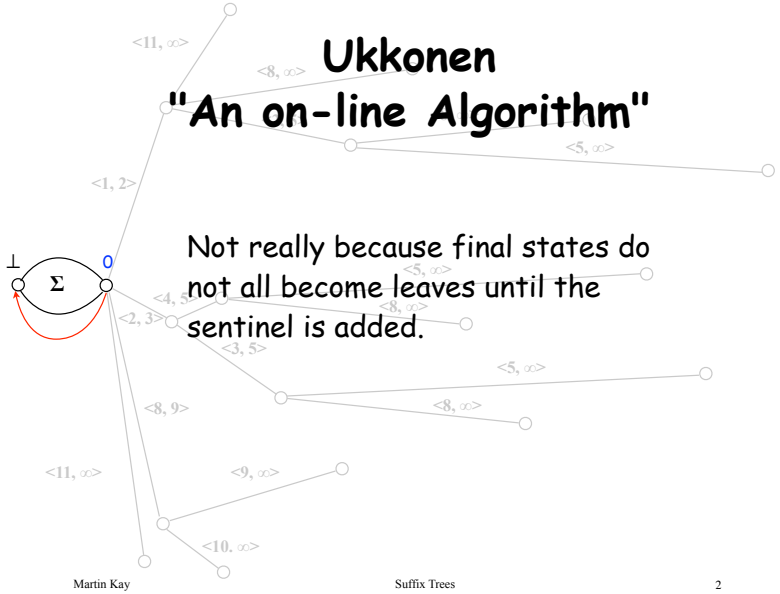
Suffix Trees

Ukkonen's Algorithm

Martin Kay

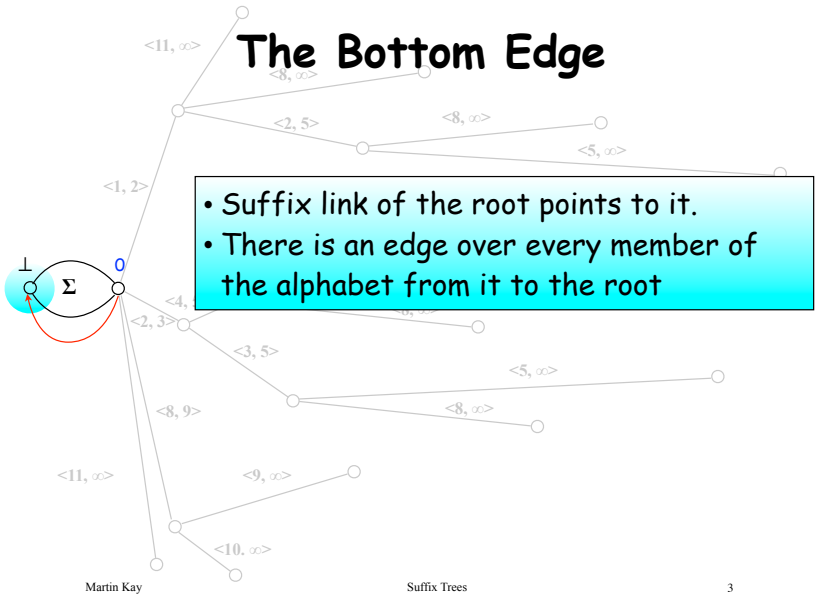
Stanford University
and the University of the Saarland

Ukkonen
"An on-line Algorithm"



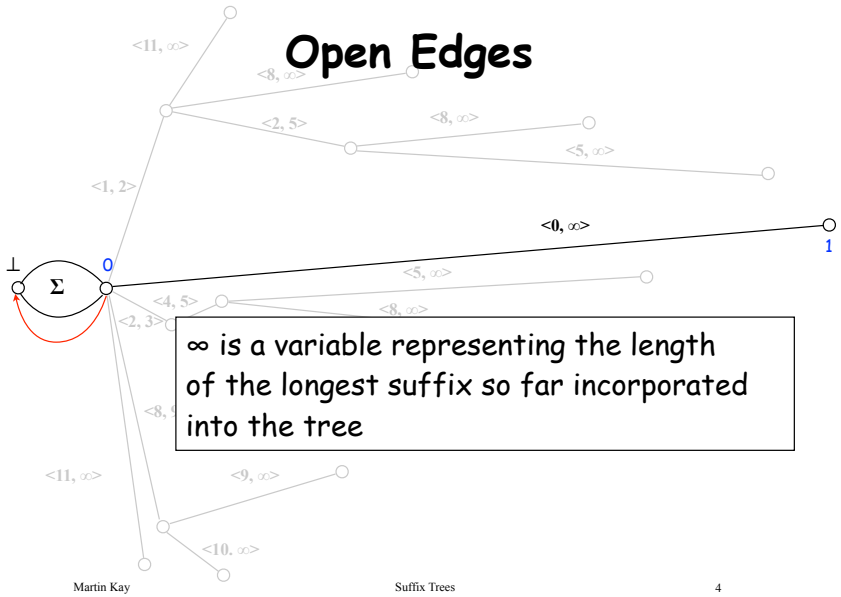
Not really because final states do not all become leaves until the sentinel is added.

The Bottom Edge



- Suffix link of the root points to it.
- There is an edge over every member of the alphabet from it to the root

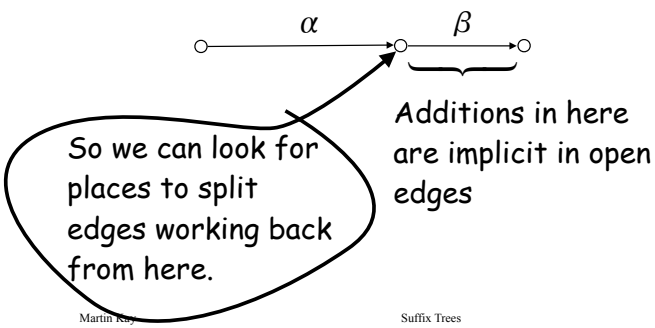
Open Edges



∞ is a variable representing the length of the longest suffix so far incorporated into the tree

The Active Point

The longest string (α) that is in the tree **both** as a suffix and elsewhere.



The End Point

If the current character already begins the label on some edge on the boundary path, then it will begin the label on every subsequent edge on the boundary path because, if x follows aα, it also follows α.

So we can stop looking for places to split once we encounter such a node.

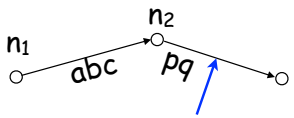
Reference Pairs

References to real and virtual nodes.

(n, s) where n is a real node and s is a string.

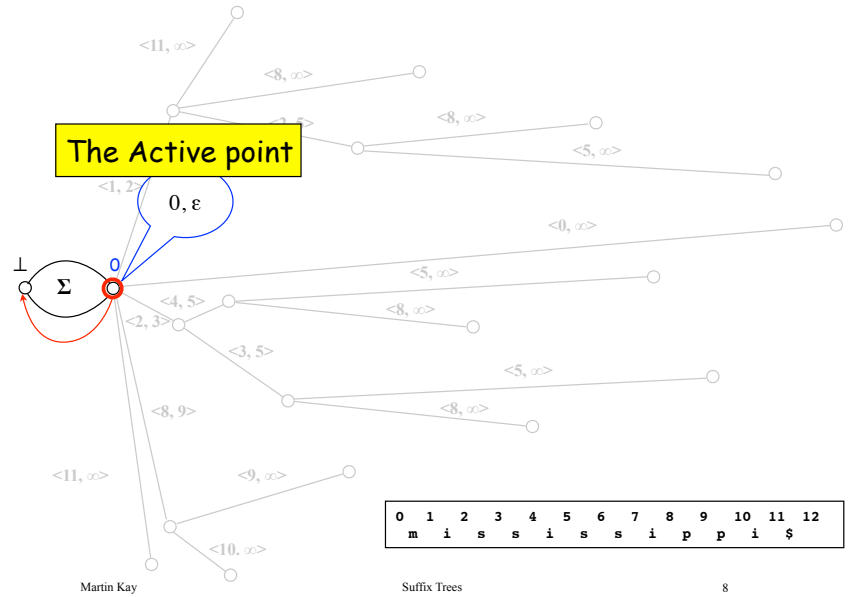
If s = ε, then (n, s) is a real node.

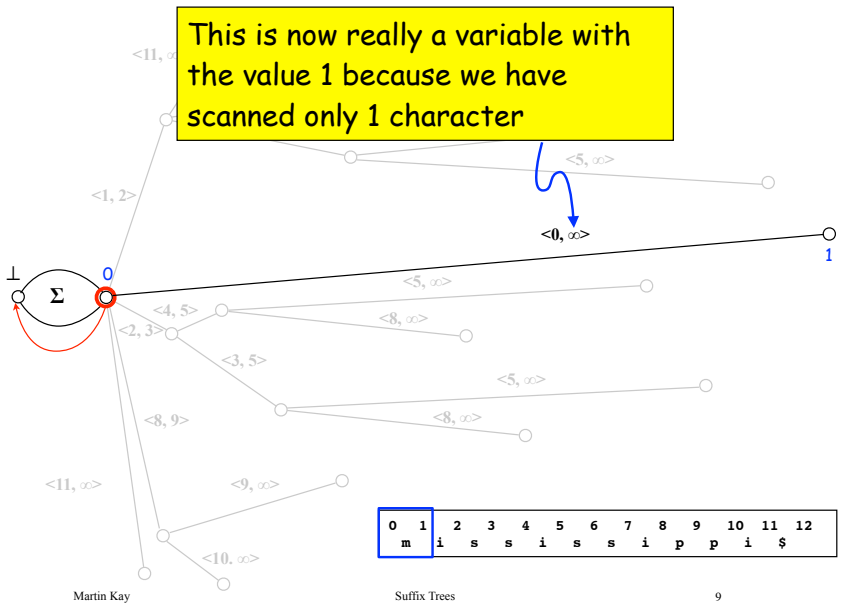
The reverse does not hold:

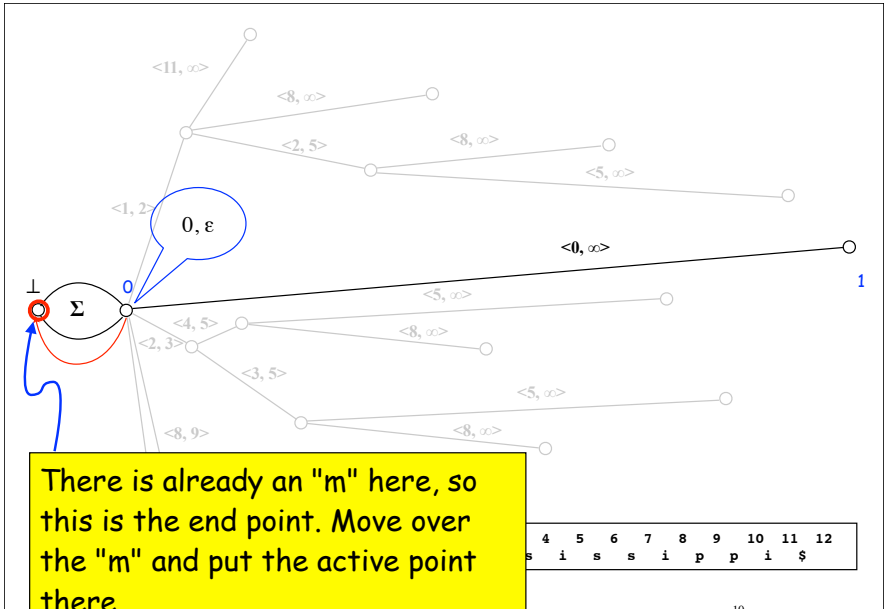


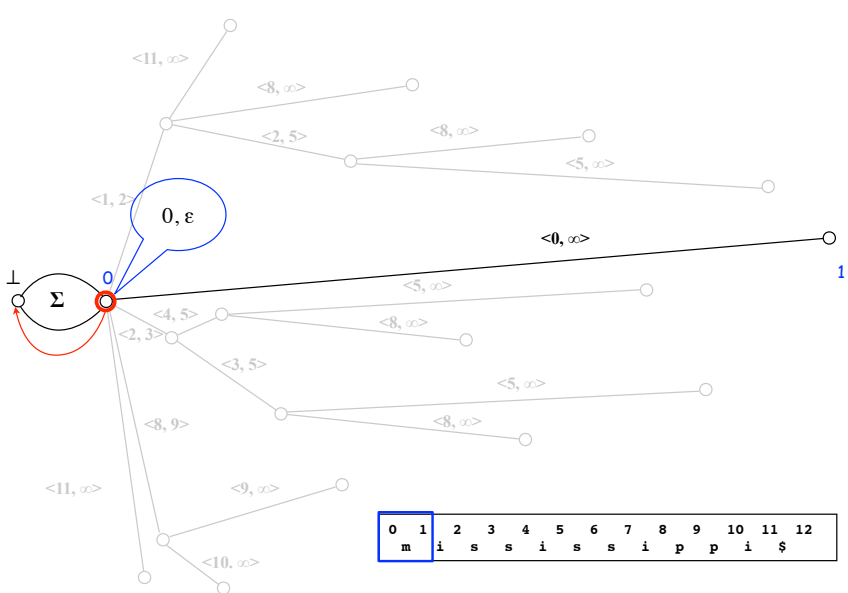
$(n_1\ abcpq) = (n_2\ pq)$

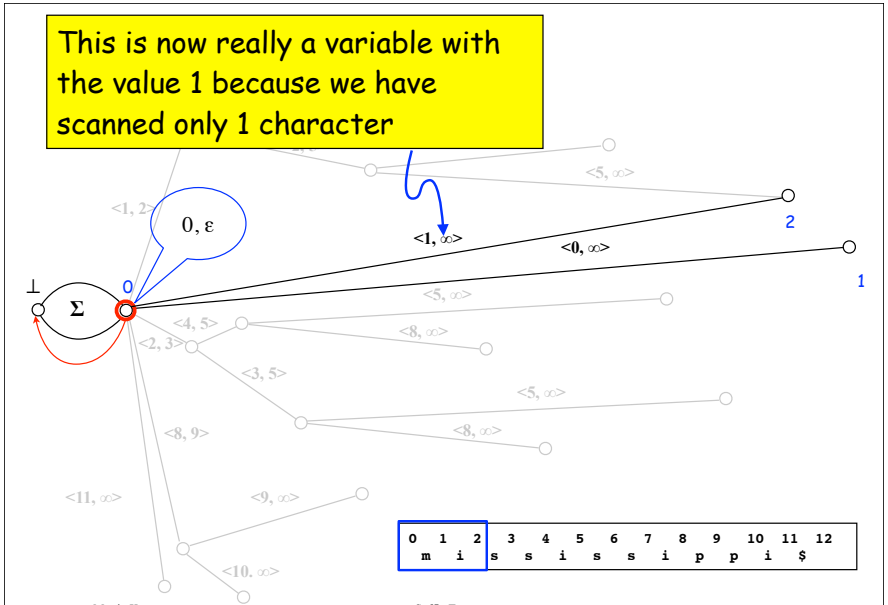
Canonical form: $(n_2\ pq)$ — shortest string

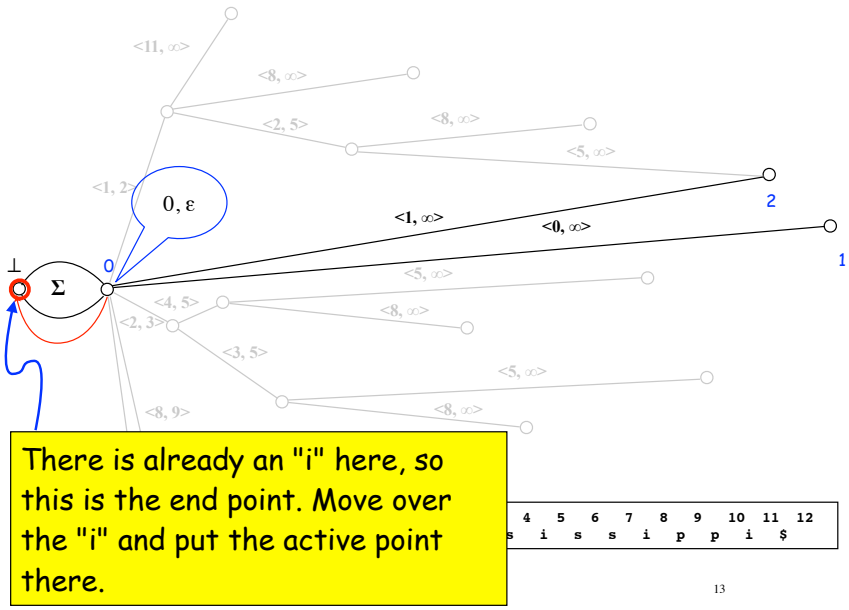



This is now really a variable with the value 1 because we have scanned only 1 character
<11, ∞>
<8, ∞>
<2, 5>
<8, ∞>
<5, ∞>
<1, 2>
<0, ∞>
1
⊥
Σ
0
<5, ∞>
<4, 5>
<2, 3>
<8, ∞>
<3, 5>
<5, ∞>
<8, 9>
<8, ∞>
<11, ∞>
<9, ∞>
<10, ∞>
0 1 2 3 4 5 6 7 8 9 10 11 12
m i s s i s s i p p i $

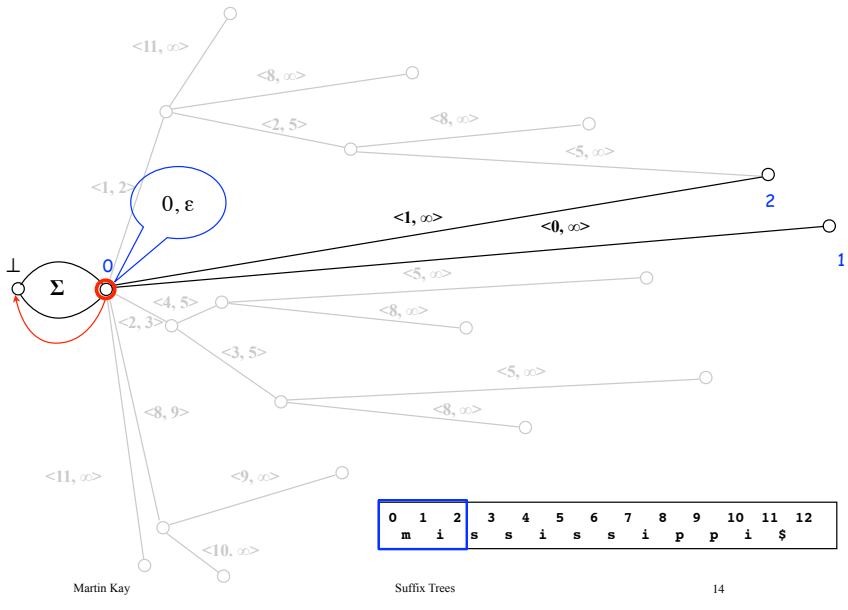

0, ε
<11, ∞>
<8, ∞>
<2, 5>
<8, ∞>
<5, ∞>
<1, 2>
<0, ∞>
1
⊥
Σ
0
<5, ∞>
<4, 5>
<2, 3>
<8, ∞>
<3, 5>
<5, ∞>
<8, 9>
<8, ∞>
There is already an "m" here, so this is the end point. Move over the "m" and put the active point there.
4 5 6 7 8 9 10 11 12
s i s s i p p i $

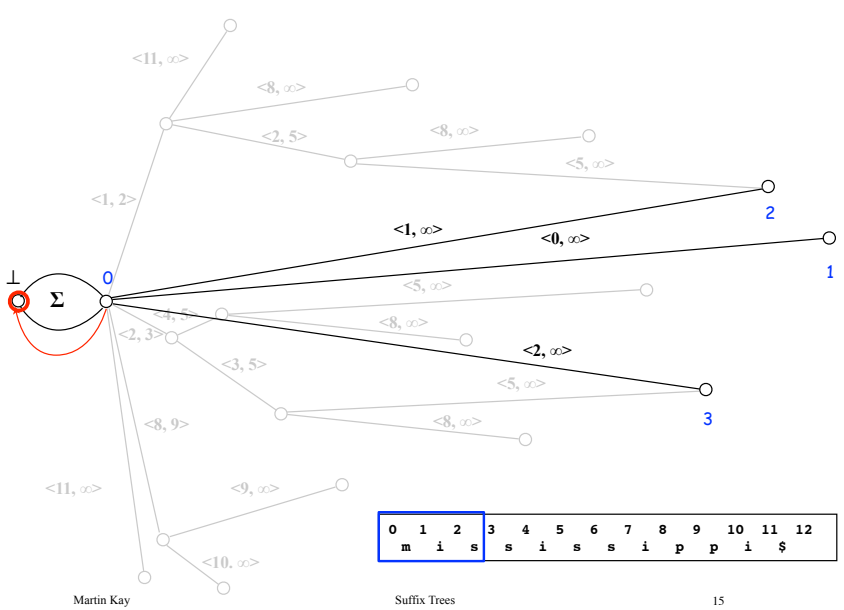

0, ε
<11, ∞>
<8, ∞>
<2, 5>
<8, ∞>
<5, ∞>
<1, 2>
<0, ∞>
1
⊥
Σ
0
<5, ∞>
<4, 5>
<2, 3>
<8, ∞>
<3, 5>
<5, ∞>
<8, 9>
<8, ∞>
<11, ∞>
<9, ∞>
<10, ∞>
0 1 2 3 4 5 6 7 8 9 10 11 12
m i s s i s s i p p i $

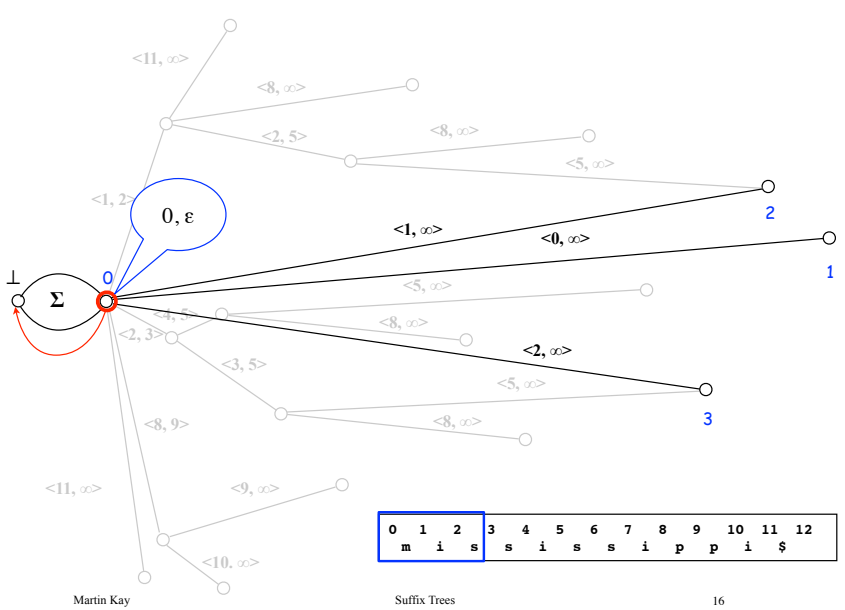

This is now really a variable with the value 1 because we have scanned only 1 character
0, ε
<11, ∞>
<8, ∞>
<2, 5>
<8, ∞>
<5, ∞>
<1, 2>
<1, ∞>
<0, ∞>
2
1
⊥
Σ
0
<5, ∞>
<4, 5>
<2, 3>
<8, ∞>
<3, 5>
<5, ∞>
<8, 9>
<8, ∞>
<11, ∞>
<9, ∞>
<10, ∞>
0 1 2 3 4 5 6 7 8 9 10 11 12
m i s s i s s i p p i $

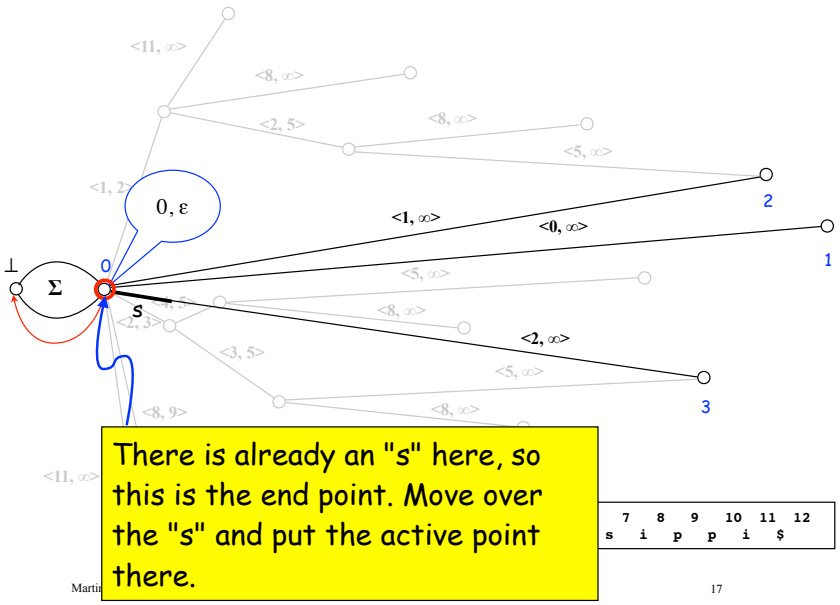


There is already an "i" here, so this is the end point. Move over the "i" and put the active point there.

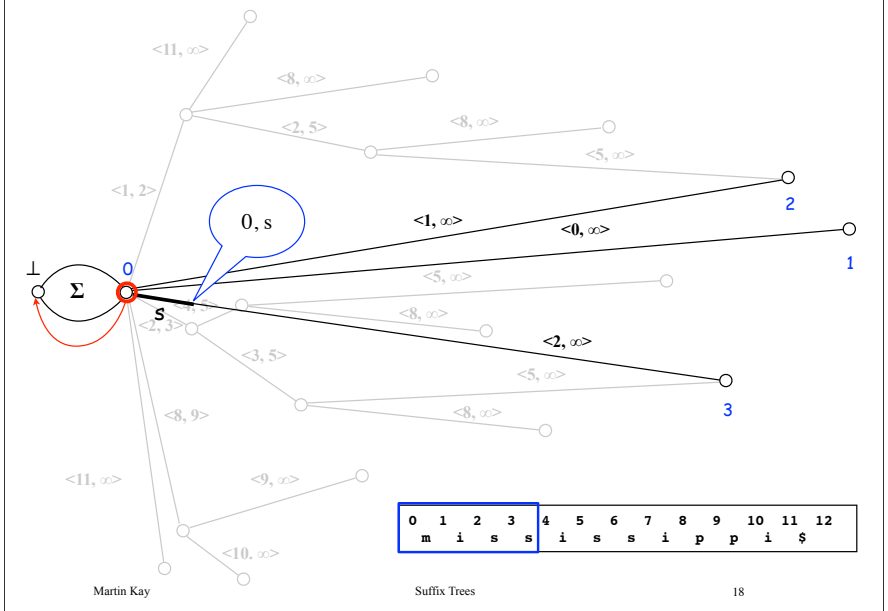


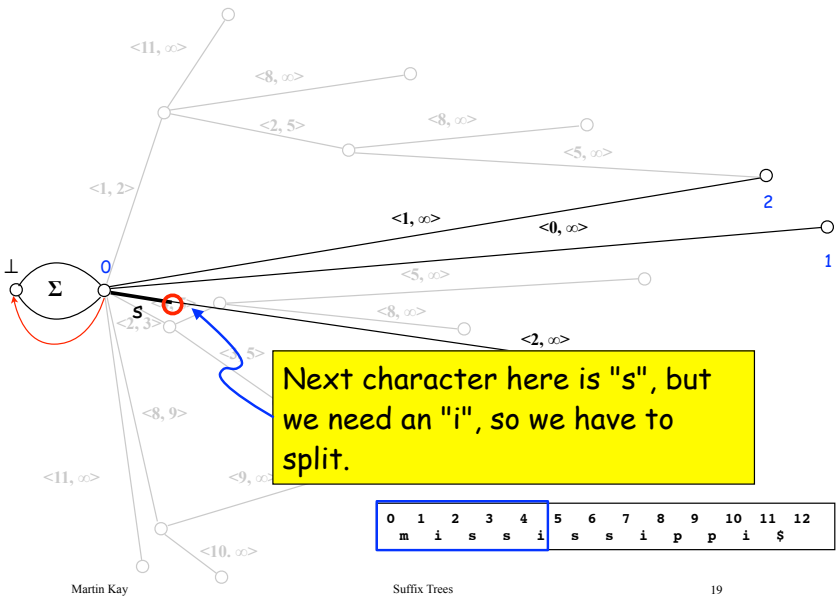


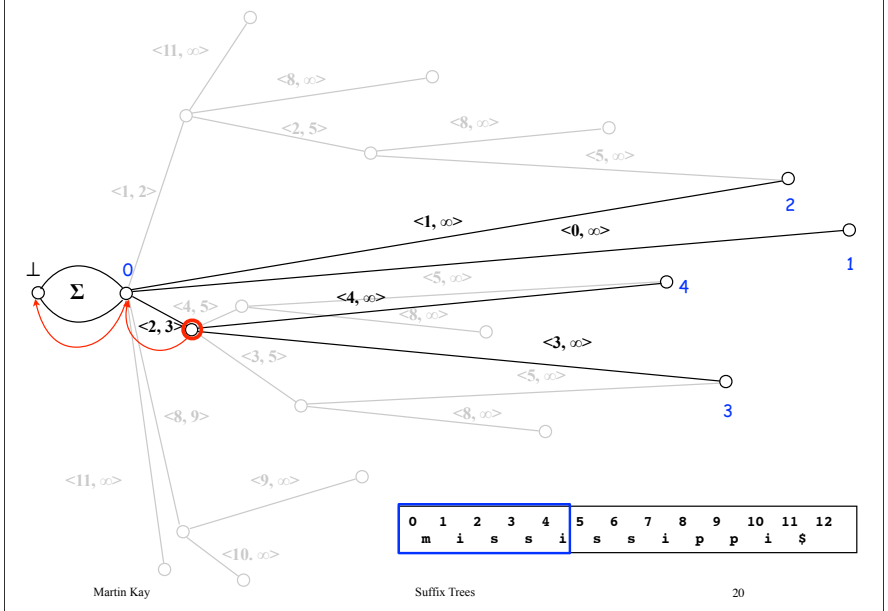


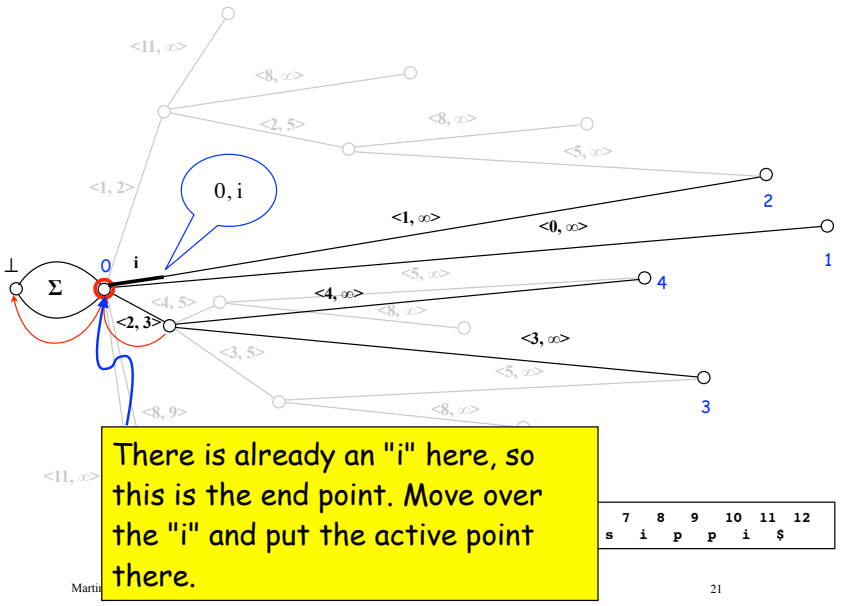


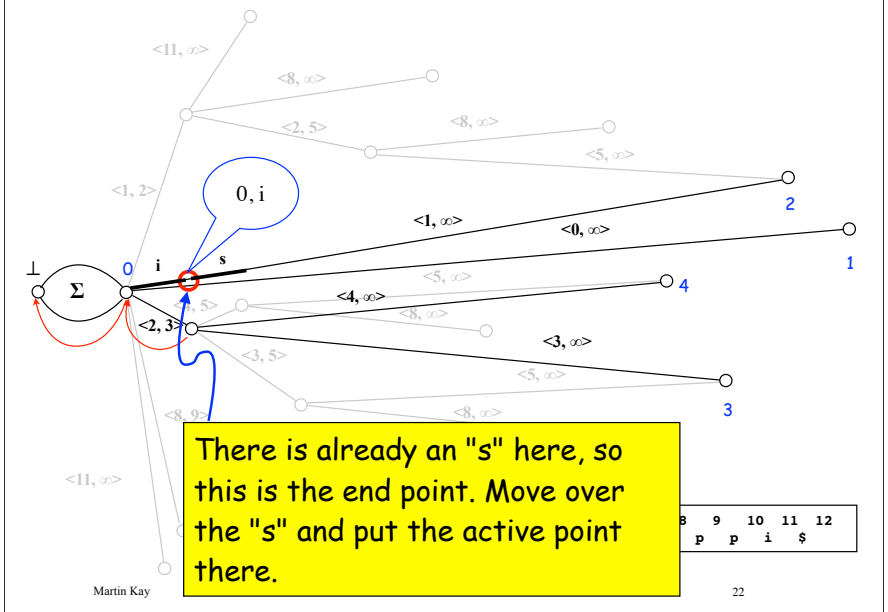
There is already an "s" here, so this is the end point. Move over the "s" and put the active point there.

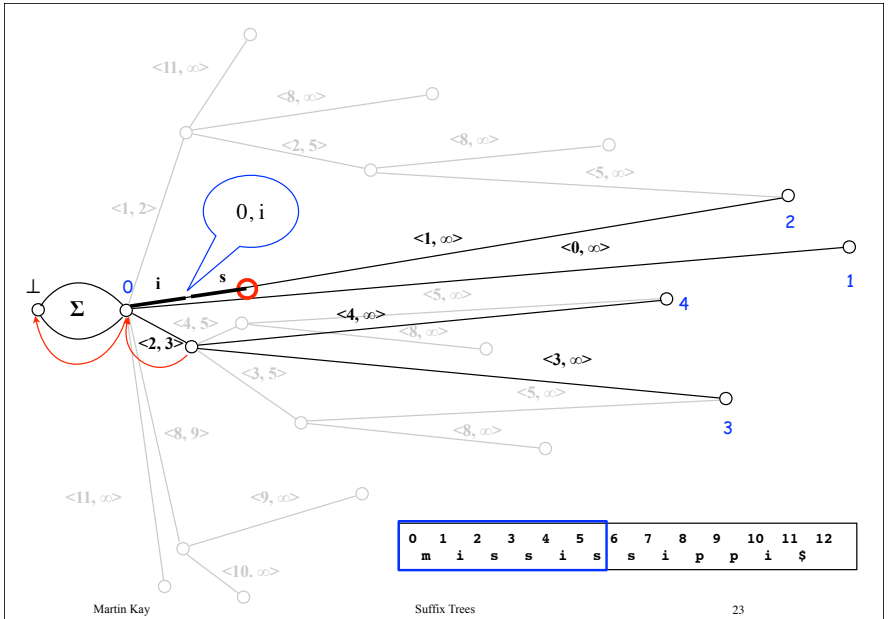


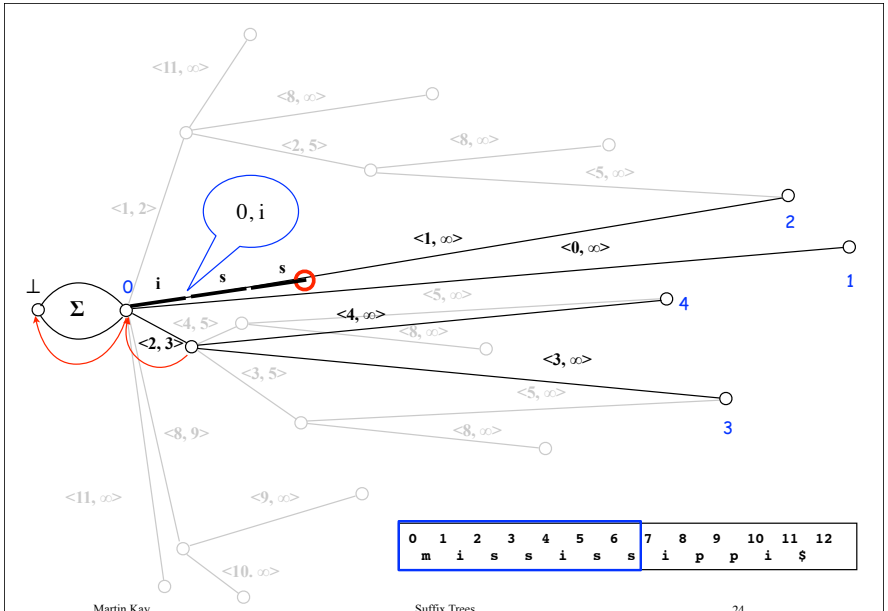


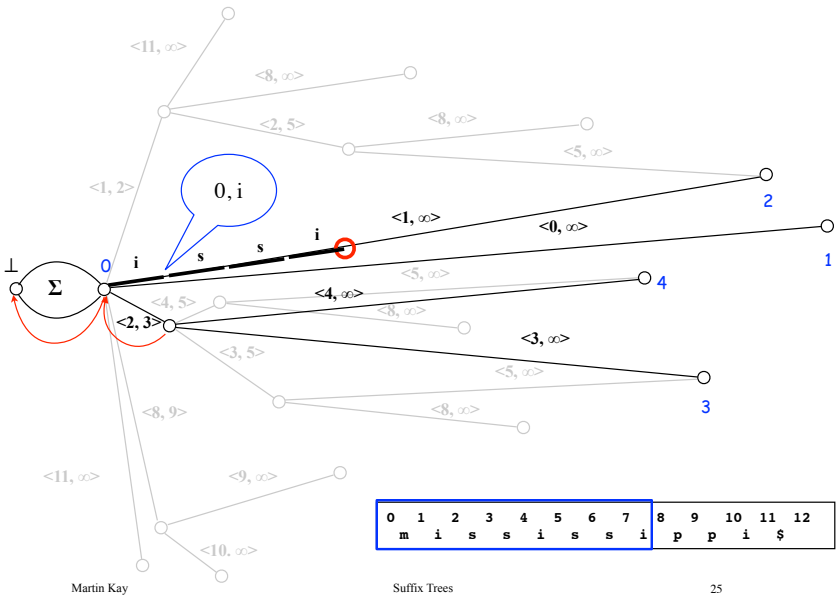


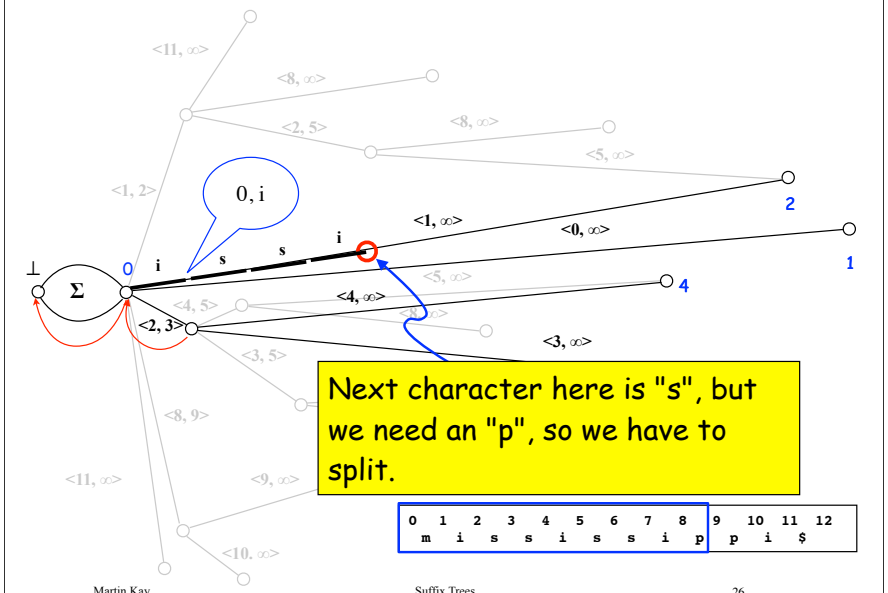

0, i
<1, ∞>
<0, ∞>
<4, ∞>
<2, 3>
<3, ∞>
There is already an "i" here, so this is the end point. Move over the "i" and put the active point there.
7 8 9 10 11 12
s i p p i $

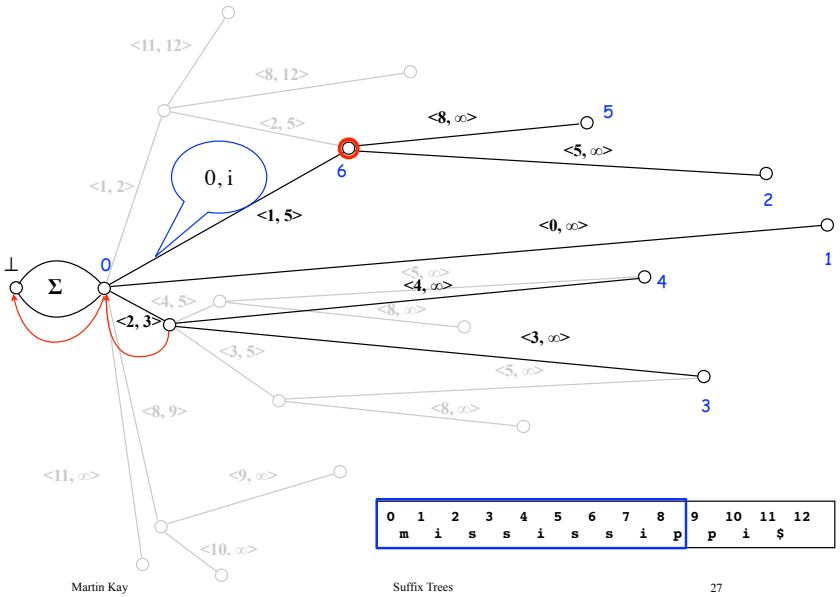

0, i
<1, ∞>
<0, ∞>
<4, ∞>
<2, 3>
<3, ∞>
There is already an "s" here, so this is the end point. Move over the "s" and put the active point there.
8 9 10 11 12
p p i $

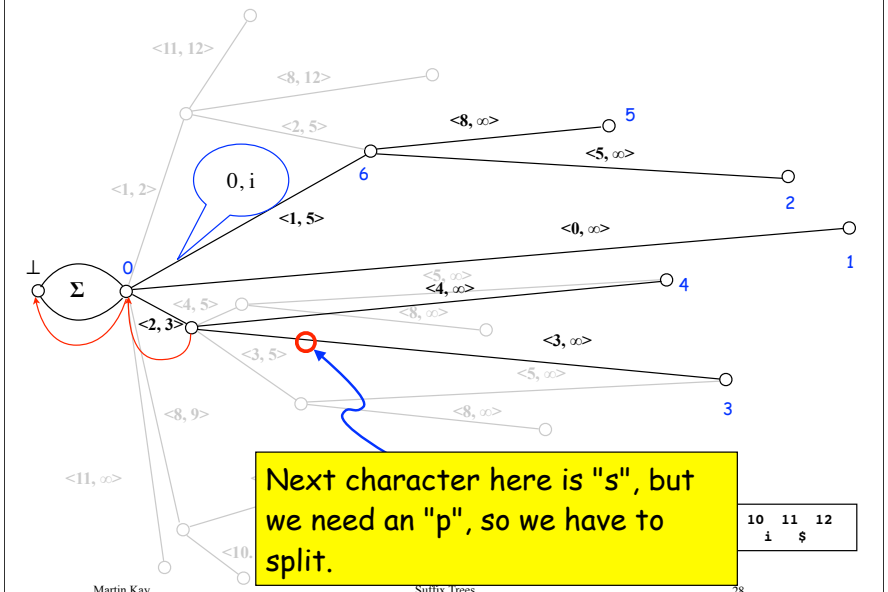

0, i
<1, ∞>
<0, ∞>
<4, ∞>
<2, 3>
<3, ∞>
0 1 2 3 4 5 6 7 8 9 10 11 12
m i s s i s s i p p i $

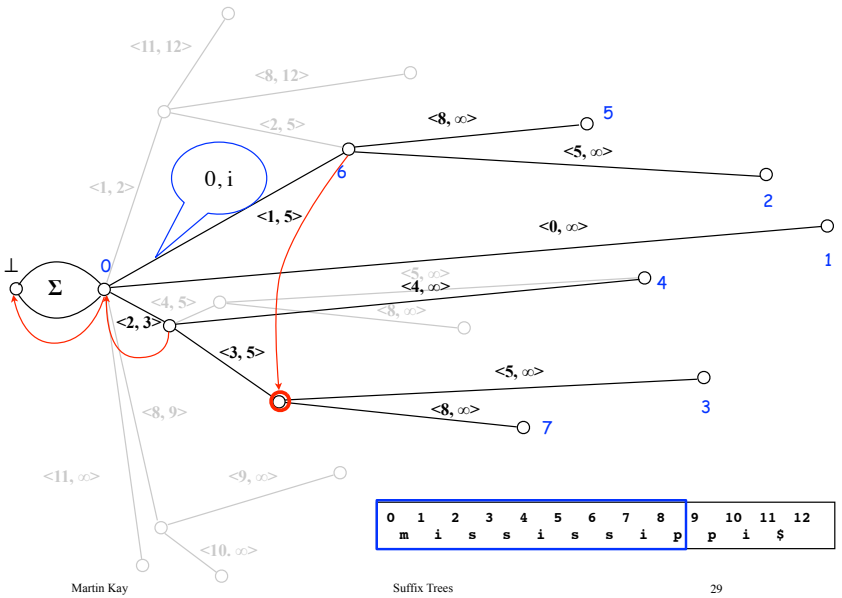

0, i
<1, ∞>
<0, ∞>
<4, ∞>
<2, 3>
<3, ∞>
0 1 2 3 4 5 6 7 8 9 10 11 12
m i s s i s s i p p i $

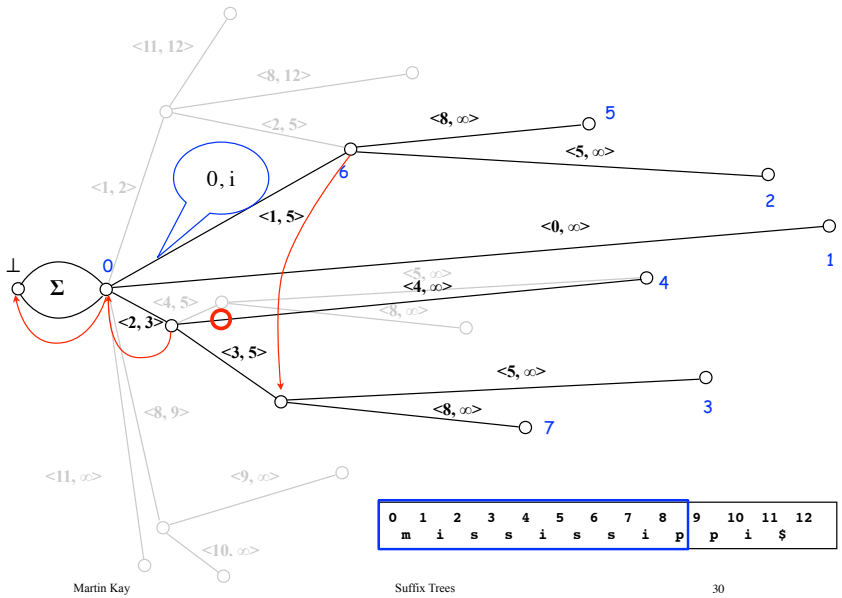

0, i
<11, ∞>
<8, ∞>
<2, 5>
<8, ∞>
<5, ∞>
<1, 2>
<1, ∞>
<0, ∞>
<4, ∞>
<2, 3>
<3, ∞>
<4, 5>
<8, ∞>
<5, ∞>
<3, 5>
<8, 9>
<11, ∞>
<9, ∞>
<10, ∞>
⊥
Σ
0
i
s
s
i
1
2
3
4
0 1 2 3 4 5 6 7 8 9 10 11 12
m i s s i s s i p p i $

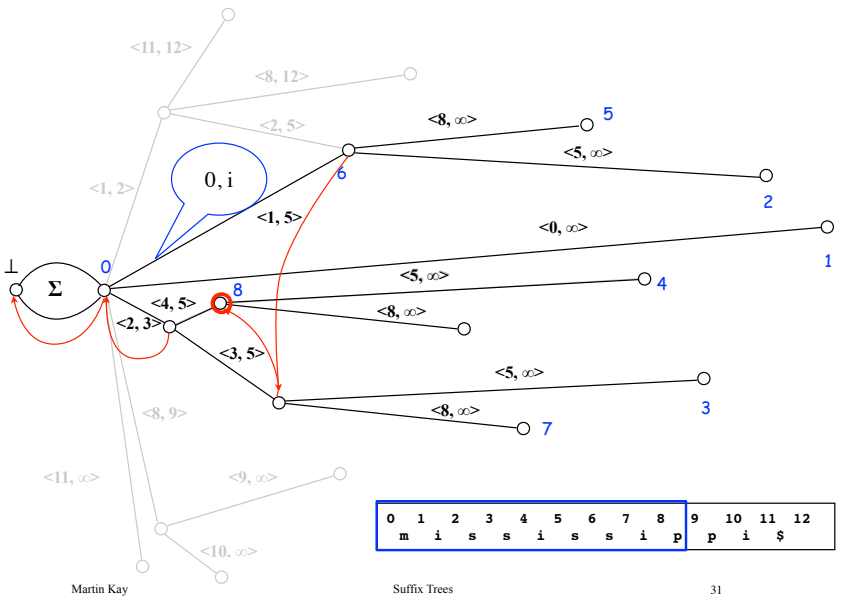

0, i
<11, ∞>
<8, ∞>
<2, 5>
<8, ∞>
<5, ∞>
<1, 2>
<1, ∞>
<0, ∞>
<4, ∞>
<2, 3>
<3, ∞>
<4, 5>
<8, ∞>
<5, ∞>
<3, 5>
<8, 9>
<11, ∞>
<9, ∞>
<10, ∞>
⊥
Σ
0
i
s
s
i
1
2
3
4
Next character here is "s", but we need an "p", so we have to split.
0 1 2 3 4 5 6 7 8 9 10 11 12
m i s s i s s i p p i $

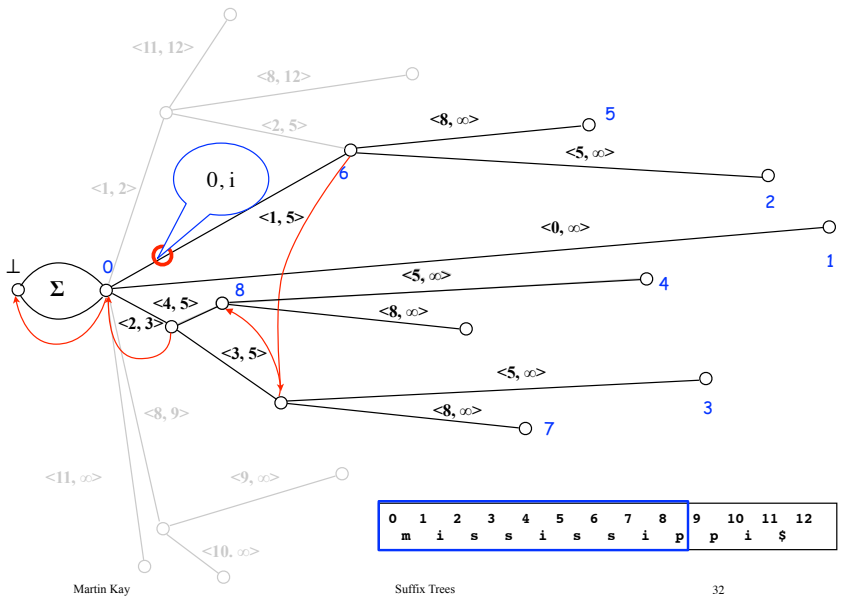

0, i
<11, 12>
<8, 12>
<2, 5>
<8, ∞>
5
<5, ∞>
6
<1, 2>
<1, 5>
<0, ∞>
<4, ∞>
<2, 3>
<3, ∞>
<4, 5>
<8, ∞>
<5, ∞>
<3, 5>
<8, 9>
<11, ∞>
<9, ∞>
<10, ∞>
⊥
Σ
0
1
2
3
4
0 1 2 3 4 5 6 7 8 9 10 11 12
m i s s i s s i p p i $

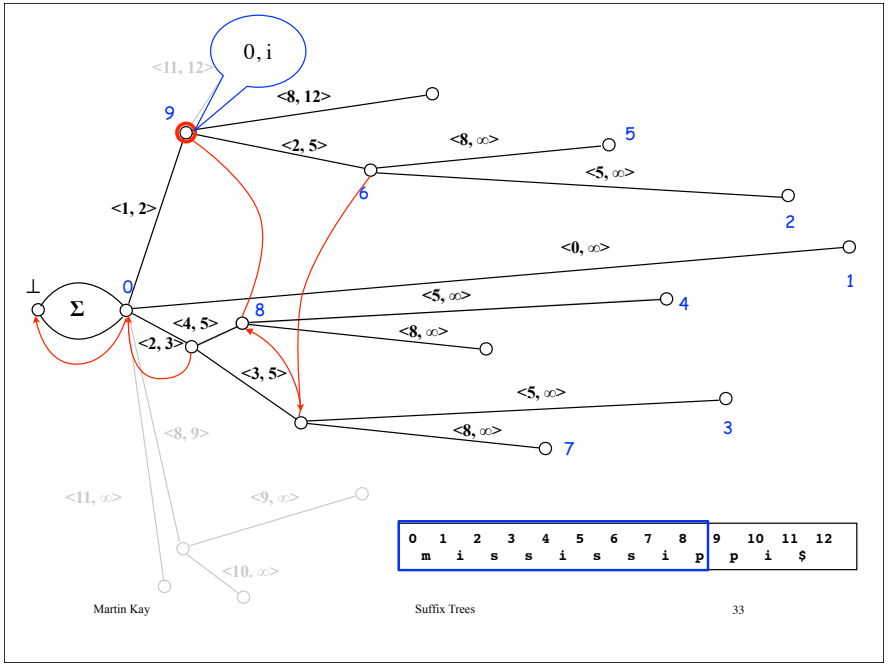

0, i
<11, 12>
<8, 12>
<2, 5>
<8, ∞>
5
<5, ∞>
6
<1, 2>
<1, 5>
<0, ∞>
<4, ∞>
<2, 3>
<3, ∞>
<4, 5>
<8, ∞>
<5, ∞>
<3, 5>
<8, 9>
<11, ∞>
<9, ∞>
<10, ∞>
⊥
Σ
0
1
2
3
4
Next character here is "s", but we need an "p", so we have to split.
10 11 12
i $

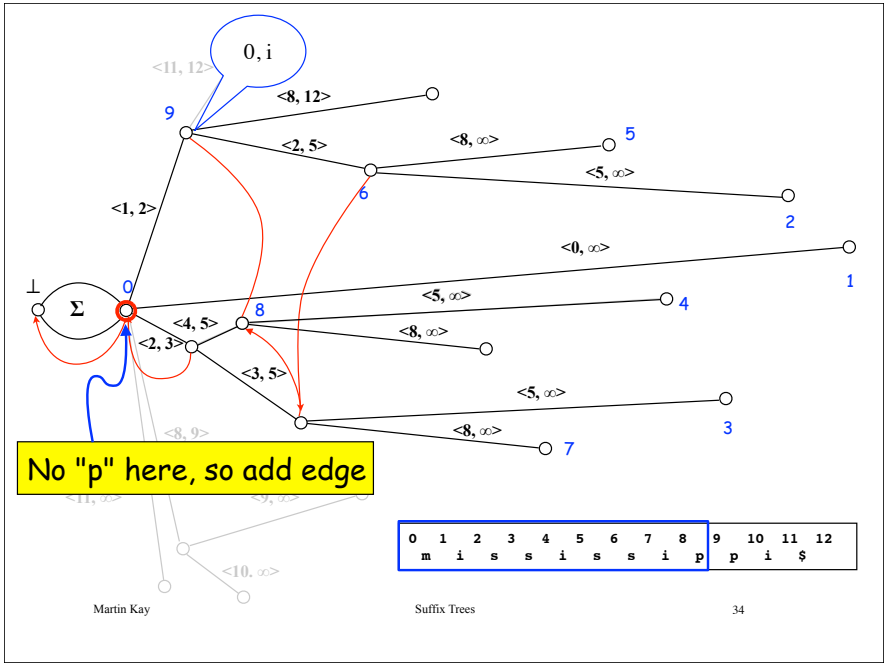

<11, 12>
<8, 12>
<2, 5>
<1, 2>
0, i
<8, ∞>
5
<5, ∞>
6
2
<1, 5>
<0, ∞>
1
⊥
Σ
0
<5, ∞>
<4, ∞>
4
<4, 5>
<8, ∞>
<2, 3>
<3, 5>
<5, ∞>
<8, 9>
<8, ∞>
3
7
<11, ∞>
<9, ∞>
<10, ∞>
0 1 2 3 4 5 6 7 8 9 10 11 12
m i s s i s s i p p i $

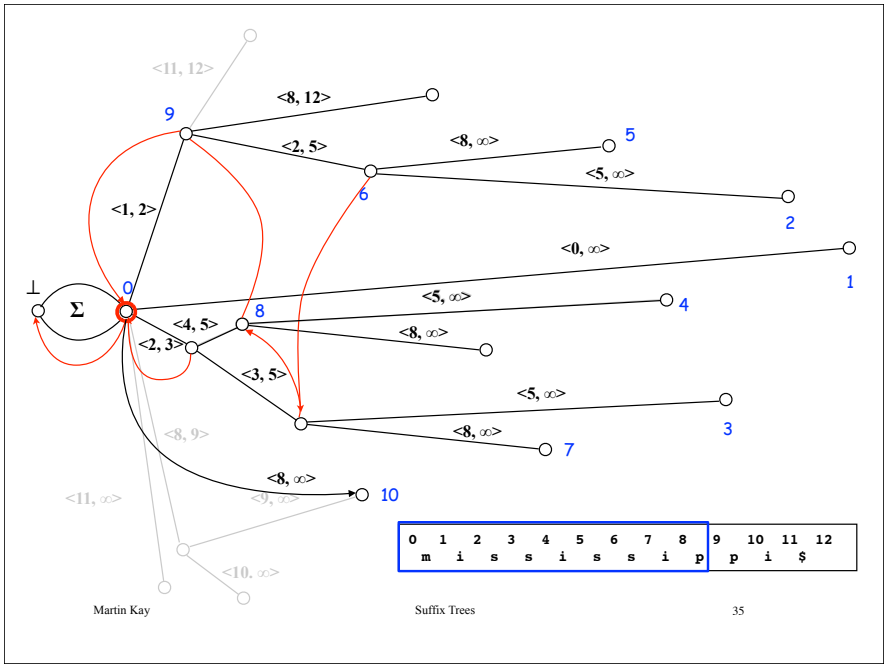

<11, 12>
<8, 12>
<2, 5>
<1, 2>
0, i
<8, ∞>
5
<5, ∞>
6
2
<1, 5>
<0, ∞>
1
⊥
Σ
0
<5, ∞>
<4, ∞>
4
<4, 5>
<8, ∞>
<2, 3>
<3, 5>
<5, ∞>
<8, 9>
<8, ∞>
3
7
<11, ∞>
<9, ∞>
<10, ∞>
0 1 2 3 4 5 6 7 8 9 10 11 12
m i s s i s s i p p i $

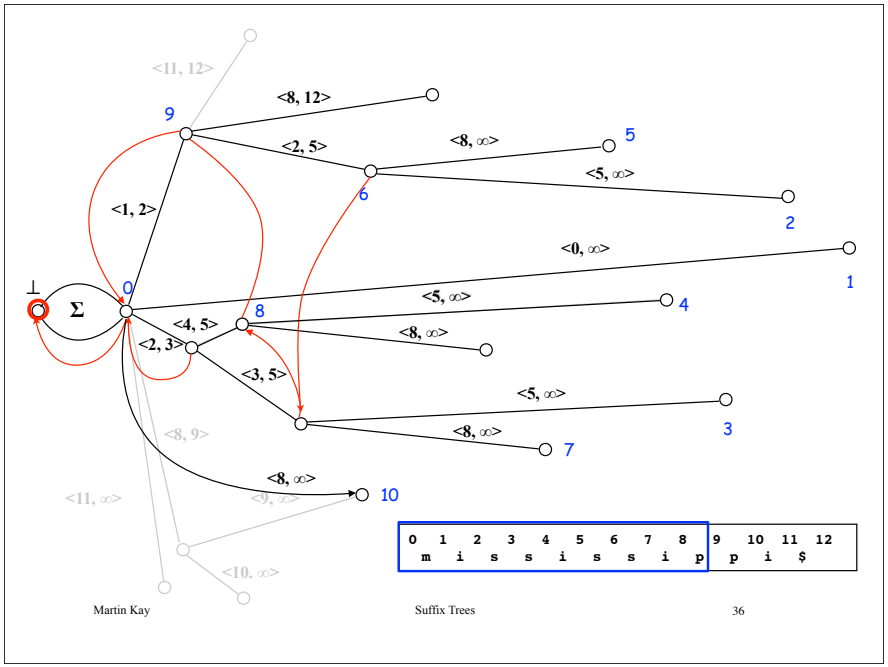

<11, 12>
<8, 12>
<2, 5>
<1, 2>
0, i
<8, ∞>
5
<5, ∞>
6
2
<1, 5>
<0, ∞>
1
⊥
Σ
0
<5, ∞>
8
4
<4, 5>
<8, ∞>
<2, 3>
<3, 5>
<5, ∞>
<8, 9>
<8, ∞>
3
7
<11, ∞>
<9, ∞>
<10, ∞>
0 1 2 3 4 5 6 7 8 9 10 11 12
m i s s i s s i p p i $

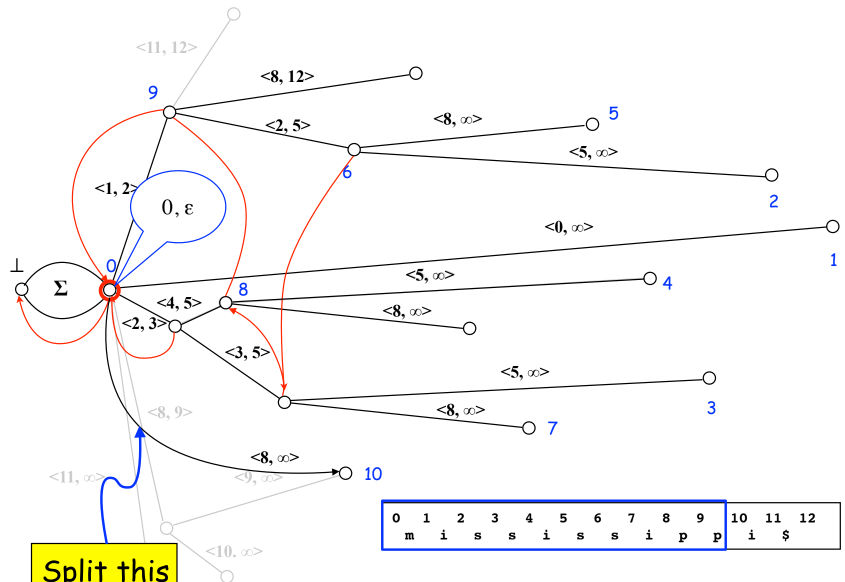

<11, 12>
<8, 12>
<2, 5>
<1, 2>
0, i
<8, ∞>
5
<5, ∞>
6
2
<1, 5>
<0, ∞>
1
⊥
Σ
0
<5, ∞>
8
4
<4, 5>
<8, ∞>
<2, 3>
<3, 5>
<5, ∞>
<8, 9>
<8, ∞>
3
7
<11, ∞>
<9, ∞>
<10, ∞>
0 1 2 3 4 5 6 7 8 9 10 11 12
m i s s i s s i p p i $


0, i
<11, 12>
<8, 12>
9
<2, 5>
<8, ∞>
5
<5, ∞>
6
2
<1, 2>
<0, ∞>
1
⊥
Σ
0
<5, ∞>
8
4
<4, 5>
<2, 3>
<8, ∞>
<3, 5>
<5, ∞>
3
<8, 9>
<8, ∞>
7
<11, ∞>
<9, ∞>
<10, ∞>
0 1 2 3 4 5 6 7 8 9 10 11 12
m i s s i s s i p p i $

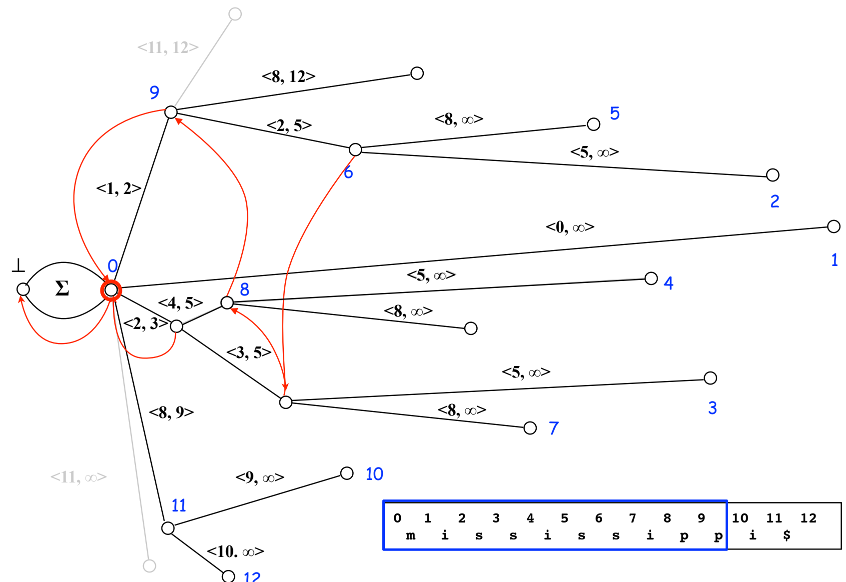

0, i
<11, 12>
<8, 12>
9
<2, 5>
<8, ∞>
5
<5, ∞>
6
2
<1, 2>
<0, ∞>
1
⊥
Σ
0
<5, ∞>
8
4
<4, 5>
<2, 3>
<8, ∞>
<3, 5>
<5, ∞>
3
<8, 9>
<8, ∞>
7
No "p" here, so add edge
<11, ∞>
<9, ∞>
<10, ∞>
0 1 2 3 4 5 6 7 8 9 10 11 12
m i s s i s s i p p i $


<11, 12>
<8, 12>
9
<2, 5>
<8, ∞>
5
<5, ∞>
6
2
<1, 2>
<0, ∞>
1
⊥
Σ
0
<5, ∞>
8
4
<4, 5>
<2, 3>
<8, ∞>
<3, 5>
<5, ∞>
3
<8, 9>
<8, ∞>
7
<8, ∞>
10
<11, ∞>
<9, ∞>
<10, ∞>
0 1 2 3 4 5 6 7 8 9 10 11 12
m i s s i s s i p p i $

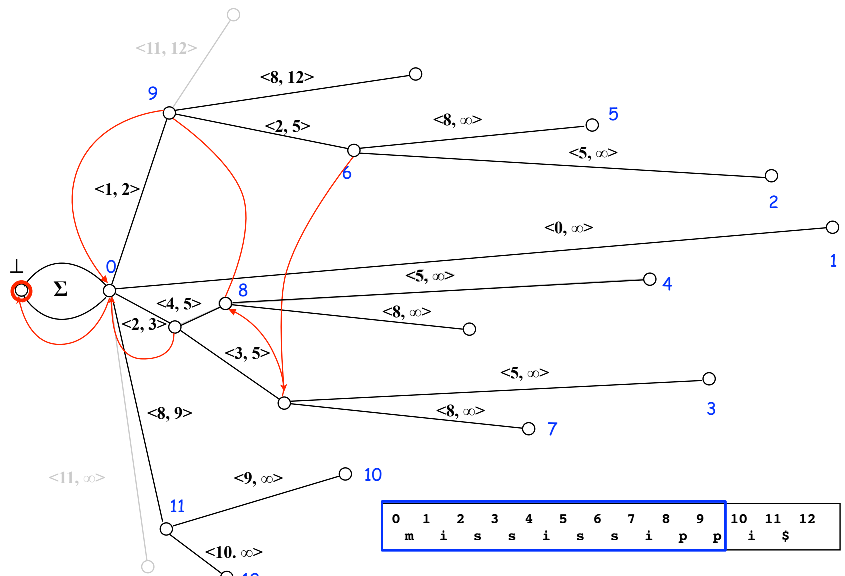

<11, 12>
<8, 12>
9
<2, 5>
<8, ∞>
5
<5, ∞>
6
2
<1, 2>
<0, ∞>
1
⊥
Σ
0
<5, ∞>
8
4
<4, 5>
<2, 3>
<8, ∞>
<3, 5>
<5, ∞>
3
<8, 9>
<8, ∞>
7
<8, ∞>
10
<11, ∞>
<9, ∞>
<10, ∞>
0 1 2 3 4 5 6 7 8 9 10 11 12
m i s s i s s i p p i $



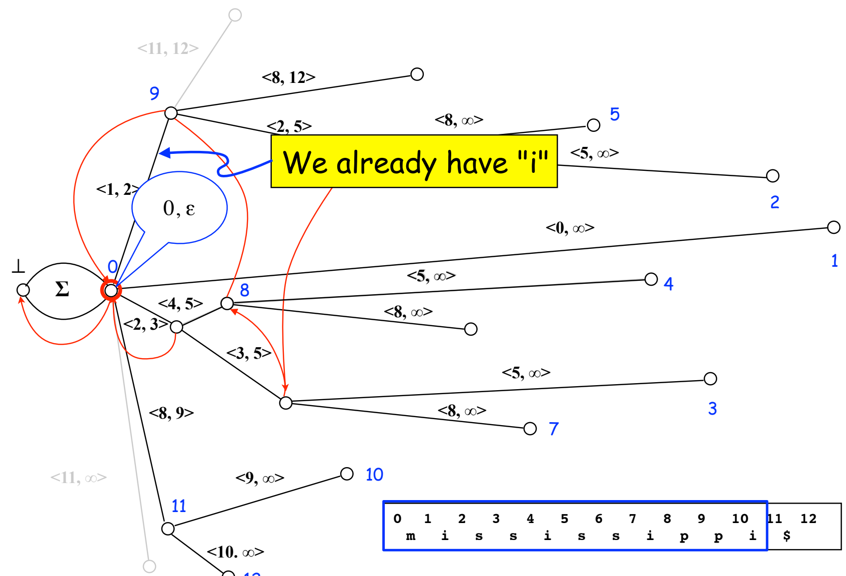
<11, 12>
<8, 12>
<2, 5>
<8, ∞>
<5, ∞>
<1, 2>
0, ε
<0, ∞>
<5, ∞>
<4, 5>
<2, 3>
<8, ∞>
<3, 5>
<5, ∞>
<8, 9>
<8, ∞>
<8, ∞>
<11, ∞>
<9, ∞>
<10, ∞>
Split this
0 1 2 3 4 5 6 7 8 9 10 11 12
m i s s i s s i p p i $



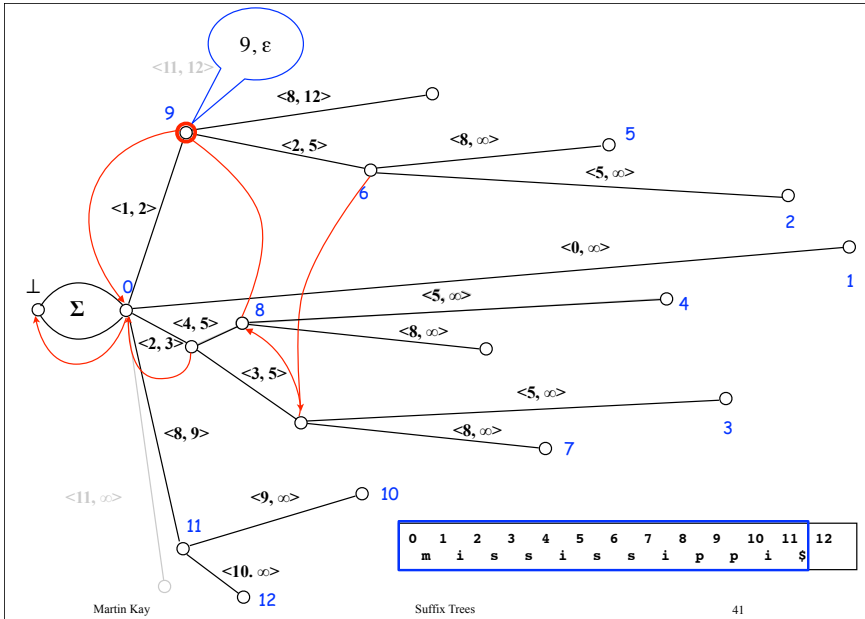
<11, 12>
<8, 12>
<2, 5>
<8, ∞>
<5, ∞>
<1, 2>
<0, ∞>
<5, ∞>
<4, 5>
<2, 3>
<8, ∞>
<3, 5>
<5, ∞>
<8, 9>
<8, ∞>
<11, ∞>
<9, ∞>
<10. ∞>
0 1 2 3 4 5 6 7 8 9 10 11 12
m i s s i s s i p p i $

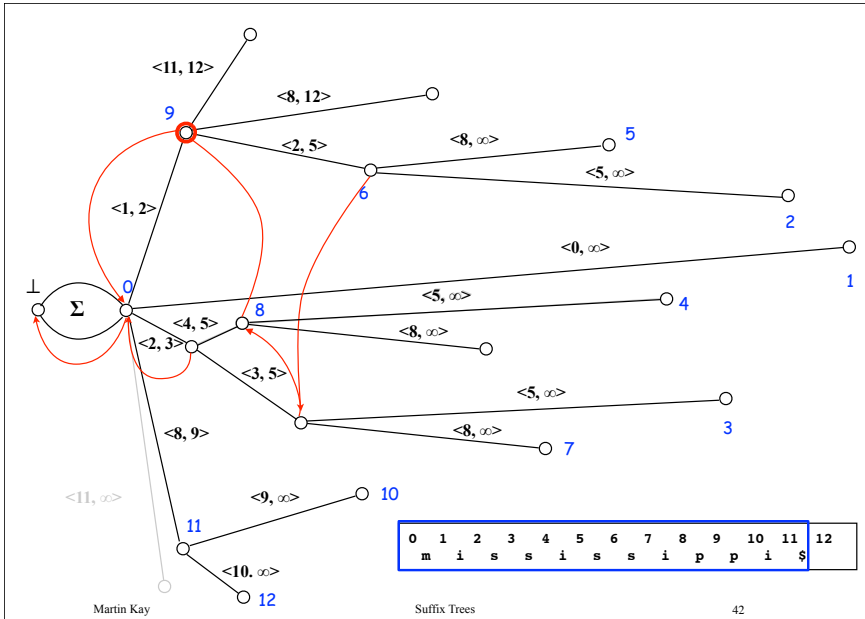


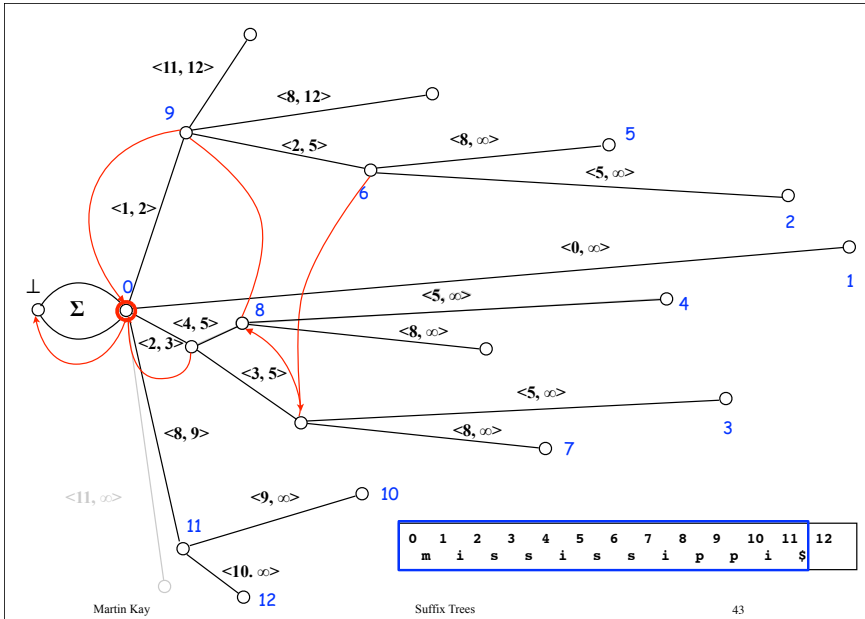
<11, 12>
<8, 12>
<2, 5>
<8, ∞>
<5, ∞>
<1, 2>
<0, ∞>
<5, ∞>
<4, 5>
<2, 3>
<8, ∞>
<3, 5>
<5, ∞>
<8, 9>
<8, ∞>
<11, ∞>
<9, ∞>
<10. ∞>
0 1 2 3 4 5 6 7 8 9 10 11 12
m i s s i s s i p p i $

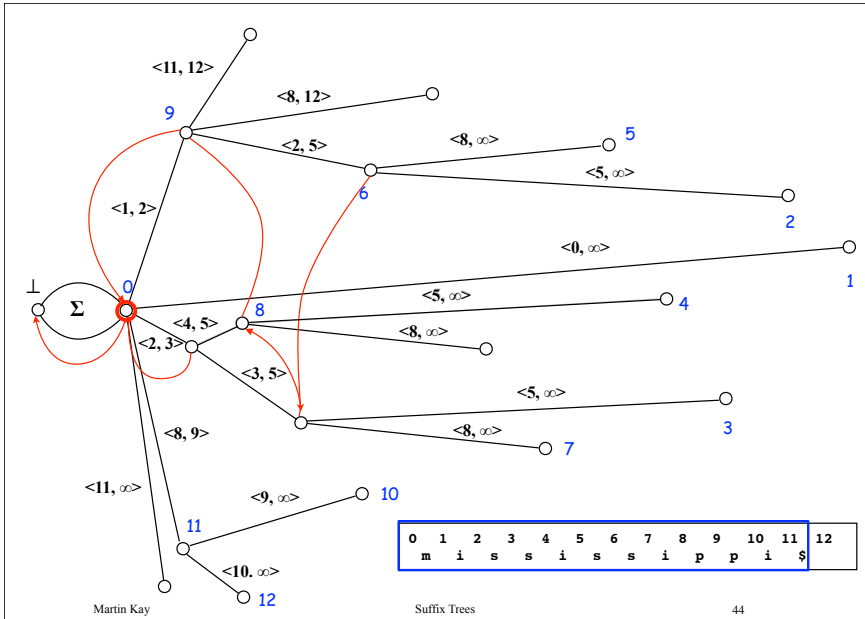


<11, 12>
<8, 12>
<2, 5>
<8, ∞>
We already have "i"
<5, ∞>
<1, 2>
0, ε
<0, ∞>
<5, ∞>
<4, 5>
<2, 3>
<8, ∞>
<3, 5>
<5, ∞>
<8, 9>
<8, ∞>
<11, ∞>
<9, ∞>
<10. ∞>
0 1 2 3 4 5 6 7 8 9 10 11 12
m i s s i s s i p p i $


9, ε
<11, 12>
<8, 12>
9
<2, 5>
<8, ∞>
5
<5, ∞>
6
2
<1, 2>
<0, ∞>
1
⊥
0
Σ
<5, ∞>
8
4
<4, 5>
<2, 3>
<8, ∞>
<3, 5>
<5, ∞>
3
<8, 9>
<8, ∞>
7
<11, ∞>
<9, ∞>
10
11
<10. ∞>
12
0 1 2 3 4 5 6 7 8 9 10 11 12
m i s s i s s i p p i $


<11, 12>
<8, 12>
9
<2, 5>
<8, ∞>
5
<5, ∞>
6
2
<1, 2>
<0, ∞>
1
⊥
0
Σ
<5, ∞>
8
4
<4, 5>
<2, 3>
<8, ∞>
<3, 5>
<5, ∞>
3
<8, 9>
<8, ∞>
7
<11, ∞>
<9, ∞>
10
11
<10. ∞>
12
0 1 2 3 4 5 6 7 8 9 10 11 12
m i s s i s s i p p i $


<11, 12>
<8, 12>
9
<2, 5>
<8, ∞>
5
<5, ∞>
6
2
<1, 2>
<0, ∞>
1
⊥
0
Σ
<5, ∞>
8
4
<4, 5>
<2, 3>
<8, ∞>
<3, 5>
<5, ∞>
3
<8, 9>
<8, ∞>
7
<11, ∞>
<9, ∞>
10
11
<10. ∞>
12
0 1 2 3 4 5 6 7 8 9 10 11 12
m i s s i s s i p p i $


<11, 12>
<8, 12>
9
<2, 5>
<8, ∞>
5
<5, ∞>
6
2
<1, 2>
<0, ∞>
1
⊥
0
Σ
<5, ∞>
8
4
<4, 5>
<2, 3>
<8, ∞>
<3, 5>
<5, ∞>
3
<8, 9>
<8, ∞>
7
<11, ∞>
<9, ∞>
10
11
<10. ∞>
12
0 1 2 3 4 5 6 7 8 9 10 11 12
m i s s i s s i p p i $