

6  
4/7/25

หาว่า keyword เหมือน/คล้ายกับ keyword ที่ไหน

# Scalar Clusters

keyword เหมือน/คล้ายกับ keyword ที่ไหน

- Idea
  - two stems with similar neighborhoods have some **synonymity relationships**

- Definition

- $c_{u,v} = c(k_u, k_v)$
- vectors of correlation values for stem  $k_u$  and  $k_v$

$$\vec{s}_u = (c_{u,1}, c_{u,2}, \dots, c_{u,t})$$

keyword 1    keyword 2

$$\vec{s}_v = (c_{v,1}, c_{v,2}, \dots, c_{v,t})$$

- scalar association matrix

$$S_{u,v} = \frac{\vec{s}_u \bullet \vec{s}_v}{|\vec{s}_u| \times |\vec{s}_v|}$$

- scalar clusters

- the set of  $k$  **largest values** of scalar association

Database , Math , Set  
Tree, Water , Fertilizer  
Flower, Letter , Lover

Step 1

# หาความสัมพันธ์ $S_{i,j}$ ( $C_{i,keyword_1}, C_{i,keyword_2}, \dots$ )

Scalar Clusters

Association / Metric

$$\vec{s}_u = (c_{u,1}, c_{u,2}, \dots, c_{u,t})$$

$$\vec{s}_v = (c_{v,1}, c_{v,2}, \dots, c_{v,t})$$

$$\vec{s}_1 = (c_{1,1}, c_{1,2}, \dots, c_{1,t})$$

$$\vec{s}_3 = (c_{3,1}, c_{3,2}, \dots, c_{3,t})$$

$$\vec{s}_1 = (c_{1,1}, c_{1,2}, c_{1,3}) = (5, 6, 1)$$

$C_{\text{Database, Algebra}}, C_{\text{Database, Math}}, C_{\text{Database, Set}}$  (keyword)

$$\vec{s}_2 = (c_{2,1}, c_{2,2}, c_{2,3}) = (6, 9, 0)$$

$C_{\text{AI, Algebra}}, C_{\text{AI, Math}}, C_{\text{AI, Set}}$

$$\vec{s}_3 = (c_{3,1}, c_{3,2}, c_{3,3}) = (1, 0, 2)$$

$C_{\text{Network, Algebra}}, C_{\text{Network, Math}}, C_{\text{Network, Set}}$



ได้ network - ความสัมพันธ์อีกกลุ่ม

Network = {Set(2), Algebra (1), Math(0)} \*\*\*idea

Step 2

Normalize

# Scalar Clusters

Normalize

$$\vec{s}_u = (c_{u,1}, c_{u,2}, \dots, c_{u,t})$$

$$\vec{s}_1 = (c_{1,1}, c_{1,2}, \dots, c_{1,t})$$

$$\vec{s}_1 = (c_{1,1}, c_{1,2}, c_{1,3}) = (5, 6, 1)$$

$$\vec{s}_2 = (c_{2,1}, c_{2,2}, c_{2,3}) = (6, 9, 0)$$

$$\vec{s}_3 = (c_{3,1}, c_{3,2}, c_{3,3}) = (1, 0, 2)$$

1 unnormalized  $\sqrt{5^2 + 6^2 + 1^2}$

$$|\vec{s}_1| = \sqrt{25 + 36 + 1} = 7.874$$

$$|\vec{s}_2| = \sqrt{36 + 81 + 0} = 10.817$$

$$|\vec{s}_3| = \sqrt{1 + 0 + 4} = 2.236$$

$$\vec{s}_v = (c_{v,1}, c_{v,2}, \dots, c_{v,t})$$

$$\vec{s}_3 = (c_{3,1}, c_{3,2}, \dots, c_{3,t})$$

$$S_{u,v} = \frac{\vec{s}_u \bullet \vec{s}_v}{|\vec{s}_u| \times |\vec{s}_v|}$$

$$S_{1,3} = \frac{\vec{s}_1 \bullet \vec{s}_3}{|\vec{s}_1| \times |\vec{s}_3|} = \frac{(5 \times 1) + (6 \times 0) + (1 \times 2)}{7.874 \times 2.236}$$

$$S_{1,3} = \frac{7}{7.874 \times 2.236} = 0.398$$

Step 3

# 181 0. Stem + ឯកសារ Query 9.2 Scalar Clusters

### Normalized Correlation Matrix (S)

$S$	$S_1$	$S_2$	$S_3$
$S_1$	1	0.986	0.398
$S_2$	0.986	1	0.248
$S_3$	0.398	0.248	1

1) ឯកសារ Stem

ឯកសារ  $S_i$  គឺជាឯកសារ  $S_j$  ដែលមានទំនាក់ទំនង

#### Stem Relation

1.  $\{S_1, S_2\}$
2.  $\{S_2, S_1\}$
3.  $\{S_3, S_1\}$

2) កំណត់ ឯកសារ Query 9.2

#### Original Query

$$q = 3S_1 + S_3$$

Database

Network

#### New Query

$$\begin{aligned} q' &= 3*(S_1 + 0.986S_2) + (0.398S_1 + S_3) \\ &= 3.398S_1 + 2.958S_2 + S_3 \end{aligned}$$

កំណត់ ឯកសារ 9.2