

# Association Clusters Example

$q = A+B$  (query item)

keyword association rule  
 $\{B,D,C\} \rightarrow A$

$d_1 = A,A,B,D$

$d_2 = B,A,C,C,D$

$d_3 = A,B$

$d_4 = B,C,D$

$d_5 = D$

$d_6 = A,B,D$

$d_7 = B,B,A$

|     | $d_1$ | $d_2$ | $d_3$ | $d_4$ | $d_5$ | $d_6$ | $d_7$ |
|-----|-------|-------|-------|-------|-------|-------|-------|
| $A$ | 2     | 1     | 1     | 0     | 0     | 1     | 1     |
| $B$ | 1     | 1     | 1     | 1     | 0     | 1     | 2     |
| $C$ | 0     | 2     | 0     | 1     | 0     | 0     | 0     |
| $D$ | 1     | 1     | 0     | 1     | 1     | 1     | 0     |

$$C_{u,v} = \sum_{dj \in DI} f_{s_u,j} \times f_{s_v,j}$$

$$\begin{aligned}
 C_{1,4} &= (f_{1,1} * f_{4,1}) + (f_{1,2} * f_{4,2}) + (f_{1,3} * f_{4,3}) + (f_{1,4} * f_{4,4}) + (f_{1,5} * f_{4,5}) + (f_{1,6} * f_{4,6}) + (f_{1,7} * f_{4,7}) \\
 &= 2*1 + 1*1 + 1*0 + 0*1 + 0*1 + 1*1 + 1*0 \\
 &= 4
 \end{aligned}$$

ค. สัมพันธ์ระหว่าง A กับ D คือ 4

Step 1

พหุน. (C) keyword + ฟิลเตอร์

# Association Clusters Example

Correlation Matrix (C)

|   | A | B | C | D |
|---|---|---|---|---|
| A | 8 | 7 | 2 | 4 |
| B | 7 | 9 | 3 | 4 |
| C | 2 | 3 | 5 | 3 |
| D | 4 | 4 | 3 | 5 |

② ฟิลเตอร์

พหุน.

①

$$C_{u,v} = \sum_{dj \in DI} f_{s_u,j} \times f_{s_v,j}$$

$$C_{1,4} = (f_{1,1} * f_{4,1}) + (f_{1,2} * f_{4,2}) + (f_{1,3} * f_{4,3}) + (f_{1,4} * f_{4,4}) + (f_{1,5} * f_{4,5}) + (f_{1,6} * f_{4,6}) + (f_{1,7} * f_{4,7})$$

$$= 2*1 + 1*1 + 1*0 + 0*1 + 0*1 + 1*1 + 1*0$$

$$= 4$$

พหุน.สัมพันธ์ระหว่าง A กับ D คือ 4

Step 2

Normalize

# Association Clusters Example

## Normalized Correlation Matrix (S)

เพื่อหาความสัมพันธ์

ก. key u กับ key v

$$S_{u,v} = \frac{C_{u,v}}{C_{u,u} + C_{v,v} - C_{u,v}}$$

ก. key u
ก. key v
ก. key u กับ key v

C item 1, item 2

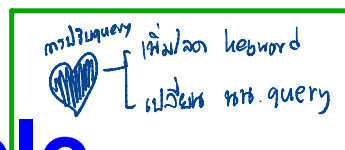
$$s_{1,2} = \frac{C_{1,2}}{C_{1,1} + C_{2,2} - C_{1,2}} = \frac{7}{8 + 9 - 7} = 0.70$$

|   | A | B | C | D |
|---|---|---|---|---|
| A | 8 | 7 | 2 | 4 |
| B | 7 | 9 | 3 | 4 |
| C | 2 | 3 | 5 | 3 |
| D | 4 | 4 | 3 | 5 |

Step 3

คล้าย Query

\*



# Association Clusters Example

## Normalized Correlation Matrix

ตัวสีเทา คือค่าที่ตรงกับตัวแถวที่ 1

|   | A    | B    | C    | D    |
|---|------|------|------|------|
| A | 1    | 0.70 | 0.18 | 0.44 |
| B | 0.70 | 1    | 0.27 | 0.40 |
| C | 0.18 | 0.27 | 1    | 0.43 |
| D | 0.44 | 0.40 | 0.43 | 1    |

Take u-th row

เลือก A, A, B, B, C, C, D, D

Return the set of n **largest values**  $s_{u,v}$  ( $u \neq v$ )

## Term Relation

สำหรับ A ที่อยู่ใน B เพราะ B มีค่าสูง (ไม่มองตัวแถว)

B, A

1. {A, B}

2. {B, A}

C, D

3. {C, D}

D, A

4. {D, A}

1

มี  $k_i$  ... ที่อยู่ใน  $k_j$

## Original Query

$$q = A + B$$

2

กำหนด query ใหม่

## New Query

$$\begin{aligned} q' &= (A + 0.7B) + (0.7A + B) \\ &= 1.7A + 1.7B \\ &= A + B \end{aligned}$$