

# 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

`figure/mississippi.pdf`

# Algorithm of Constructing Suffix Tree

figure/mississippi.pdf



figure/gray.pdf

Node (id:generation)

# Algorithm of Constructing Suffix Tree

figure/mississippi.pdf

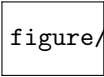
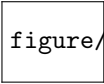
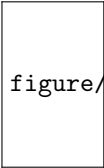
- |                 |
|-----------------|
| figure/gray.pdf |
|-----------------|

 Node (id:generation)
- |                |
|----------------|
| figure/box.pdf |
|----------------|

 Leaf

# Algorithm of Constructing Suffix Tree

figure/mississippi.pdf

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

# Algorithm of Constructing Suffix Tree

figure/mississippi.pdf

- figure/gray.pdf

Node (id:generation)
- figure/box.pdf

Leaf
- figure/line.pdf

Child-relation
- figure/dotted.pdf

suffix-link relation

# Experiment – mississippi

figure/m.pdf



## Experiment – mississippi

figure/mi.pdf

# Experiment – mississippi

figure/mis.pdf

# Experiment – mississippi

figure/miss.pdf

## Experiment – mississippi

figure/missi.pdf

## Experiment – mississippi

figure/missis.pdf

## Experiment – mississippi

figure/mississ.pdf

## Experiment – mississippi

figure/mississi.pdf

# Experiment – mississippi

figure/mississip.pdf



# Experiment – mississippi

figure/mississipp.pdf

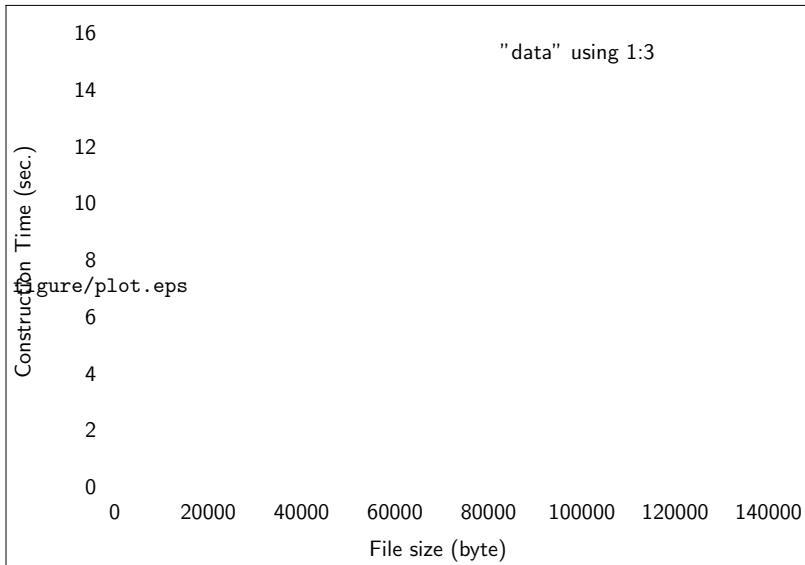
# Experiment – mississippi

`figure/mississippi.pdf`

# Experiment – mississippi

figure/mississippi\_.pdf

# 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.