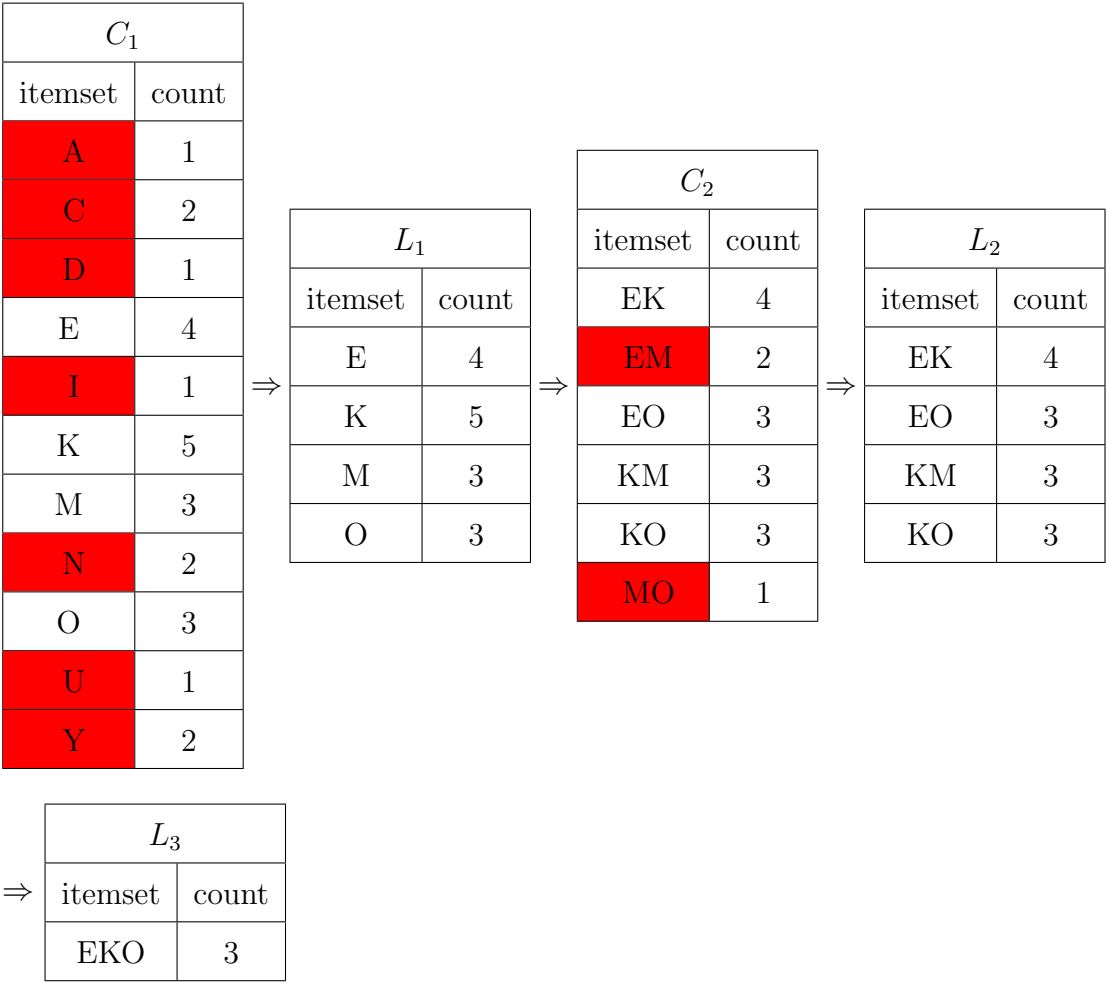


1.  $sup_{min} = 60\% = 3$ ,  $conf_{min} = 80\%$

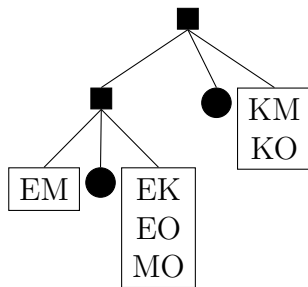
(a) Find all frequent itemsets by following the Apriori Algorithm.



(b) Identify closed and maximal frequent itemsets.

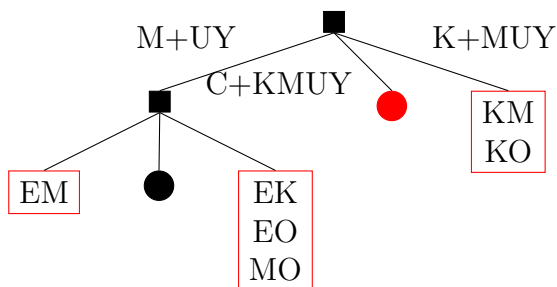
closed	K, EK, KM
maximal	KM, EKO

(c) Draw the hash tree for candidate 2-itemsets, assuming that the hash function will map A, E, M, Y to the left branch, C, I, N to the center branch, and D, K, O, U to the right branch. (Circles are empty leaf nodes, squares are internal hash table nodes)



(d) Identify the hash tree nodes visited when counting the support of 2-itemsets for transaction T400. (Red means node was visited)

T400  $\Rightarrow$  {M, U, C, K, Y}  $\Rightarrow$  {C, K, M, U, Y}

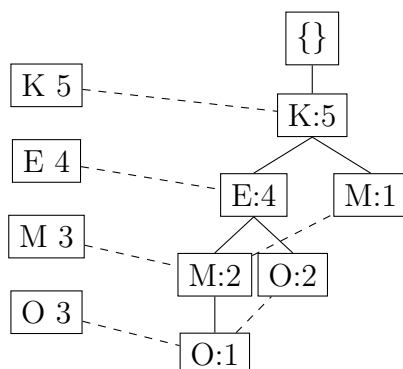


2. Let  $sup_{min} = 60\% = 3$  and  $conf_{min} = 80\%$ . Find all frequent itemsets by following the FP-growth algorithm.

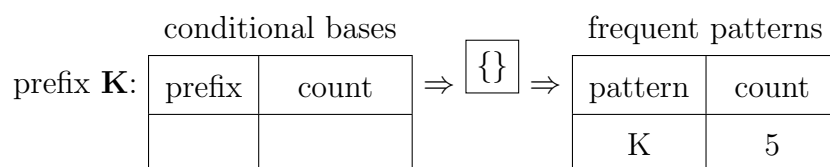
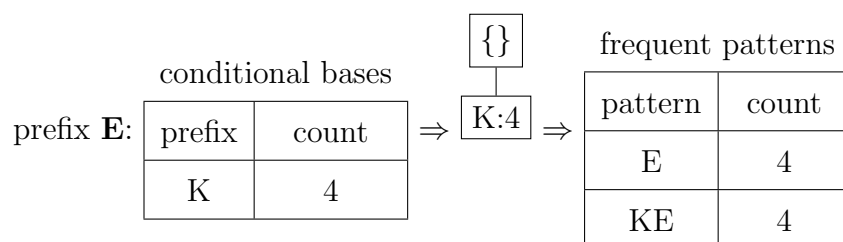
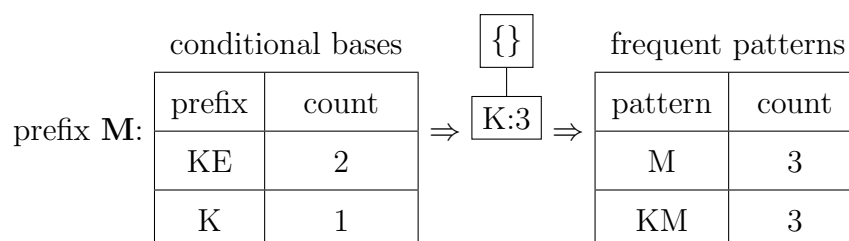
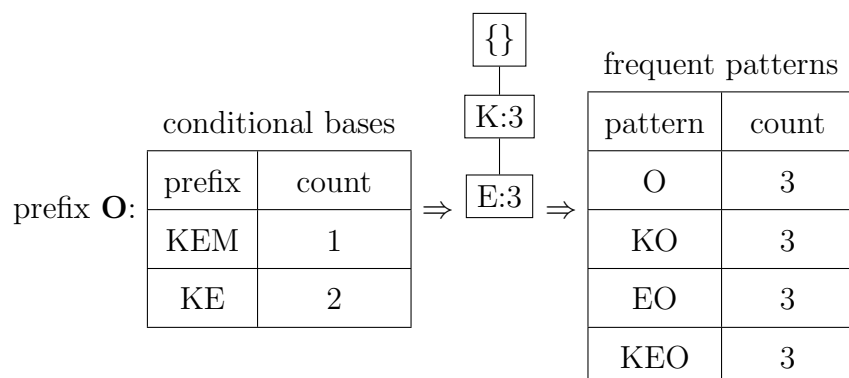
(a) Find the F-list and draw the FP-tree.

F-list

F-tree



(b) For each projected database, list all frequent itemsets that can be enumerated from the database. (Asked professor, show conditional base and conditional tree for each)



3. Find all association rules.

(a) Identify all strong association rules.

(b) Compute the lift measure for each strong association rule. (List both a and b together)

1  $buys(E), buys(K) \Rightarrow buys(O)[60\%, 75\%]$  lift = 1.25

2  $buys(E), buys(O) \Rightarrow buys(K)[60\%, 100\%]$  lift = 1

3  $buys(K), buys(O) \Rightarrow buys(E)[60\%, 100\%]$  lift = 1.25

4  $buys(E) \Rightarrow buys(K), buys(O)[60\%, 75\%]$  lift = 1.25

$5buys(K) \Rightarrow buys(E), buys(O)[60\%, 60\%] lift = 1$   
 $6buys(O) \Rightarrow buys(E), buys(K)[60\%, 100\%] lift = 1.25$   
 $7buys(E) \Rightarrow buys(K)[80\%, 100\%] lift = 1$   
 $8buys(K) \Rightarrow buys(E)[80\%, 80\%] lift = 1$   
 $9buys(E) \Rightarrow buys(O)[60\%, 75\%] lift = 1.25$   
 $10buys(O) \Rightarrow buys(E)[60\%, 100\%] lift = 1.25$   
 $11buys(K) \Rightarrow buys(M)[60\%, 60\%] lift = 1$   
 $12buys(M) \Rightarrow buys(K)[60\%, 100\%] lift = 1$   
 $13buys(K) \Rightarrow buys(O)[60\%, 80\%] lift = 1$   
 $14buys(O) \Rightarrow buys(K)[60\%, 100\%] lift = 1$

4. Convert the transaction data table into an hwk02.arff file, in which each item is a column with either a value t or ? (that is, with a type t). Use Weka Explore to load the data, and run the Apriori and FPGrowth algorithms to learn interesting association rules. Do it for at least three sets of minimum support and confidence. For each algorithm, report the following items.

(a) The commands used to run the learning tasks (Labeled as 'Scheme' in the output)

(b) The output from the runs

**Apriori:**

1.

=== Run information ===

Scheme: weka.associations.Apriori -I -N 20 -T 0 -C 0.8 -D 0.2 -U 1.0 -M 0.6  
 Relation: hwk02  
 Instances: 5  
 Attributes: 11

A  
 C  
 D  
 E  
 I  
 K  
 M  
 N  
 O  
 U  
 Y

=== Associator model (full training set) ===

Apriori

=====

Minimum support: 0.6 (3 instances)

Minimum metric <confidence>: 0.8

Number of cycles performed: 3

Generated sets of large itemsets:

Size of set of large itemsets L(1): 4

Large Itemsets L(1):

E=t 4

K=t 5

M=t 3

O=t 3

Size of set of large itemsets L(2): 4

Large Itemsets L(2):

E=t K=t 4

E=t O=t 3

K=t M=t 3

K=t O=t 3

Size of set of large itemsets L(3): 1

Large Itemsets L(3):

E=t K=t O=t 3

Best rules found:

1. E=t 4 ==> K=t 4 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. O=t 3 ==> E=t 3 <conf:(1)> lift:(1.25) lev:(0.12) [0] conv:(0.6)
3. M=t 3 ==> K=t 3 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
4. O=t 3 ==> K=t 3 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
5. K=t O=t 3 ==> E=t 3 <conf:(1)> lift:(1.25) lev:(0.12) [0] conv:(0.6)
6. E=t O=t 3 ==> K=t 3 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
7. O=t 3 ==> E=t K=t 3 <conf:(1)> lift:(1.25) lev:(0.12) [0] conv:(0.6)
8. K=t 5 ==> E=t 4 <conf:(0.8)> lift:(1) lev:(0) [0] conv:(0.5)

2.

=== Run information ===

Scheme: weka.associations.Apriori -I -N 50 -T 0 -C 0.9 -D 0.25 -U 1.0 -M 0.5  
Relation: hwk02  
Instances: 5  
Attributes: 11

A  
C  
D  
E  
I  
K  
M  
N  
O  
U  
Y

=== Associator model (full training set) ===

Apriori  
=====

Minimum support: 0.5 (3 instances)  
Minimum metric <confidence>: 0.9  
Number of cycles performed: 2

Generated sets of large itemsets:

Size of set of large itemsets L(1): 4

Large Itemsets L(1):

E=t 4  
K=t 5  
M=t 3  
O=t 3

Size of set of large itemsets L(2): 4

Large Itemsets L(2):

E=t K=t 4  
E=t O=t 3  
K=t M=t 3  
K=t O=t 3

Size of set of large itemsets L(3): 1

Large Itemsets L(3):

E=t K=t O=t 3

Best rules found:

1. E=t 4 ==> K=t 4 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. O=t 3 ==> E=t 3 <conf:(1)> lift:(1.25) lev:(0.12) [0] conv:(0.6)
3. M=t 3 ==> K=t 3 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
4. O=t 3 ==> K=t 3 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
5. K=t O=t 3 ==> E=t 3 <conf:(1)> lift:(1.25) lev:(0.12) [0] conv:(0.6)
6. E=t O=t 3 ==> K=t 3 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
7. O=t 3 ==> E=t K=t 3 <conf:(1)> lift:(1.25) lev:(0.12) [0] conv:(0.6)

3.

=== Run information ===

Scheme: weka.associations.Apriori -I -N 50 -T 0 -C 0.7 -D 0.1 -U 1.0 -M 0.4

Relation: hwk02

Instances: 5

Attributes: 11

A

C

D

E

I

K

M

N

O

U

Y

=== Associator model (full training set) ===

Apriori

=====

Minimum support: 0.4 (2 instances)

Minimum metric <confidence>: 0.7

Number of cycles performed: 7

Generated sets of large itemsets:

Size of set of large itemsets L(1): 7

Large Itemsets L(1):

C=t 2  
E=t 4  
K=t 5  
M=t 3  
N=t 2  
O=t 3  
Y=t 2

Size of set of large itemsets L(2): 10

Large Itemsets L(2):

C=t K=t 2  
E=t K=t 4  
E=t M=t 2  
E=t N=t 2  
E=t O=t 3  
K=t M=t 3  
K=t N=t 2  
K=t O=t 3  
K=t Y=t 2  
N=t O=t 2

Size of set of large itemsets L(3): 5

Large Itemsets L(3):

E=t K=t M=t 2  
E=t K=t N=t 2  
E=t K=t O=t 3  
E=t N=t O=t 2  
K=t N=t O=t 2

Size of set of large itemsets L(4): 1

Large Itemsets L(4):

E=t K=t N=t O=t 2

Best rules found:

1. E=t 4 ==> K=t 4 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. O=t 3 ==> E=t 3 <conf:(1)> lift:(1.25) lev:(0.12) [0] conv:(0.6)
3. M=t 3 ==> K=t 3 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
4. O=t 3 ==> K=t 3 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
5. K=t O=t 3 ==> E=t 3 <conf:(1)> lift:(1.25) lev:(0.12) [0] conv:(0.6)
6. E=t O=t 3 ==> K=t 3 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)



```

7. O=t 3 ==> E=t K=t 3    <conf:(1)> lift:(1.25) lev:(0.12) [0] conv:(0.6)
8. C=t 2 ==> K=t 2        <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
9. N=t 2 ==> E=t 2        <conf:(1)> lift:(1.25) lev:(0.08) [0] conv:(0.4)
10. N=t 2 ==> K=t 2        <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
11. Y=t 2 ==> K=t 2        <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
12. N=t 2 ==> O=t 2        <conf:(1)> lift:(1.67) lev:(0.16) [0] conv:(0.8)
13. E=t M=t 2 ==> K=t 2    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
14. K=t N=t 2 ==> E=t 2    <conf:(1)> lift:(1.25) lev:(0.08) [0] conv:(0.4)
15. E=t N=t 2 ==> K=t 2    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
16. N=t 2 ==> E=t K=t 2    <conf:(1)> lift:(1.25) lev:(0.08) [0] conv:(0.4)
17. N=t O=t 2 ==> E=t 2    <conf:(1)> lift:(1.25) lev:(0.08) [0] conv:(0.4)
18. E=t N=t 2 ==> O=t 2    <conf:(1)> lift:(1.67) lev:(0.16) [0] conv:(0.8)
19. N=t 2 ==> E=t O=t 2    <conf:(1)> lift:(1.67) lev:(0.16) [0] conv:(0.8)
20. N=t O=t 2 ==> K=t 2    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
21. K=t N=t 2 ==> O=t 2    <conf:(1)> lift:(1.67) lev:(0.16) [0] conv:(0.8)
22. N=t 2 ==> K=t O=t 2    <conf:(1)> lift:(1.67) lev:(0.16) [0] conv:(0.8)
23. K=t N=t O=t 2 ==> E=t 2    <conf:(1)> lift:(1.25) lev:(0.08) [0] conv:(0.4)
24. E=t N=t O=t 2 ==> K=t 2    <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
25. E=t K=t N=t 2 ==> O=t 2    <conf:(1)> lift:(1.67) lev:(0.16) [0] conv:(0.8)
26. N=t O=t 2 ==> E=t K=t 2    <conf:(1)> lift:(1.25) lev:(0.08) [0] conv:(0.4)
27. K=t N=t 2 ==> E=t O=t 2    <conf:(1)> lift:(1.67) lev:(0.16) [0] conv:(0.8)
28. E=t N=t 2 ==> K=t O=t 2    <conf:(1)> lift:(1.67) lev:(0.16) [0] conv:(0.8)
29. N=t 2 ==> E=t K=t O=t 2    <conf:(1)> lift:(1.67) lev:(0.16) [0] conv:(0.8)
30. K=t 5 ==> E=t 4        <conf:(0.8)> lift:(1) lev:(0) [0] conv:(0.5)
31. E=t 4 ==> O=t 3        <conf:(0.75)> lift:(1.25) lev:(0.12) [0] conv:(0.8)
32. E=t K=t 4 ==> O=t 3    <conf:(0.75)> lift:(1.25) lev:(0.12) [0] conv:(0.8)
33. E=t 4 ==> K=t O=t 3    <conf:(0.75)> lift:(1.25) lev:(0.12) [0] conv:(0.8)

```

## FP-Growth:

1.

=== Run information ===

```

Scheme:      weka.associations.FPGrowth -P 2 -I -1 -N 50 -T 0 -C 0.6 -D 0.05 -U 1
Relation:    hwk02
Instances:    5
Attributes:   11
A
C
D
E
I
K
M
N

```

O  
U  
Y

=== Associator model (full training set) ===

FPGrowth found 2 rules

1. [E=t]: 4 ==> [K=t]: 4 <conf:(1)> lift:(1) lev:(0) conv:(0)
2. [K=t]: 5 ==> [E=t]: 4 <conf:(0.8)> lift:(1) lev:(0) conv:(0.5)

2.

=== Run information ===

Scheme: weka.associations.FPGrowth -P 2 -I -1 -N 50 -T 0 -C 0.6 -D 0.05 -U 1  
Relation: hwk02  
Instances: 5  
Attributes: 11

A  
C  
D  
E  
I  
K  
M  
N  
O  
U  
Y

=== Associator model (full training set) ===

FPGrowth found 48 rules

1. [E=t]: 4 ==> [K=t]: 4 <conf:(1)> lift:(1) lev:(0) conv:(0)
2. [O=t]: 3 ==> [K=t]: 3 <conf:(1)> lift:(1) lev:(0) conv:(0)
3. [M=t]: 3 ==> [K=t]: 3 <conf:(1)> lift:(1) lev:(0) conv:(0)
4. [Y=t]: 2 ==> [K=t]: 2 <conf:(1)> lift:(1) lev:(0) conv:(0)
5. [N=t]: 2 ==> [K=t]: 2 <conf:(1)> lift:(1) lev:(0) conv:(0)
6. [C=t]: 2 ==> [K=t]: 2 <conf:(1)> lift:(1) lev:(0) conv:(0)
7. [O=t]: 3 ==> [E=t]: 3 <conf:(1)> lift:(1.25) lev:(0.12) conv:(0.6)
8. [N=t]: 2 ==> [E=t]: 2 <conf:(1)> lift:(1.25) lev:(0.08) conv:(0.4)
9. [N=t]: 2 ==> [O=t]: 2 <conf:(1)> lift:(1.67) lev:(0.16) conv:(0.8)
10. [O=t]: 3 ==> [K=t, E=t]: 3 <conf:(1)> lift:(1.25) lev:(0.12) conv:(0.6)
11. [K=t, O=t]: 3 ==> [E=t]: 3 <conf:(1)> lift:(1.25) lev:(0.12) conv:(0.6)
12. [E=t, O=t]: 3 ==> [K=t]: 3 <conf:(1)> lift:(1) lev:(0) conv:(0)
13. [E=t, M=t]: 2 ==> [K=t]: 2 <conf:(1)> lift:(1) lev:(0) conv:(0)

14. [N=t]: 2 ==> [K=t, E=t]: 2 <conf:(1)> lift:(1.25) lev:(0.08) conv:(0.4)  
 15. [K=t, N=t]: 2 ==> [E=t]: 2 <conf:(1)> lift:(1.25) lev:(0.08) conv:(0.4)  
 16. [E=t, N=t]: 2 ==> [K=t]: 2 <conf:(1)> lift:(1) lev:(0) conv:(0)  
 17. [N=t]: 2 ==> [K=t, O=t]: 2 <conf:(1)> lift:(1.67) lev:(0.16) conv:(0.8)  
 18. [K=t, N=t]: 2 ==> [O=t]: 2 <conf:(1)> lift:(1.67) lev:(0.16) conv:(0.8)  
 19. [O=t, N=t]: 2 ==> [K=t]: 2 <conf:(1)> lift:(1) lev:(0) conv:(0)  
 20. [N=t]: 2 ==> [E=t, O=t]: 2 <conf:(1)> lift:(1.67) lev:(0.16) conv:(0.8)  
 21. [E=t, N=t]: 2 ==> [O=t]: 2 <conf:(1)> lift:(1.67) lev:(0.16) conv:(0.8)  
 22. [O=t, N=t]: 2 ==> [E=t]: 2 <conf:(1)> lift:(1.25) lev:(0.08) conv:(0.4)  
 23. [N=t]: 2 ==> [K=t, E=t, O=t]: 2 <conf:(1)> lift:(1.67) lev:(0.16) conv:(0.8)  
 24. [K=t, N=t]: 2 ==> [E=t, O=t]: 2 <conf:(1)> lift:(1.67) lev:(0.16) conv:(0.8)  
 25. [E=t, N=t]: 2 ==> [K=t, O=t]: 2 <conf:(1)> lift:(1.67) lev:(0.16) conv:(0.8)  
 26. [K=t, E=t, N=t]: 2 ==> [O=t]: 2 <conf:(1)> lift:(1.67) lev:(0.16) conv:(0.8)  
 27. [O=t, N=t]: 2 ==> [K=t, E=t]: 2 <conf:(1)> lift:(1.25) lev:(0.08) conv:(0.4)  
 28. [K=t, O=t, N=t]: 2 ==> [E=t]: 2 <conf:(1)> lift:(1.25) lev:(0.08) conv:(0.4)  
 29. [E=t, O=t, N=t]: 2 ==> [K=t]: 2 <conf:(1)> lift:(1) lev:(0) conv:(0)  
 30. [K=t]: 5 ==> [E=t]: 4 <conf:(0.8)> lift:(1) lev:(0) conv:(0.5)  
 31. [E=t]: 4 ==> [O=t]: 3 <conf:(0.75)> lift:(1.25) lev:(0.12) conv:(0.8)  
 32. [E=t]: 4 ==> [K=t, O=t]: 3 <conf:(0.75)> lift:(1.25) lev:(0.12) conv:(0.8)  
 33. [K=t, E=t]: 4 ==> [O=t]: 3 <conf:(0.75)> lift:(1.25) lev:(0.12) conv:(0.8)  
 34. [M=t]: 3 ==> [E=t]: 2 <conf:(0.67)> lift:(0.83) lev:(-0.08) conv:(0.3)  
 35. [O=t]: 3 ==> [N=t]: 2 <conf:(0.67)> lift:(1.67) lev:(0.16) conv:(0.9)  
 36. [M=t]: 3 ==> [K=t, E=t]: 2 <conf:(0.67)> lift:(0.83) lev:(-0.08) conv:(0.3)  
 37. [K=t, M=t]: 3 ==> [E=t]: 2 <conf:(0.67)> lift:(0.83) lev:(-0.08) conv:(0.3)  
 38. [O=t]: 3 ==> [K=t, N=t]: 2 <conf:(0.67)> lift:(1.67) lev:(0.16) conv:(0.9)  
 39. [K=t, O=t]: 3 ==> [N=t]: 2 <conf:(0.67)> lift:(1.67) lev:(0.16) conv:(0.9)  
 40. [O=t]: 3 ==> [E=t, N=t]: 2 <conf:(0.67)> lift:(1.67) lev:(0.16) conv:(0.9)  
 41. [E=t, O=t]: 3 ==> [N=t]: 2 <conf:(0.67)> lift:(1.67) lev:(0.16) conv:(0.9)  
 42. [O=t]: 3 ==> [K=t, E=t, N=t]: 2 <conf:(0.67)> lift:(1.67) lev:(0.16) conv:(0.9)  
 43. [K=t, O=t]: 3 ==> [E=t, N=t]: 2 <conf:(0.67)> lift:(1.67) lev:(0.16) conv:(0.9)  
 44. [E=t, O=t]: 3 ==> [K=t, N=t]: 2 <conf:(0.67)> lift:(1.67) lev:(0.16) conv:(0.9)  
 45. [K=t, E=t, O=t]: 3 ==> [N=t]: 2 <conf:(0.67)> lift:(1.67) lev:(0.16) conv:(0.9)  
 46. [K=t]: 5 ==> [O=t]: 3 <conf:(0.6)> lift:(1) lev:(0) conv:(0.67)  
 47. [K=t]: 5 ==> [M=t]: 3 <conf:(0.6)> lift:(1) lev:(0) conv:(0.67)  
 48. [K=t]: 5 ==> [E=t, O=t]: 3 <conf:(0.6)> lift:(1) lev:(0) conv:(0.67)

3.

=== Run information ===

Scheme: weka.associations.FPGrowth -P 2 -I -1 -N 50 -T 0 -C 0.9 -D 0.05 -U 1  
 Relation: hwk02  
 Instances: 5  
 Attributes: 11  
 A  
 C

D  
E  
I  
K  
M  
N  
O  
U  
Y

=== Associator model (full training set) ===

FPGrowth found 7 rules

1. [E=t]: 4 ==> [K=t]: 4 <conf:(1)> lift:(1) lev:(0) conv:(0)
2. [O=t]: 3 ==> [K=t]: 3 <conf:(1)> lift:(1) lev:(0) conv:(0)
3. [M=t]: 3 ==> [K=t]: 3 <conf:(1)> lift:(1) lev:(0) conv:(0)
4. [O=t]: 3 ==> [E=t]: 3 <conf:(1)> lift:(1.25) lev:(0.12) conv:(0.6)
5. [O=t]: 3 ==> [K=t, E=t]: 3 <conf:(1)> lift:(1.25) lev:(0.12) conv:(0.6)
6. [K=t, O=t]: 3 ==> [E=t]: 3 <conf:(1)> lift:(1.25) lev:(0.12) conv:(0.6)
7. [E=t, O=t]: 3 ==> [K=t]: 3 <conf:(1)> lift:(1) lev:(0) conv:(0)