jupyter Untitled Last Checkpoint: 3 hours ago (autosaved)

File Edit View Insert Cell Kernel Help Python 3

Code CellToolbar

In [45]:
```
import pandas as pd
import numpy as np
```

In [53]:
```
data = pd.read_csv('a.txt', sep='\t')
m = data.values
```

In [82]:
```
file = open('C:/Users/Лев/Downloads/Новая папка/input.txt', 'w')
for i in range(m.shape[0]):
    for j in range(m.shape[1]):
        if m[i,j] == 1:
            file.write(str(j) + ' ')
    file.write('\n')
file.close()

# Поддержка 35 означает, что 35/2000 поддерживают, т.е. minsupp = 0.0175
```

1.

а) 20910 частых множеств с minsupp=35

б) 13812 частых замкнутых множеств с minsupp=35

в) 4002 частых максимальных множества с minsupp=35

г) Рассмотрим любое множество, например, полученное в 1в. К примеру, (1021 1906 995 1888 1011 989 1001 1882 1913). Получаем, что это очень близкие по смыслу словосочетания:

In [89]:
```
print(data.axes[1][1021])
print(data.axes[1][1906])
print(data.axes[1][995])
print(data.axes[1][1888])
print(data.axes[1][1011])
print(data.axes[1][989])
print(data.axes[1][1001])
print(data.axes[1][1882])
print(data.axes[1][1913])
```

```
casino online
gambling online
casino gambling online
gambling internet
casino internet
casino gambling
casino game online
gambling
gambling site
```

2.

а) 10940 ассоциативных правил для minsupp=35 и minconf=1

б) 13812 замкнутых ассоциативных правил для minsupp=35 и minconf=1

в) см. ниже, правил оказалось не 5, а 6 (можно не рассматривать последнее, достоверность которого меньше чем у остальных)

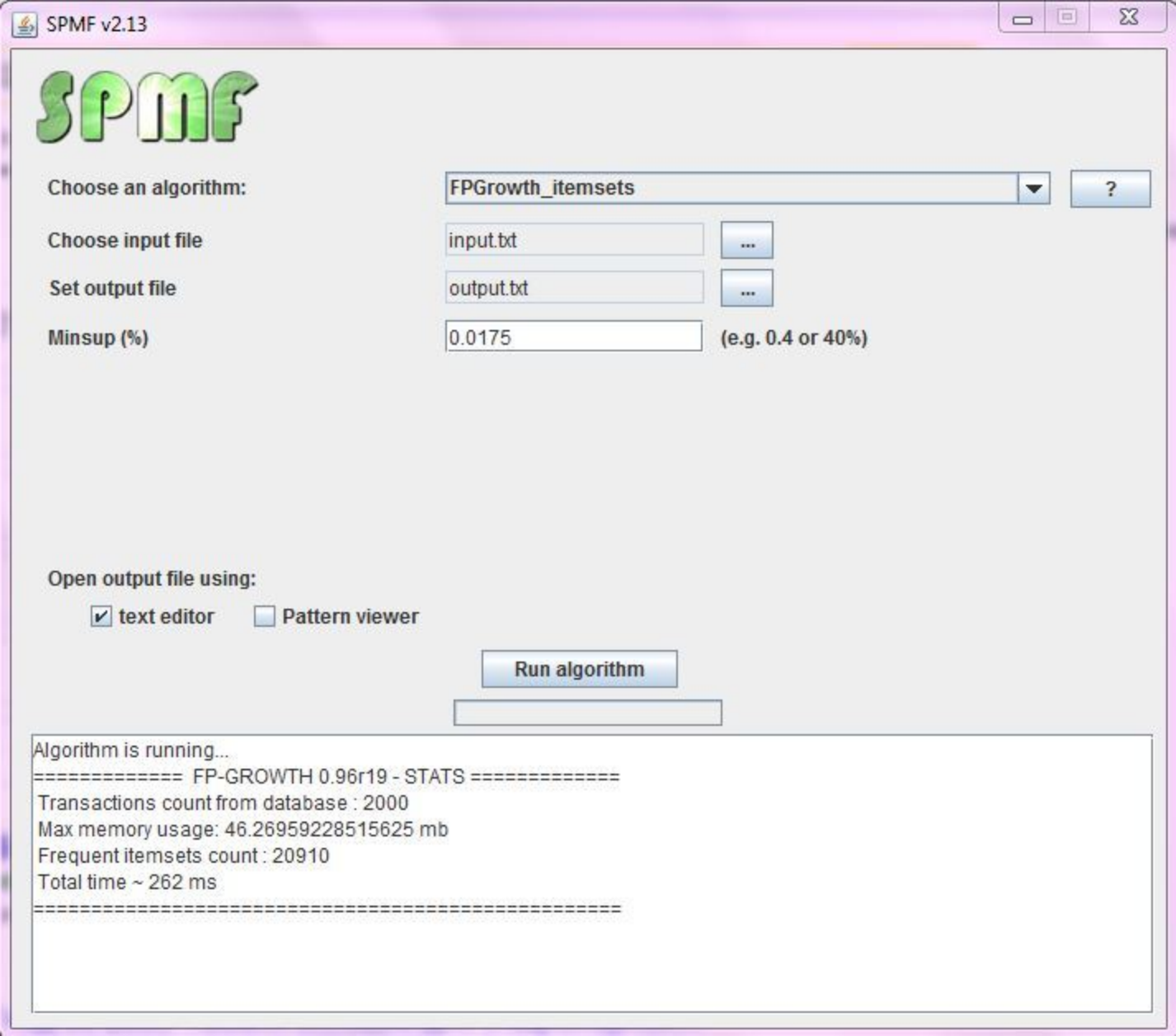
In [80]:
```
# 663 674  ==> 345  #SUP: 90 #CONF: 0.8256880733944955
# 345 674  ==> 663  #SUP: 90 #CONF: 0.8490566037735849
# 2536  ==> 2336  #SUP: 91 #CONF: 0.8666666666666667
# 355  ==> 345  #SUP: 102 #CONF: 0.8292682926829268
# 355  ==> 674  #SUP: 105 #CONF: 0.8536585365853658
# 2159  ==> 2166  #SUP: 109 #CONF: 0.8074074074074075

# Посмотрим, какие словосочетания находятся по выданным номерам:
```

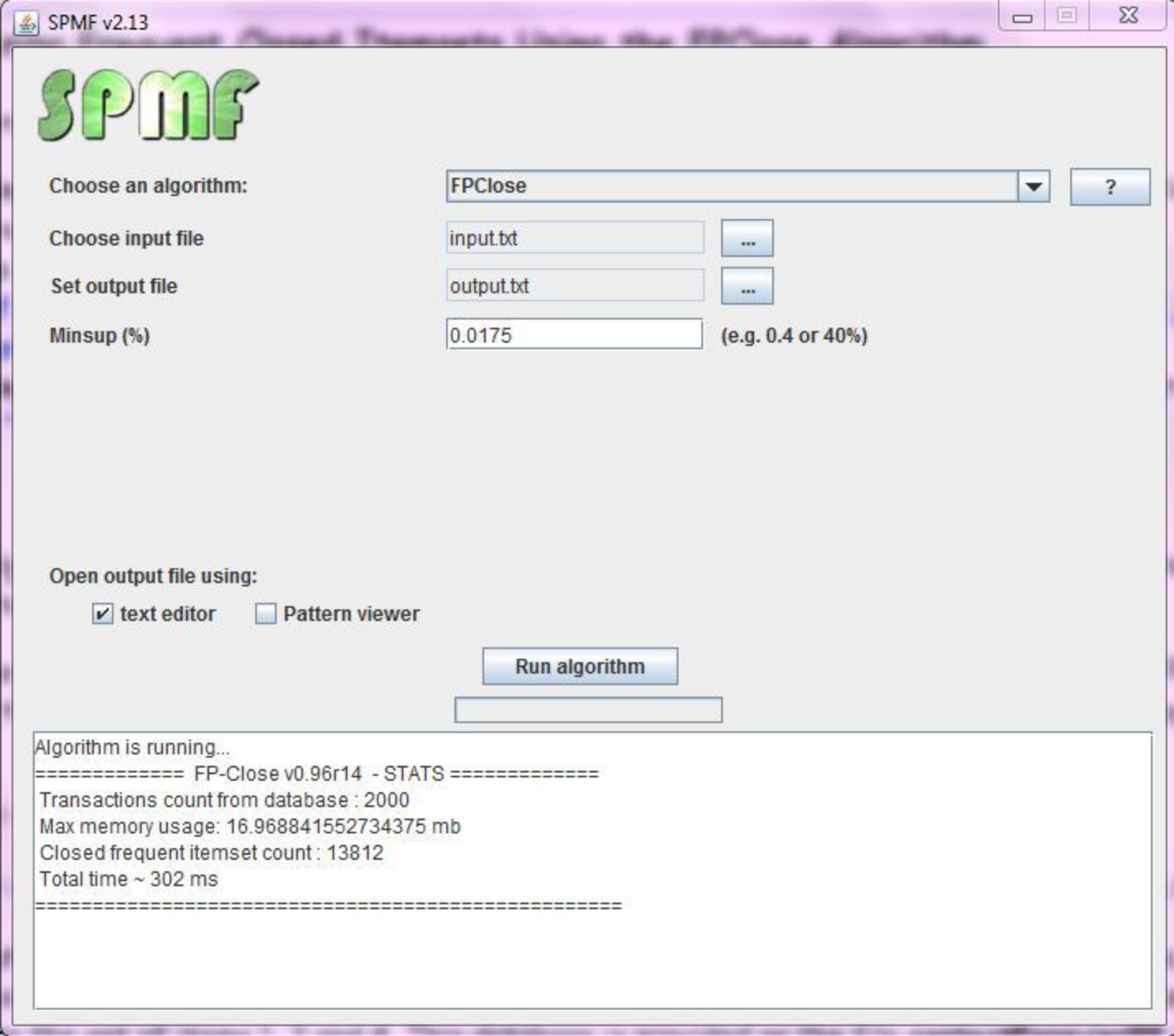
In [81]:
```
print(data.axes[1][663])
print(data.axes[1][674])
print(data.axes[1][345], '\n')

print(data.axes[1][345])
print(data.axes[1][674])
print(data.axes[1][663], '\n')

print(data.axes[1][2536])
print(data.axes[1][2336], '\n')
```

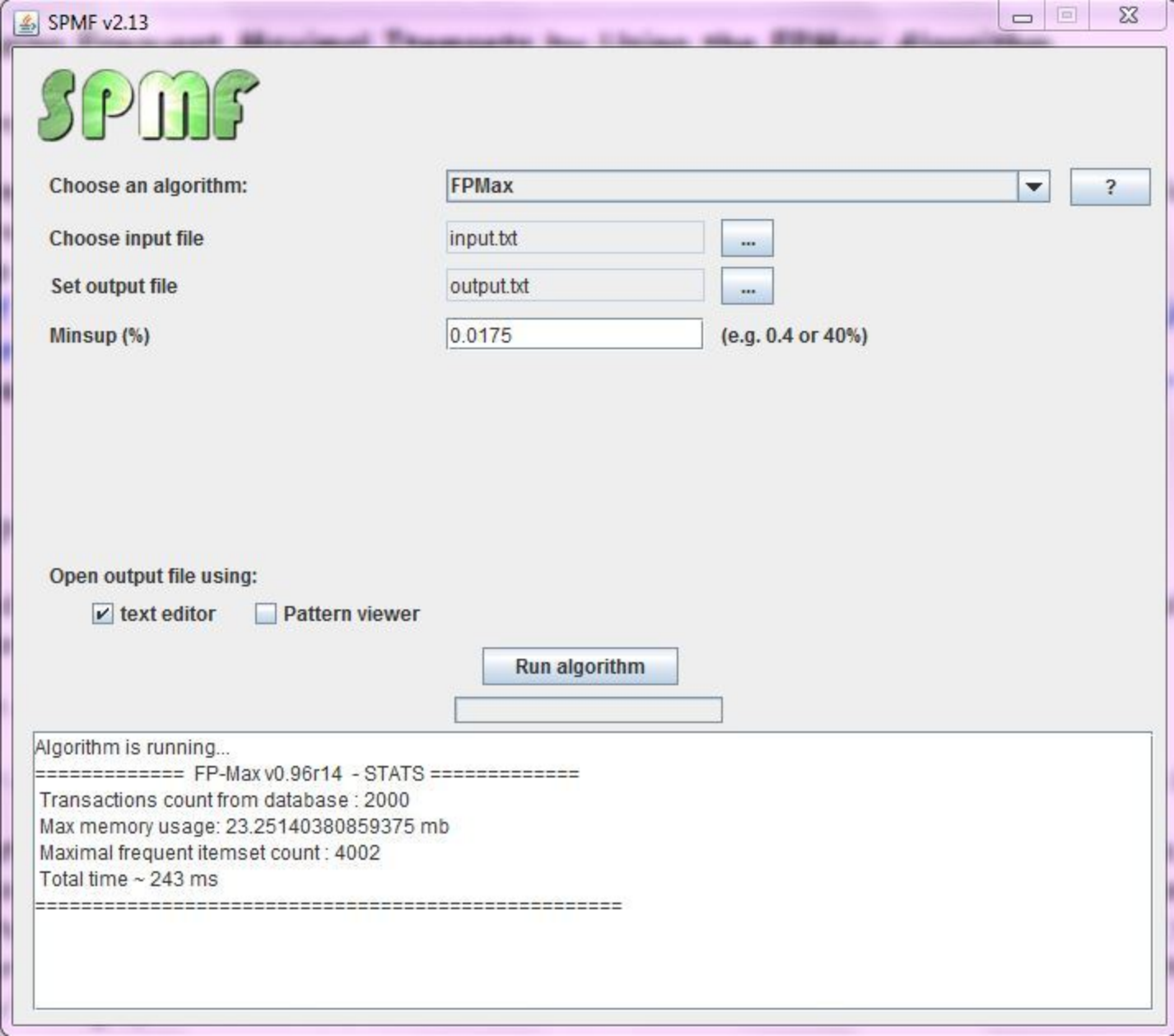
```
print(data.axes[1][355])
print(data.axes[1][345], '\n')

print(data.axes[1][355])
print(data.axes[1][674], '\n')

print(data.axes[1][2159])
print(data.axes[1][2166], '\n')
```

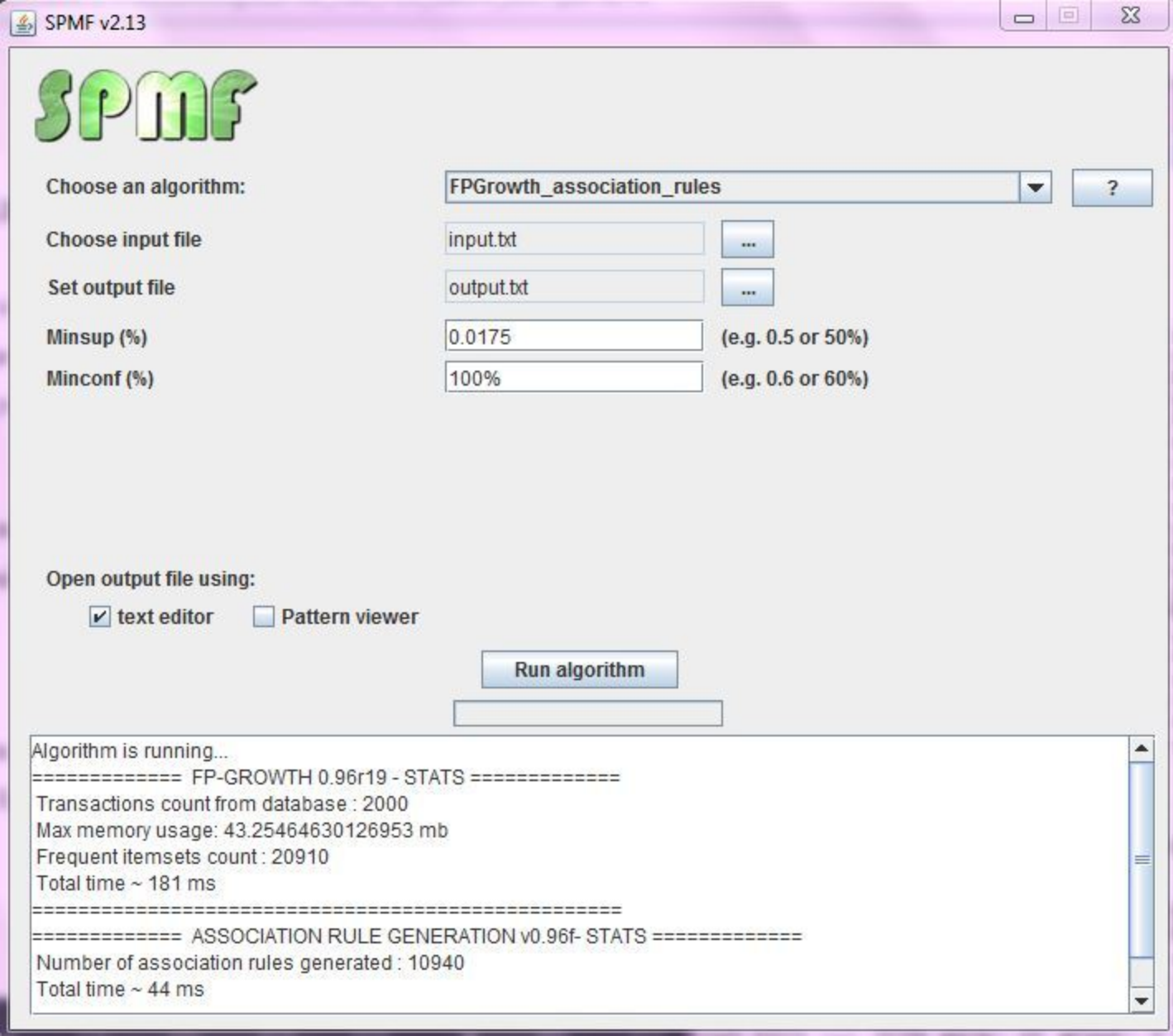
```
business home
business home opportunity
based business home

based business home
business home opportunity
business home

marketing online
internet marketing

based business home opportunity
based business home

based business home opportunity
business home opportunity

hosting site web
hosting web
```

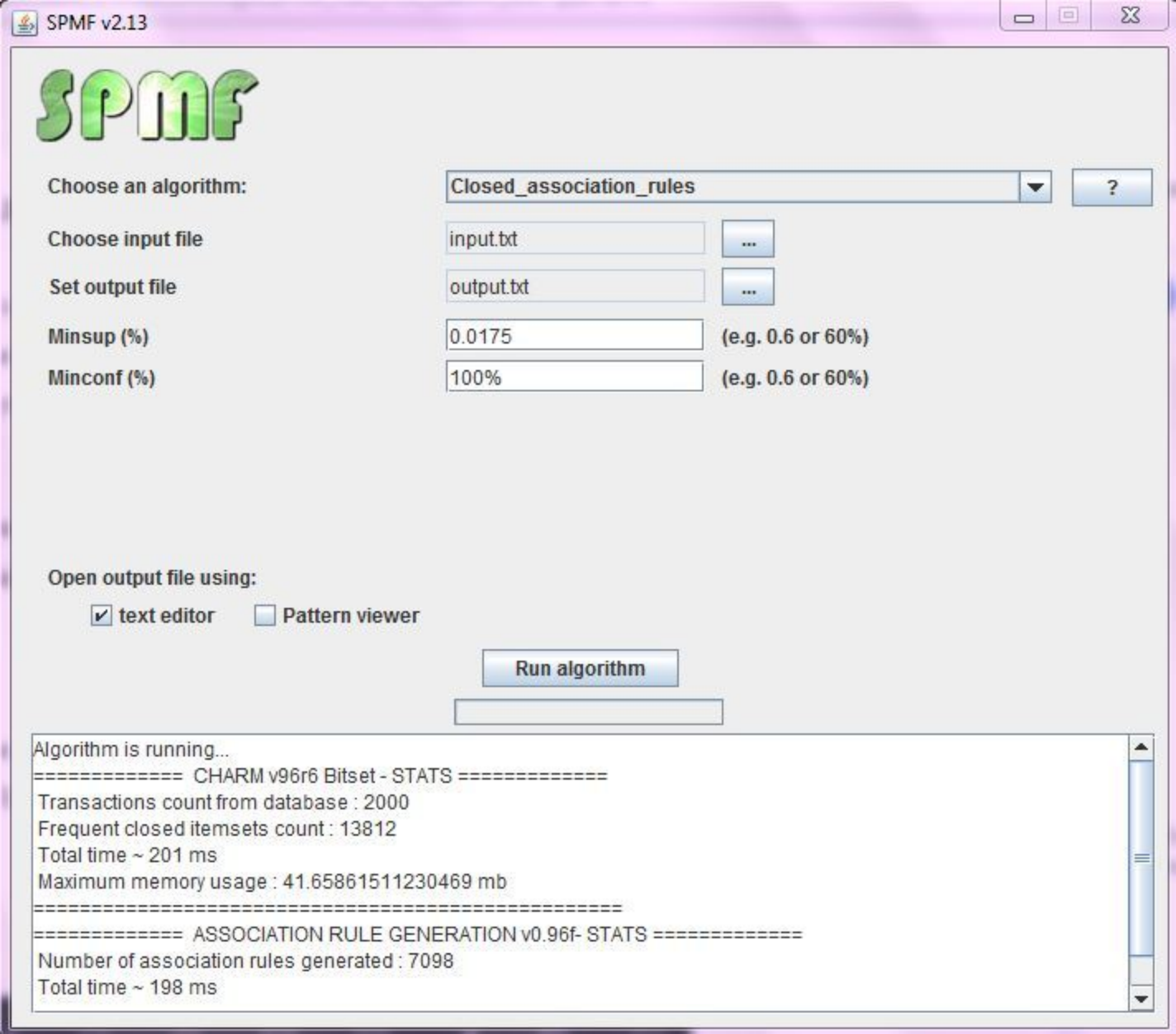
Эти правила указывают на словосочетания, которые являются синонимами, поэтому они так часто встречаются вместе, и поэтому достоверность такая высокая.

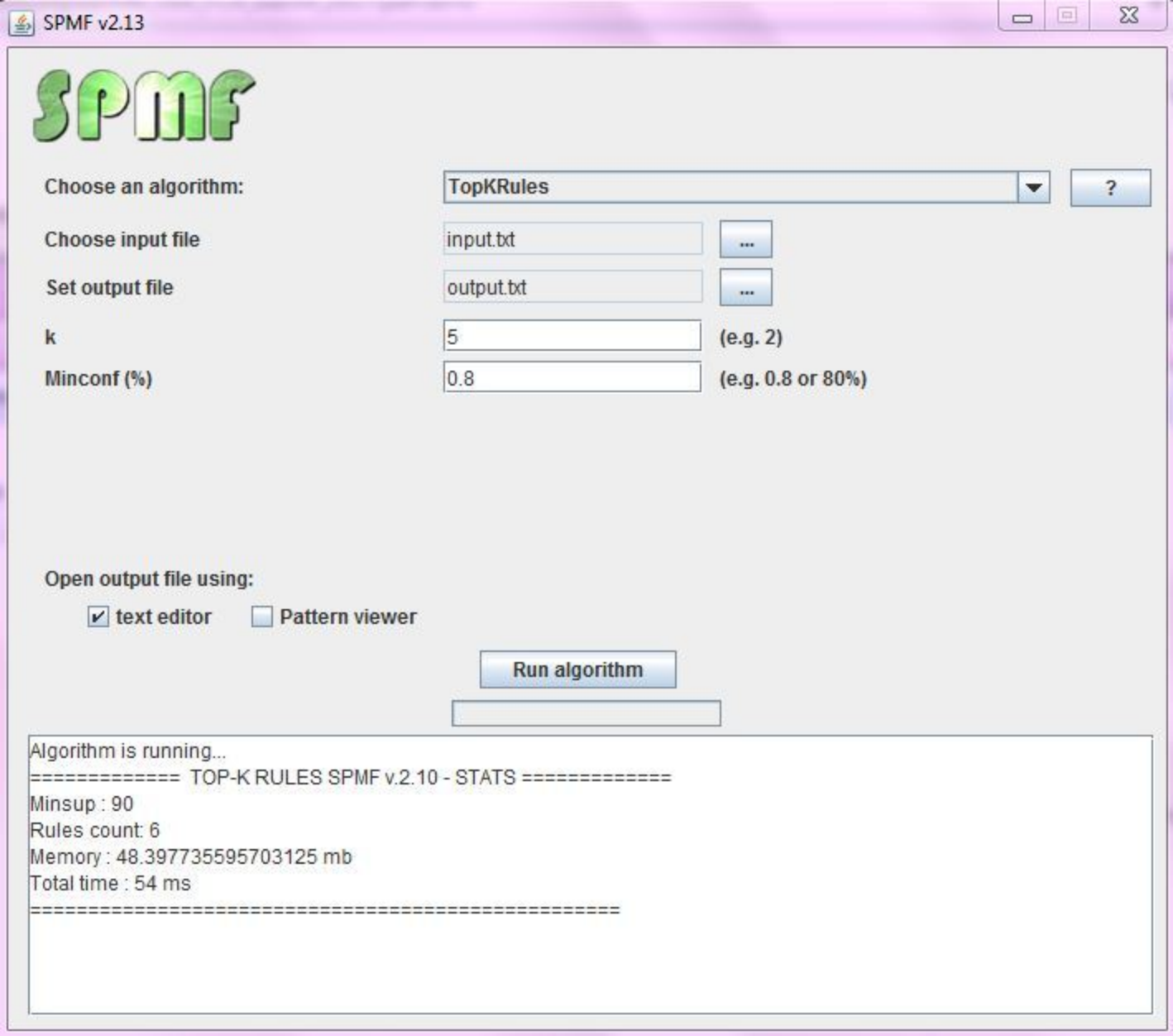

SPMF v2.13
SPMF
Choose an algorithm:
FPGrowth_itemsets
?
Choose input file
input.txt
...
Set output file
output.txt
...
Minsup (%)
0.0175
(e.g. 0.4 or 40%)
Open output file using:
text editor
Pattern viewer
Run algorithm
Algorithm is running...
========== FP-GROWTH 0.96r19 - STATS ==========
Transactions count from database : 2000
Max memory usage: 46.26959228515625 mb
Frequent itemsets count : 20910
Total time ~ 262 ms
==========


SPMF v2.13
Choose an algorithm:
FPClose
?
Choose input file
input.txt
...
Set output file
output.txt
...
Minsup (%)
0.0175
(e.g. 0.4 or 40%)
Open output file using:
text editor
Pattern viewer
Run algorithm
Algorithm is running...
========== FP-Close v0.96r14 - STATS ==========
Transactions count from database : 2000
Max memory usage: 16.968841552734375 mb
Closed frequent itemset count : 13812
Total time ~ 302 ms
==========


SPMF v2.13
Choose an algorithm:
FPMax
?
Choose input file
input.txt
...
Set output file
output.txt
...
Minsup (%)
0.0175
(e.g. 0.4 or 40%)
Open output file using:
text editor
Pattern viewer
Run algorithm
Algorithm is running...
========== FP-Max v0.96r14 - STATS ==========
Transactions count from database : 2000
Max memory usage: 23.25140380859375 mb
Maximal frequent itemset count : 4002
Total time ~ 243 ms
==========


SPMF v2.13
SPMF
Choose an algorithm:
FPGrowth_association_rules
?
Choose input file
input.txt
...
Set output file
output.txt
...
Minsup (%)
0.0175
(e.g. 0.5 or 50%)
Minconf (%)
100%
(e.g. 0.6 or 60%)
Open output file using:
text editor
Pattern viewer
Run algorithm
Algorithm is running...
========== FP-GROWTH 0.96r19 - STATS ==========
Transactions count from database : 2000
Max memory usage: 43.25464630126953 mb
Frequent itemsets count : 20910
Total time ~ 181 ms
==========
========== ASSOCIATION RULE GENERATION v0.96f- STATS ==========
Number of association rules generated : 10940
Total time ~ 44 ms


SPMF v2.13
SPMF
Choose an algorithm:
Closed_association_rules
?
Choose input file
input.txt
...
Set output file
output.txt
...
Minsup (%)
0.0175
(e.g. 0.6 or 60%)
Minconf (%)
100%
(e.g. 0.6 or 60%)
Open output file using:
text editor
Pattern viewer
Run algorithm
Algorithm is running...
========== CHARM v96r6 Bitset - STATS ==========
Transactions count from database : 2000
Frequent closed itemsets count : 13812
Total time ~ 201 ms
Maximum memory usage : 41.65861511230469 mb
==========
========== ASSOCIATION RULE GENERATION v0.96f- STATS ==========
Number of association rules generated : 7098
Total time ~ 198 ms


SPMF v2.13
SPMF
Choose an algorithm:
TopKRules
?
Choose input file
input.txt
...
Set output file
output.txt
...
k
5
(e.g. 2)
Minconf (%)
0.8
(e.g. 0.8 or 80%)
Open output file using:
text editor
Pattern viewer
Run algorithm
Algorithm is running...
========== TOP-K RULES SPMF v.2.10 - STATS ==========
Minsup : 90
Rules count: 6
Memory : 48.397735595703125 mb
Total time : 54 ms
==========