Report of Progress

by Jiachen Yang

Jiachen Yang[1]

[1]Research Student in Osaka University

October 17, 2011

Part Outline

Algorithm of Finding Common Subsequence on Suffix Tree

Find All common subsequences in a given String

```
$ echo mississippi#|./common_filter.py
issi 4:{1, 4}
si 2:{3, 6}
ssi 3:{2, 5}
```

Find common subsequences between two Strings

```
$ echo abcdef_xcdefgh#|./common_filter.py
cdef 4:{8, 2}
ef 2:{10, 4}
def 3:{9, 3}
```

Filter result

Filter common sequence longer than 40 chars:

```
for length , start_set in apply_filter(st ,
                [lambda length ,start_set:length >40]):
    print(length ,start_set)
```

Filter common sequence occered in more than 3 positions:

```
for length , start_set in apply_filter(st ,
        [lambda length ,start_set:len(start_set)>4]):
    print(length ,start_set)
```

```
VM_S390_ 8:{2842, 915, 4301, 4261, 4282}
VM_GET_ 7:{2023, 719, 3666, 787, 3638, 3612}
```

Tokenize in Python Code

```
farseer ~/pytoken $  ./tokenfc.py tokenfc.py | head -n20
!57Encoding utf-8
!55T #!/usr/bin/python
#56NL

#56NL

!1N import
!1N tokenize
!53O ,
!1N token
!4n

!1N from
!1N functools
!1N import
!1N total_ordering
!4n
```

Combine Tokenizer with Suffix Tree

```
farseer ~/pytoken $  ./st_token.py st_token.py
6:{35, 70} [!1N, !1N, !1N, !53O, !1N, !53O]
8:{8, 3} [!1N, !1N, !1N, !53O, !4n, !1N, !1N, !1N]
6:{15, 31} [!1N, !1N, !4n, !1N, !1N, !1N]
6:{90, 71} [!1N, !1N, !53O, !1N, !53O, !53O]
7:{9, 4} [!1N, !1N, !53O, !4n, !1N, !1N, !1N]
7:{54, 47} [!1N, !53O, !1N, !53O, !1N, !53O, !4n]
7:{122, 100} [!1N, !53O, !1N, !53O, !1N, !53O, !53O]
6:{120, 100, 78, 47} [!1N, !53O, !1N, !53O, !1N, !53O]
11:{80, 37} [!1N, !53O, !1N, !53O, !21, !53O, !53O, !53O, !
7:{91, 102} [!1N, !53O, !1N, !53O, !53O, !21, !53O]
9:{82, 39} [!1N, !53O, !21, !53O, !53O, !53O, !4n, !5>==>,
6:{10, 5} [!1N, !53O, !4n, !1N, !1N, !1N]
7:{41, 84} [!21, !53O, !53O, !53O, !4n, !5>==>, !1N]
6:{17, 7} [!4n, !1N, !1N, !1N, !53O, !4n]
7:{99, 53} [!4n, !1N, !53O, !1N, !53O, !1N, !53O]
6:{48, 55} [!53O, !1N, !53O, !1N, !53O, !4n]
6:{123, 101} [!53O, !1N, !53O, !1N, !53O, !53O]
10:{81, 38} [!53O, !1N, !53O, !21, !53O, !53O, !53O, !4n, !5
```