

## TD n° 2

### Règles d'association

**Exercice 1** Le but de cet exercice est de dérouler l'algorithme **Apriori** sur les données d'une épicerie suivante, en utilisant comme seuils de support et de confiance  $minsup=33,3\%$  et  $minconf=60\%$ .

TID	Items
T1	Fromage, Pain, Ketchup
T2	Fromage, Pain
T3	Fromage, Coca, Chips
T4	Chips, Coca
T5	Chips, Ketchup
T6	Fromage, Coca, Chips

1. Calculez d'abord les ensembles de k-itemsets candidats et fréquents, pour chaque étape de l'algorithme.
2. Montrez ensuite l'ensemble des règles d'associations générées, en les triant par confiance décroissante.

**Exercice 2** Soit l'ensemble d'items  $I = \{A, B, C, D, E\}$ . Supposons que l'algorithme **Apriori** soit appliquée sur les données ci-dessous avec le seuil  $minsup=30\%$ .

- Dessinez le treillis des itemsets. Pour chaque noeud du treillis, joignez lui la lettre suivante, selon le cas :
  - **N** : si l'itemset ne fait pas partie des itemsets candidats (s'il n'est jamais généré par Apriori).
  - **F** : si c'est un itemset fréquent.
  - **R** : si c'est un itemset candidat, qui s'est révélé non-fréquent (itemset rare).

TID	Items
T1	A, B, D, E
T2	B, C, D
T3	A, B, D, E
T4	A, C, D, E
T5	B, C, D, E
T6	B, D, E
T7	C, D
T8	A, B, C
T9	A, D, E
T10	B, D

- Quel est le pourcentage d'itemsets qui sont fréquents ?
- Quel est le «taux d'élagage» d'Apriori sur ces données ( le pourcentage d'itemsets non candidats) ?
- Quel est le «taux de rejet» (le pourcentage d'itemsets non-fréquents parmi les candidats) ?