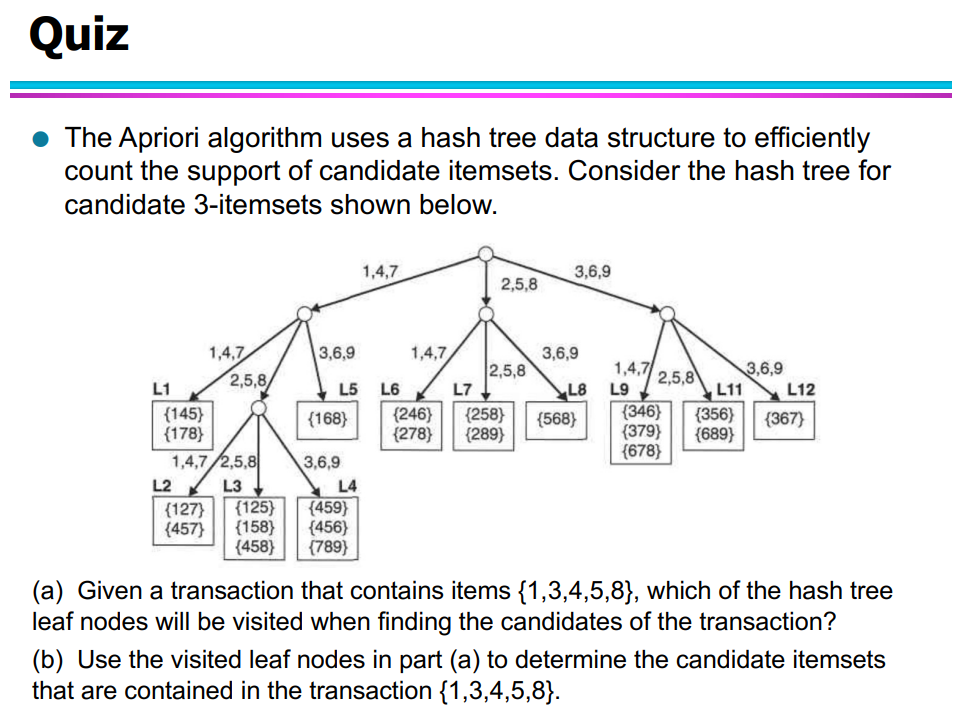
**10215501435-杨茜雅-数据挖掘Quiz**



题目1：

Apriori算法使用哈希树数据结构有效地计数候选项集的支持度。请考虑下面显示的候选3-项集的哈希树。

(a) 给定一个包含物品 {1,3,4,5,8} 的交易记录，当寻找该交易的候选项集时，哪些哈希树的叶节点将会被访问？

(b) 使用（a）部分中访问的叶节点来确定包含在交易记录 {1,3,4,5,8} 中的候选项集。

**题目答案：**

(a) L1 L3 L5 L9 L11

(b) {145}，{158}，{458}

**解答：**

a)

首先，生成交易{1,3,4,5,8}所有可能的3-项集：

{1,3,4} {1,3,5} {1,3,8} {1,4,5} {1,4,8} {1,5,8} {3,4,5} {3,4,8} {3,5,8} {4,5,8}

然后，我们根据哈希树的分支逻辑来检查这些3-项集会访问哪些叶节点。

{1,3,4} 将会走到L5

{1,3,5} 将会走到L5

{1,3,8} 将会走到L5

{1,4,5} 将会走到L1

{1,4,8} 将会走到L1

{1,5,8} 将会走到L3

{3,4,5} 将会走到L9

{3,4,8} 将会走到L9

{3,5,8} 将会走到L11

{4,5,8} 将会走到L3

所以，会访问到**L1 L3 L5 L9 L11**这些节点。

(b) 接下来，我们需要检查在步骤(a)中确定的叶节点（L1, L3, L5, L9, L11）来找出哪些候选项集实际包含在交易{1,3,4,5,8}中。

由图可知：

L1 包含 **{145}** 和 {178}

L3 包含 {125}、**{158}** 和 **{458}**

L5 包含 {168}

L9 包含 {346}、{379} 和 {678}

L11 包含 {356}、{689}

所以最终的答案是**{145}，{158}，{458}。**

