Quiz #3: Frequent Itemsets Week 1

10/10

Name: Yijun Lin ID: 3689281438

1) Consider the entire set of items contains: A, B, C, D, E,…, Z (a total of 26 items), how much memory do you need if you use a hash table (triples) to count the occurrence of each possible pair assuming only 1/4 of the pairs have an occurrence > 0? (you can assume that a counter is 4 bytes) (1 pts)

$\binom{26}{2} \times \frac{1}{4} \times 4 \times 3 = 975$ bytes (1)

2) For the **A Priori algorithm**, consider the following input file of basket data and a support threshold **s = 3**, answer the following questions.

***Basket data: {a, b, c, d, e} {a, d, e, b, c} {a, b, c} {a, b, c} {a, b, e} {a, b}**

2.1) What are the **item counts** produced in pass 1 and which of these items are frequent? (1pt)

Item	Count
a	6
b	6
c	4
d	2
e	3

a.b.c.e are frequent (1)

2.2) For pass 2, **which are the candidate pairs** for each basket? (Only include the pairs that **will be counted**.) (2 pts)

Basket	Candidate pairs
1	{a,b} {a,c} {a,e} {b,c} {b,e} {c,e}
2	{a,c} {a,b} {a,e} {b,c} {b,e} {c,e}
3	~~{a,e}~~ {a,b} {a,c} {b,c}
4	{b,c} {a,b} {a,c}
5	{b,e} {a,b} {a,e}
6	~~{c,e}~~ {a,b}

(2)

2.3) What is the **count for each candidate pair and** which of the **candidate pairs are frequent**? (2pt)

Candidate pair	Count
{a,b}	6
{a,c}	4
{a,e}	3

{a,b} {a,c} {a,e} {b,c} {b,e} are frequent