Course: Data Mining

Homework 3

Student ID: 20C14001

Student Name: Le Duong Tuan Anh

Transaction Dataset

TID	Item
1	A C T W
2	C D W
3	A C T W
4	A C D W
5	A C D T W
6	C D T

minsup = 60% (s)

minconf = 80% (c)

1. Find all frequent itemsets by using Apriori.

Total transaction: 6.

1st scan

Items	Support (count)	Support (%)	Acceptable minsup=60%
A	4	4 / 6 = 67%	Y
C	6	6 / 6 = 100%	Y
T	4	4 / 6 = 67%	Y
W	5	5 / 6 = 83%	Y
D	4	4 / 6 = 67%	Y

2nd scan

Items	Support (count)	Support (%)	Acceptable minsup=60%
{A, C}	4	4 / 6 (67%)	Y
{A, T}	3	3 / 6 (50%)	N
{A, W}	4	4 / 6 (67%)	Y
{A, D}	2	2 / 6 (33%)	N
{C, T}	4	4 / 6 (67%)	Y
{C, W}	5	5 / 6 (83%)	Y
{C, D}	4	4 / 6 (67%)	Y
{T, W}	3	3 / 6 (50%)	N
{T, D}	2	2 / 6 (33%)	N
{W, D}	3	3 / 6 (50%)	N

3rd scan

Items	Support (count)	Support (%)	Acceptable minsup=60%
{A, C, W}	4	4 / 6 (67%)	Y
{C, T, W}	3	3 / 6 (50%)	N
{C, T, D}	2	2 / 6 (33%)	N

{C, W, D}	3	3 / 6 (50%)	N

4th scan (end)

Items	Support (count)
{A, C, W}	4

Frequent Itemsets:

Items	Support (count)	Support (%)
{A, C, W}	4	4 / 6 (67%)
{C, T}	4	4 / 6 (67%)
{C, W}	5	5 / 6 (83%)
{C, D}	4	4 / 6 (67%)

2. Find all frequent itemsets by using FP-Growth.

Total transaction: 6.

Frequent 1-itemset

Items	Support (count)	Support (%)	Acceptable minsup=60%
A	4	4 / 6 = 67%	Y
C	6	6 / 6 = 100%	Y
T	4	4 / 6 = 67%	Y
W	5	5 / 6 = 83%	Y
D	4	4 / 6 = 67%	Y

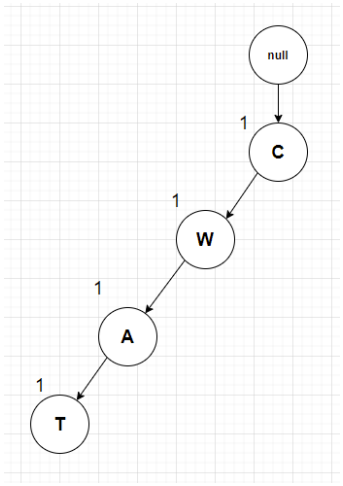
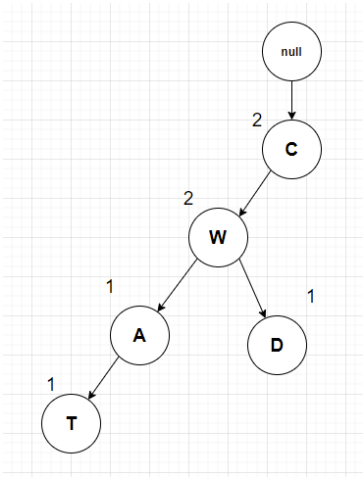
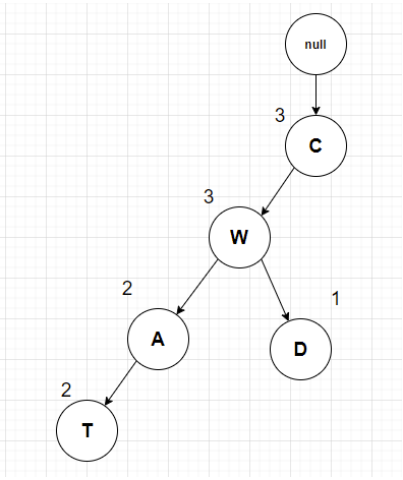
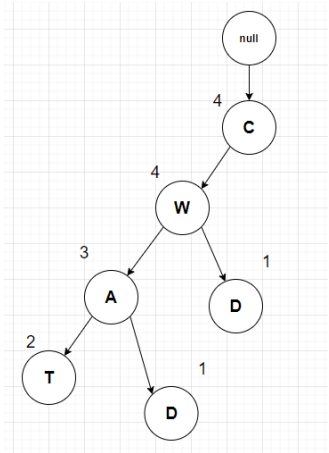
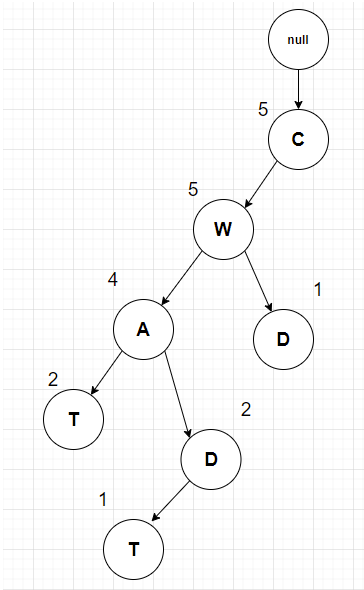
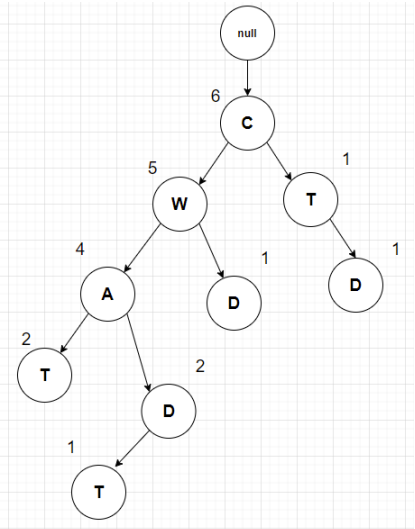
F-list, sorted by support.

F-list	C	W	A	T	D

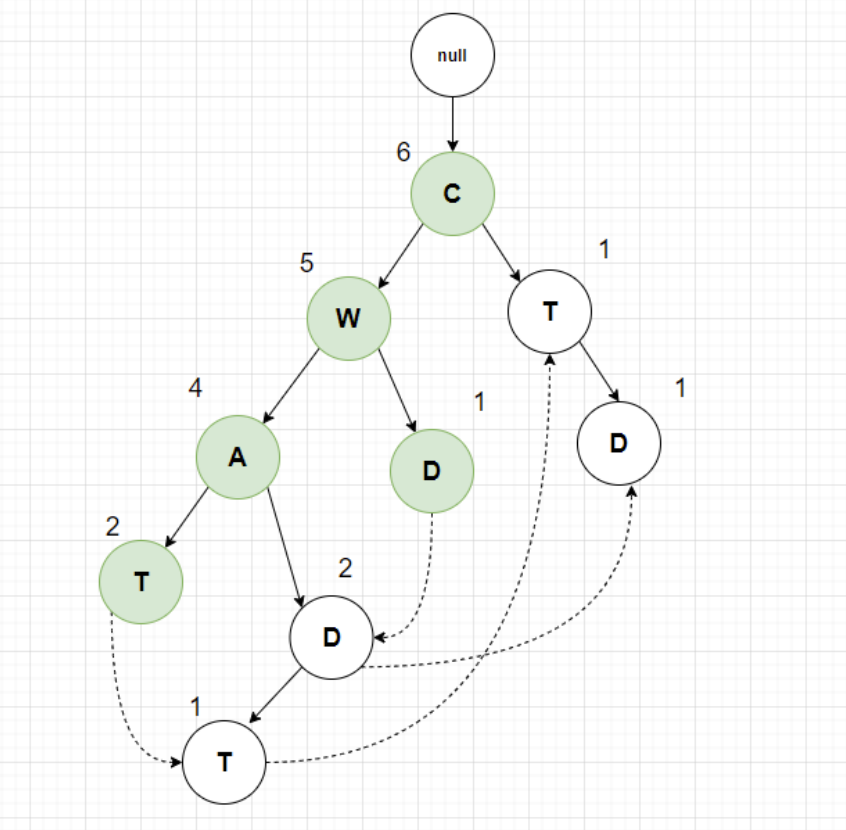
Sort frequent items in transactions, based on F-list

TID	Item
1	C W A T
2	C W D
3	C W A T
4	C W A D
5	C W A D T
6	C T D

Construct FP-Tree

Transaction 1	Transaction 2	Transaction 3
		
Transaction 4	**Transaction 5**	**Transaction 6**
		

Final FP-Tree



Conditional Database

Item	**Conditional Database**	**Acceptable with minsup = 60% (3.6)**
W	C:5	CW (5 = 83%)
A	CW: 4	CWA (4 = 67%)
T	CWA: 2, CWAD: 1, C:1	CT (4 = 67%)
D	CW: 1, CWAD: 2, CT: 1	CD (4 = 67%)

Frequent Itemsets:

Items	**Support (count)**	**Support (%)**
{C, W}	5	5 / 6 (83%)
{C, W, A}	4	4 / 6 (67%)
{C, T}	4	4 / 6 (67%)
{C, D}	4	4 / 6 (67%)

3. File all association rules that satisfy [s,c]

Items / Support	**Association Rule X → Y**	**Confidence (%) | X ∪ Y | / | X |**	**Acceptable minconf=80%**
{A, W} / 67%	A → W	4 / 5 (80%)	Y
{W, A} / 67%	W → A	4 / 4 (100%)	Y
{A, C} / 67%	A → C	4 / 6 (67%)	N
{C, A} / 67%	C → A	4 / 4 (100%)	Y
{C, W} / 83%	C → W	5 / 6 (83%)	Y
{C, W} / 83%	W → C	5 / 5 (100%)	Y
{A, C, W} / 67%	A → C, W	4 / 4 (100%)	Y
{A, C, W} / 67%	A, C → W	4 / 4 (100%)	Y
{A, C, W} / 67%	A, W → C	4 / 4 (100%)	Y
{A, C, W} / 67%	C → A, W	4 / 6 (67%)	N
{A, C, W} / 67%	C, W → A	4 / 5 (80%)	Y
{A, C, W} / 67%	W → A, C	4 / 5 (80%)	Y
{C, T} / 67%	C → T	4 / 6 (67%)	N
{C, T} / 67%	T → C	4 / 4 (100%)	Y
{C, D} / 67%	C → D	4 / 6 (67%)	N
{C, D} / 67%	D → C	4 / 4 (100%)	Y

4) For each rule of question 3, use Lift measure to assess the degree to which the occurrence of the left part "lifts" the occurrence of the right part.

Association Rule **X → Y**	**Confidence (%)** **| X ∪ Y | / | X |**	**Lift** **conf(X → Y) / sup(Y)**	**Measure association rule**
A → W	4 / 5 (80%)	0.8 / 0.67 = 1.19	Positively correlated
W → A	4 / 4 (100%)	1 / 0.67 = 1.49	Positively correlated
C → A	4 / 4 (100%)	1 / 0.67 = 1.49	Positively correlated
C → W	5 / 6 (83%)	0.83 / 0.83 = 1	Independent
W → C	5 / 5 (100%)	1 / 1 = 1	Independent
A → C, W	4 / 4 (100%)	1 / 0.83 = 1.2	Positively correlated
A, C → W	4 / 4 (100%)	1 / 0.83 = 1.2	Positively correlated
A, W → C	4 / 4 (100%)	1 / 1 = 1	Independent
C, W → A	4 / 5 (80%)	0.8 / 0.67 = 1.19	Positively correlated
W → A, C	4 / 5 (100%)	1 / 0.67 = 1.49	Positively correlated
T → C	4 / 4 (100%)	1 / 1 = 1	Independent
D → C	4 / 4 (100%)	1 / 0.67 = 1.49	Positively correlated