

## Fouille de Données - Extractions de motifs

M2

## 1 Extraction des ensembles fréquents

Soit la base de données de transactions représentée dans la figure 1.

<i>Id</i>	items
1	{ <i>d, e, f</i> }
2	{ <i>b, c, e</i> }
3	{ <i>a, b, c, e</i> }
4	{ <i>b, c, e</i> }
5	{ <i>a, b, c, e</i> }
6	{ <i>b, c, e</i> }

TABLE 1 – Une base de données de transactions

1. Extraire les motifs fréquents, dont le support est supérieur ou égal à 2
2. Après avoir identifier les fermés et les générateurs minimaux de support supérieur ou égal à 2, les énumérer directement.

## 2 Fouille de motifs sous contraintes

Soit la base de transactions suivante.

TID	Transactions
$T_1$	a,b,c,d,f
$T_2$	b,c,d,e,g
$T_3$	a,c,d,f
$T_4$	a,b,c,e,g
$T_5$	c,d,f,h

item	prix
a	10
b	21
c	15
d	12
e	30
f	15
g	22
h	101

1. On souhaite extraire tous les motifs  $X$  qui apparaissent dans au moins deux transactions ( $support(X) \geq 2$ ) et dont la somme est inférieure à 40 ( $sum(X) < 40$ ).
  - (a) Quel est le type de cette contrainte ?
  - (b) Faire l'extraction.
  - (c) Y a-t-il des règles ayant une confiance de 100% ?
2. On souhaite extraire tous les motifs  $X$  qui apparaissent dans au moins deux transactions ( $support(X) \geq 2$ ) et dont le prix moyen est supérieur à 24 ( $average(X) > 24$ ).
  - (a) Quel est le type de cette contrainte ?
  - (b) Essayez de la convertir en une contrainte anti-monotone, puis faire l'extraction.