

# Algorithm of Suffix Tree Construction

by farseerfc@gmail.com

Jiachen Yang<sup>1</sup>

<sup>1</sup>Research Student in Osaka University

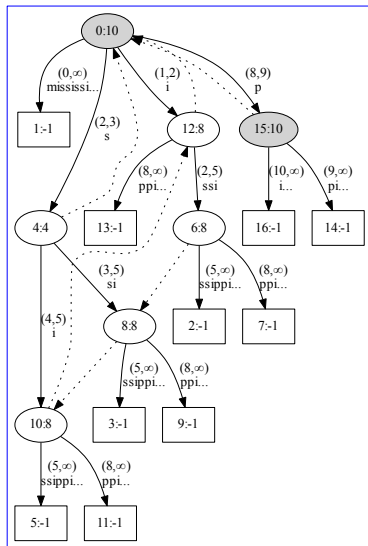
October 14, 2011

# Part Outline

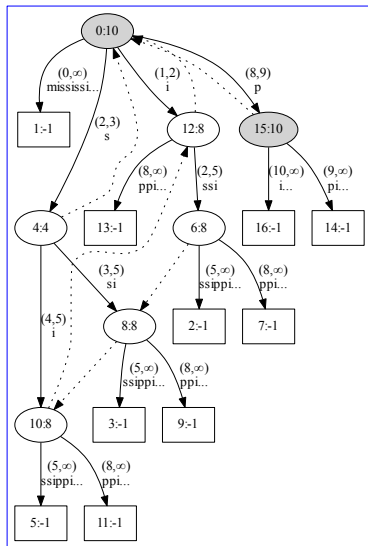
1 What I did

2 Future works

# Algorithm of Constructing Suffix Tree

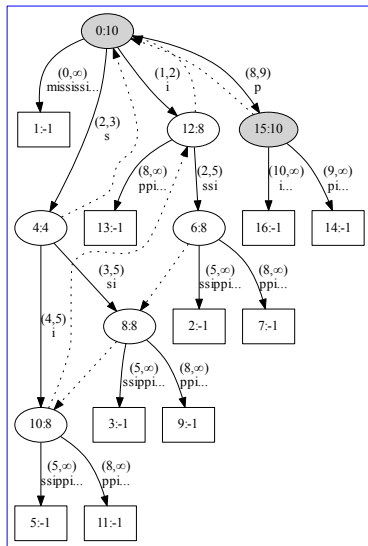


# Algorithm of Constructing Suffix Tree



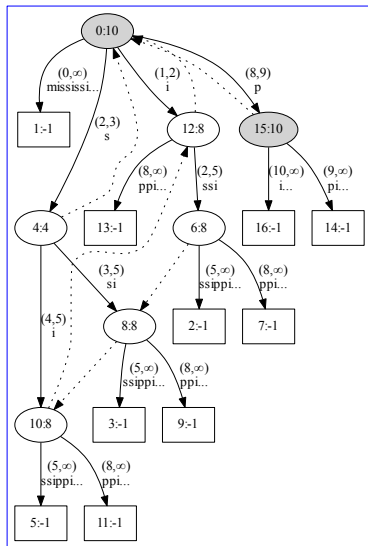
Node (id:generation)



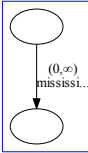
# Algorithm of Constructing Suffix Tree



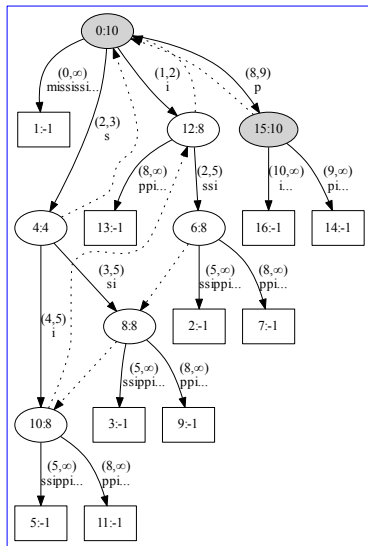
- Node (id:generation)
- Leaf

# Algorithm of Constructing Suffix Tree



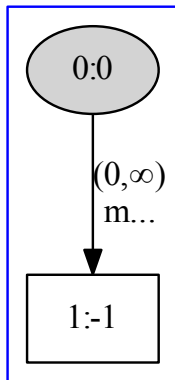
-  Node (id:generation)
-  Leaf
-  Child-relation

# Algorithm of Constructing Suffix Tree



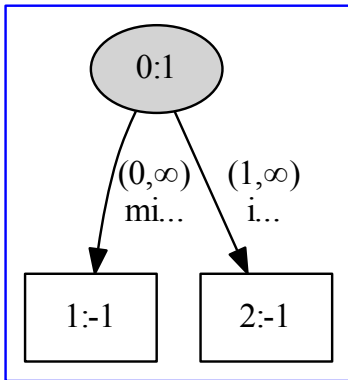
- Node (id:generation)
- Leaf
- Child-relation
- suffix-link relation

# Experiment – mississippi

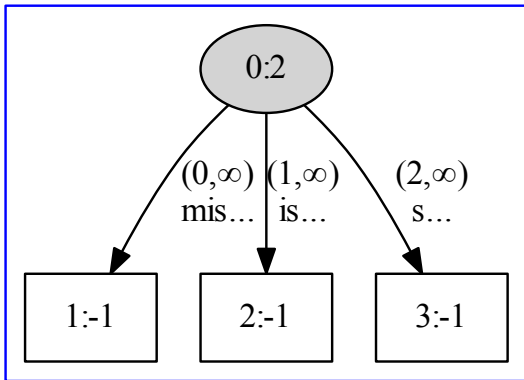




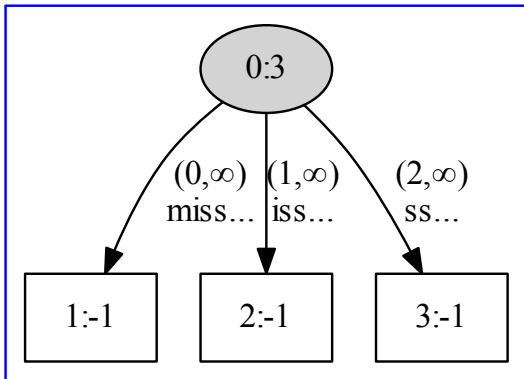
# Experiment – mississippi



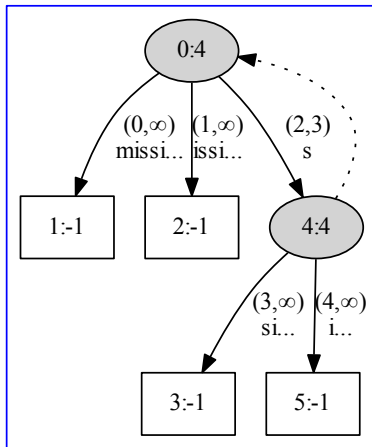
# Experiment – mississippi



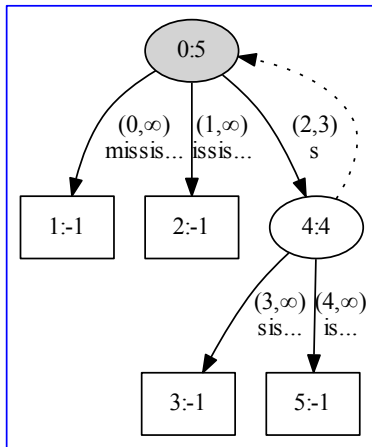
# Experiment – mississippi



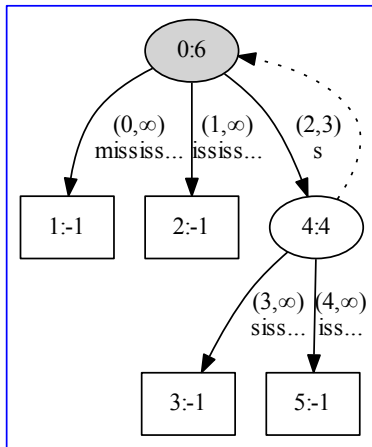
# Experiment – mississippi



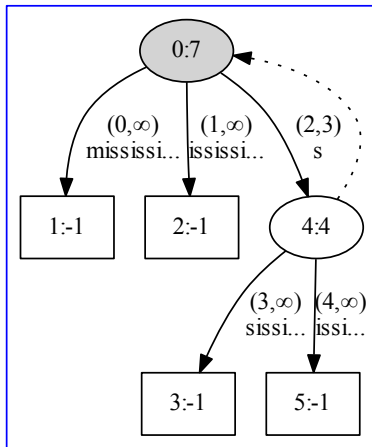
# Experiment – mississippi



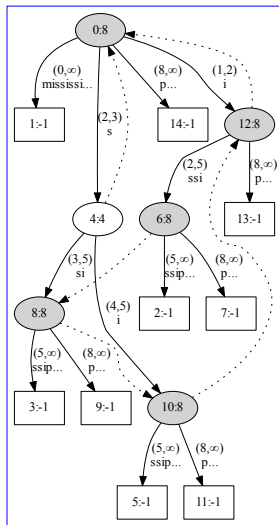
# Experiment – mississippi



# Experiment – mississippi

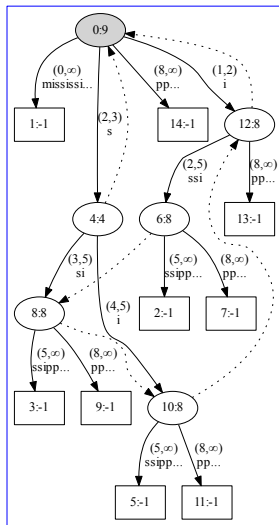


# Experiment – mississippi

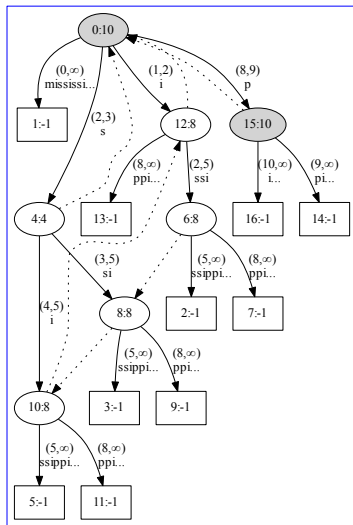




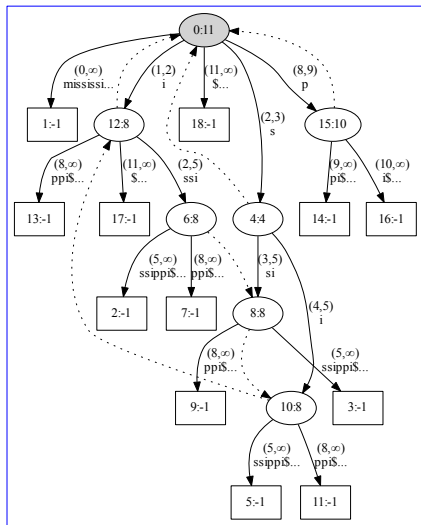
# Experiment – mississippi



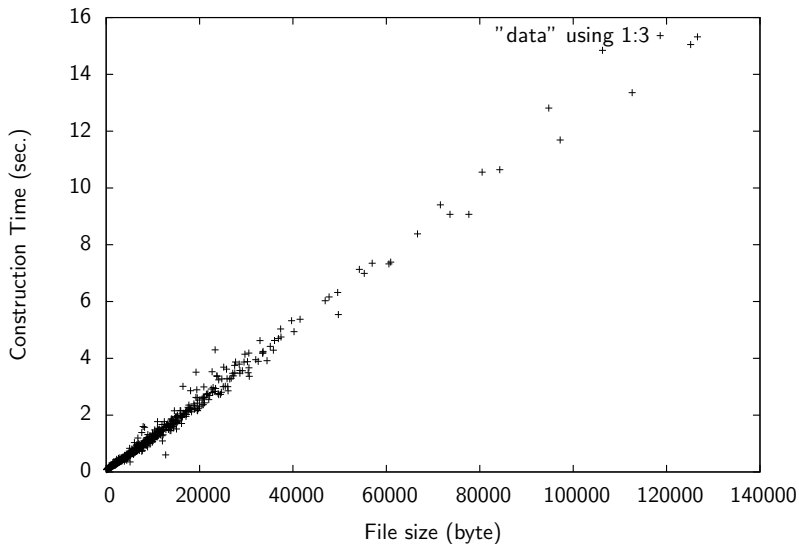
# Experiment – mississippi



# Experiment – mississippi



# Experiment – English text



# Part Outline

1 What I did

2 Future works

# Works in next week

- Implementing Longest Common Substring algorithm using Generalised Suffix Tree based on Ukk's algorithm.

# Works in next week

- Implementing Longest Common Substring algorithm using Generalised Suffix Tree based on Ukk's algorithm.
- Implementing Tokenization on Python codes.

# Works in next week

- Implementing Longest Common Substring algorithm using Generalised Suffix Tree based on Ukk's algorithm.
- Implementing Tokenization on Python codes.
- Implementing DC3 algorithm of Suffix Array and compare their performance.



# Works in next week

- Implementing Longest Common Substring algorithm using Generalised Suffix Tree based on Ukk's algorithm.
- Implementing Tokenization on Python codes.
- Implementing DC3 algorithm of Suffix Array and compare their performance.
- Reading and Understanding n-gram algorithm and its application in code-clone detection.