

โจทย์

Step 0 กำหนดรากศัพท์ +
ที่ทราบความหมาย

Answer

เลือกใช้ Metric Cluster เนื่องจากต้องการปรับเปลี่ยน Query ดั้งเดิม โดยใช้เอกสารที่ถูกส่งออกมาและพิจารณาเฉพาะระยะห่างของ Keyword และโจทย์ไม่ต้องการให้ใช้วิธีอื่น

Metric Clusters Example

มีรากศัพท์เข้ามาเกี่ยวข้อง

รากศัพท์ คืออะไร

$$q = A + 2D$$

$$k_n = A, B, C, D, E, F$$

A, B, C base on S_1 stem
 D, E base on S_2 stem
 F base on S_3 stem

Then

รากศัพท์ที่ 1 มีมา 2 keyword

$$V(S_1) = \{A, B, C\}$$

$$V(S_2) = \{D, E\}$$

$$V(S_3) = \{F\}$$

① คัดรากศัพท์

② ตัดรากศัพท์ "สำคัญ" จากทุก Doc

ทุกเอกสารที่ return ออกมา ... keyword ... keyword ... เท่าไร (บอกค่าจุด)

ระยะห่าง	A	B	C	D	E	F
A	0	5	∞	∞	1	2
B	5	0	3	2	1	1
C	∞	3	0	3	4	∞
D	∞	2	3	0	∞	5
E	1	1	4	∞	0	1
F	2	1	∞	5	1	0

เมื่อใช้งานคำเรียกค้น Q = 7Dog-2Cat+5Bird ระบบส่งผลลัพธ์ออกมาคือเอกสาร 5 เอกสารดังนี้

D1: "Dog is a animal like to fight cat that same tiger. Cats eat fish. Dogs eat ham"

D2: "Birds fly over tiger. A Dog run to catch a Bird. A cat smile beside the river"

D3: "A tiger is crying for a lost bird. A cat is being hungry while the dog eat all fishs. That cat complain to its friend"

D4: "All birds sing the song. A dog is flighting the cat. The cat crying and go to complain with mother cat"

D5: "???"

*** เอกสารหมายเลข 5 ในขั้น.ศ.เป็นประโยคเองตามใจชอบ โดยให้มี Keyword อย่างน้อย 2 Keyword และห้ามเป็นประโยคเดียวกับประโยคที่ใช้ในการเรียนการสอนวิชานี้

โดย Dog,Cat,Tiger,Bird คือ Keyword ของระบบ

หากต้องการปรับคำเรียกค้น (Query) ให้มีผลลัพธ์มีความตรงประเด็นมากยิ่งขึ้น โดยโจทย์กำหนดว่าห้ามใช้ Associate Cluster และ Scalar Cluster นักศึกษาควรใช้โมเดลใดเพราะอะไร จงแสดงวิธีคำนวณในโมเดลที่เลือกใช้ (40คะแนน)

หาความสัมพันธ์ระหว่าง Keyword โดยพิจารณาระยะห่าง

Doc1
Keyword → {Cat, Dog, Tiger}
หาระยะห่าง
Cat, Dog = 3 Dog, Tiger = 4
Cat, Tiger = 1

Doc2
Keyword → {Bird, Cat, Dog, Tiger}
หาระยะห่าง
Bird, Cat = 2 Cat, Dog = 7
Bird, Dog = 5 Cat, Tiger = 9
Bird, Tiger = 3 Dog, Tiger = 2

Doc3
Keyword → {Bird, Cat, Dog, Tiger}
หาระยะห่าง
Bird, Cat = 2 Cat, Dog = 5 Dog, Tiger = 14
Bird, Dog = 8 Cat, Tiger = 8
Bird, Tigger = 6

Doc4
Keyword → {Bird, Cat, Dog}
หาระยะห่าง
Bird, Cat = 9 Cat, Dog = 4
Bird, Dog = 5

Doc5
Keyword → {Cat, Dog}
หาระยะห่าง
Cat, Dog = 6

เลือกระยะห่างที่สั้นที่สุด ได้ข้อมูลระยะห่างตามตาราง

	Bird	Cat	Dog	Tiger
Bird	0	2	5	3
Cat	2	0	3	1
Dog	5	3	0	2
Tiger	3	1	2	0

step 1 $พหุพ. = \frac{1}{ระยะห่าง}$ ความสัมพันธ์ระหว่าง Keyword ($\frac{1}{Distance}$)

Metric Clusters Example

เอา 1 ตัว มาด้วย
↓
ทุกตัวในแถว

ระยะ ห่าง	A	B	C	D	E	F
A	0	5	∞	∞	1	2
B	5	0	3	2	1	1
C	∞	3	0	3	4	∞
D	∞	2	3	0	∞	5
E	1	1	4	∞	0	1
F	2	1	∞	5	1	0



	A	B	C	D	E	F
A	ไม่ติดด้วย	0.20	0	0	1	0.50
B	0.20	-	0.33	0.50	1	1
C	0	0.33	-	0.33	0.25	0
D	0	0.50	0.33	-	0	0.20
E	1	1	0.25	0	-	1
F	0.50	1	0	0.20	1	-

step 2

หา ค. จว ใกล้เคียง

Metric Clusters Example

9 ขั้นตอน

๑. หาพหุคูณของ ใกล้เคียง S_1 กับ S_2
 $A \times D, A \times E, B \times D, B \times E, C \times D, C \times E$

$$V(S_1) = \{A, B, C\}$$

$$V(S_2) = \{D, E\}$$

$$V(S_3) = \{F\}$$

	A	B	C	D	E	F
A	-	0.20	0	0	1	0.50
B	0.20	-	0.33	0.50	1	1
C	0	0.33	-	0.33	0.25	0
D	0	0.50	0.33	-	0	0.20
E	1	1	0.25	0	-	1
F	0.50	1	0	0.20	1	-

Correlation Matrix (C)

	S_1	S_2	S_3
S_1	0	3.08	1.50
S_2	3.08	0	1.20
S_3	1.50	1.20	0

$$C_{u,v} = \sum_{ki \in V(su)} \sum_{kj \in V(sv)} \frac{1}{r(k_i, k_j)}$$

สูตร ①

$$V(S_1) = \{A, B, C\}$$

$$V(S_2) = \{D, E\}$$

② ขั้นตอน

หาผลรวม

$$\begin{aligned}
 C_{1,2} &= c(A, D) + c(A, E) + c(B, D) + c(B, E) + c(C, D) + c(C, E) \\
 &= 0 + 1 + 0.50 + 1 + 0.33 + 0.25 \\
 &= 3.08
 \end{aligned}$$

Step 3 Normalize ค.จากศัพท์

Metric Clusters Example

Normalized Correlation Matrix (S)

	S_1	S_2	S_3
S_1	0	3.08	1.50
S_2	3.08	0	1.20
S_3	1.50	1.20	0

$V(S_1) = \{A, B, C\} = 3$
 $V(S_2) = \{D, E\} = 2$
 $V(S_3) = \{F\} = 1$

$$s_{u,v} = \frac{c_{u,v}}{|V(s_u)| \times |V(s_v)|}$$

จำนวนสัมพันธ์ระหว่างคำศัพท์
 จากคำศัพท์ (มี keyword ที่ซ้ำ)

$$s_{2,3} = \frac{c_{2,3}}{|V(s_2)| \times |V(s_3)|} = \frac{1.2}{2 \times 1} = 0.6$$

Step 4

เลือก query

Metric Clusters Example

Normalized Correlation Matrix (S)

	S_1	S_2	S_3
S_1	0	0.51	0.50
S_2	0.51	0	0.60
S_3	0.50	0.60	0

Stem Relation

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

① มีค่าที่พหุ A
คือค่าที่พหุ B

ไม่ซ้ำกัน

Original Query

$$q = A + 2D \rightarrow \text{เลือกค่า 2}$$

เลือกค่าที่ 1

② เลือก query

New Query

$$q' = (S_1 + 0.51S_2) + 2*(S_2 + 0.60S_3) \\ = S_1 + 2.51S_2 + 1.2S_3$$