

ข้อ 4. จากการส่ง Query เรียกค้น $Q = 3\text{Dog}-2\text{Cat}$ เข้าไปในระบบ ส่งผลลัพธ์ออกมาคือเอกสาร 3 เอกสารดังนี้

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

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

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. The puppy is so hungry."

โดย Bird, Cat, Dog, Kitty(ลูกแมว), Puppy(ลูกสุนัข), Tiger คือ Keyword ของระบบ

4.1 เพื่อให้ได้คำตอบในคำถาม 4.2 จงให้เหตุผลว่า"ควรเลือกใช้วิธีการปรับ Query แบบอัตโนมัติ วิธีใด
จึงอธิบาย"

A) Global Analysis B) Associate Cluster C) Metric Cluster D) Scalar Cluster

4.2 ให้นักศึกษาปรับ Query ให้เหมาะสมตามหลักการของ Automatic Analysis โดยพิจารณาเฉพาะ
ระยะห่างของ Keyword

(30 คะแนน) ข้อสอบ 2/2559

ข้อ 4. จากการส่ง Query เรียกค้น $Q = 3\text{Dog}-2\text{Cat}$ เข้าไปในระบบ ส่งผลลัพธ์ออกมาคือเอกสาร 3 เอกสารดังนี้

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

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

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. The puppy is so hungry."

โดย Bird, Cat, Dog, Kitty(ลูกแมว), Puppy(ลูกสุนัข), Tiger คือ Keyword ของระบบ

4.1 เพื่อให้ได้คำตอบในคำถาม 4.2 จงให้เหตุผลว่า"ควรเลือกใช้วิธีการปรับ Query แบบอัตโนมัติ วิธีใด
จึงอธิบาย"

A) Global Analysis B) Associate Cluster C) Metric Cluster D) Scalar Cluster

4.2 ให้นักศึกษาปรับ Query ให้เหมาะสมตามหลักการของ Automatic Analysis โดยพิจารณาเฉพาะ
ระยะห่างของ Keyword

(30 คะแนน) ข้อสอบ 2/2559

Answer

4.1 เลือกใช้ Metric Cluster เนื่องจากต้องการปรับเปลี่ยน Query ดั้งเดิม โดยใช้เอกสารที่ถูกส่งออกมาและพิจารณาเฉพาะ
ระยะห่างระหว่าง Keyword

4.2

หาความสัมพันธ์ระหว่าง Keyword โดยพิจารณาระยะห่าง

Doc1

Keyword \rightarrow {Cat, Dog, Puppy, Tiger}

หาระยะห่าง

Cat, Dog = 3	Dog, Puppy = 3
Cat, Puppy = 6	Dog, Tiger = 6
Cat, Tiger = 1	Puppy, Tiger = 7

Doc2

Keyword \rightarrow {Bird, Cat, Dog, Kitty, Tiger}

หาระยะห่าง

Bird, Cat = 2	Cat, Dog = 7	Dog, Tiger = 2
Bird, Dog = 5	Cat, Kitty = 5	Kitty, Tiger = 14
Bird, Kitty = 7	Cat, Tiger = 9	
Bird, Tiger = 3	Dog, Kitty = 12	

Doc3

Keyword \rightarrow {Bird, Cat, Dog, Puppy, Tiger}

หาระยะห่าง

Bird, Cat = 2	Cat, Dog = 5	Dog, Tiger = 14
Bird, Dog = 8	Cat, Puppy = 6	Puppy, Tiger = 25
Bird, Puppy = 19	Cat, Tiger = 8	
Bird, Tiger = 6	Dog, Puppy = 11	

เลือกกระยะห่างที่สั้นที่สุด ได้ข้อมูลระยะห่างตามตาราง

	Bird	Cat	Dog	Kitty	Puppy	Tiger
Bird	0	2	5	7	19	3
Cat	2	0	3	5	6	1
Dog	5	3	0	12	3	2
Kitty	7	5	12	0	∞	14
Puppy	19	6	3	∞	0	7
Tiger	3	1	2	14	7	0

ความสัมพันธ์ระหว่าง Keyword ($\frac{1}{Distance}$)

C	Bird	Cat	Dog	Kitty	Puppy	Tiger
Bird	-	0.500	0.200	0.143	0.053	0.333
Cat	0.500	-	0.333	0.200	0.167	1.000
Dog	0.200	0.333	-	0.083	0.333	0.500
Kitty	0.143	0.200	0.083	-	0.000	0.071
Puppy	0.053	0.167	0.333	0.000	-	0.143
Tiger	0.333	1.000	0.500	0.071	0.143	-

ความสัมพันธ์ระหว่างรากศัพท์ Stem

V1 = {Bird}	Size = 1
V2 = {Cat,Kitty}	Size = 2
V3 = {Dog,Puppy }	Size = 2
V4 = {Tiger}	Size = 1

$$c(k_u, k_v) = \sum_{i \in V(k_u)} \sum_{j \in V(k_v)} \frac{1}{r(k_i, k_j)}$$

$$\begin{aligned} C(1,2) &= C(Bird, Cat) + C(Bird, Kitty) \\ &= 0.500 + 0.143 = 0.643 \end{aligned}$$

$$\begin{aligned} C(1,3) &= C(Bird, Dog) + C(Bird, Puppy) \\ &= 0.200 + 0.053 = 0.253 \end{aligned}$$

$$C(1,4) = C(Bird, Tiger) = 0.333$$

$$\begin{aligned} C(2,3) &= C(Cat, Dog) + C(Cat, Puppy) + C(Kitty, Dog) + C(Kitty, Puppy) \\ &= 0.333 + 0.167 + 0.083 + 0.000 \\ &= 0.583 \end{aligned}$$

$$\begin{aligned} C(2,4) &= C(Cat, Tiger) + C(Kitty, Tiger) \\ &= 1 + 0.071 = 1.071 \end{aligned}$$

$$\begin{aligned} C(3,4) &= C(Dog, Tiger) + C(Puppy, Tiger) \\ &= 0.500 + 0.143 = 0.643 \end{aligned}$$

ความสัมพันธ์ระหว่าง รากศัพท์ Stem

C	s1	s2	s3	s4
s1	–	0.643	0.253	0.333
s2	0.643	–	0.583	1.071
s3	0.253	0.583	–	0.643
s4	0.333	1.071	0.643	–

Normalize ความสัมพันธ์ระหว่าง รากศัพท์

$$s(k_u, k_v) = \frac{c(k_u, k_v)}{|V(k_u)| \times |V(k_v)|}$$

V1 = {Bird}	Size = 1
V2 = {Cat, Kitty}	Size = 2
V3 = {Dog, Puppy }	Size = 2
V4 = {Tiger}	Size = 1

$$s(k_u, k_v) = \frac{c(k_u, k_v)}{|V(k_u)| \times |V(k_v)|}$$

$$\begin{aligned} S(V1, V2) &= \frac{C_{1,2}}{|V1||V2|} \\ &= \frac{0.643}{1 \times 2} = 0.322 \end{aligned}$$

$$S(V1, V3) = \frac{0.253}{1 \times 2} = 0.127$$

$$S(V1, V4) = \frac{0.333}{1 \times 1} = 0.333$$

$$S(V2, V3) = \frac{0.583}{2 \times 2} = 0.146$$

$$S(V2, V4) = \frac{1.071}{2 \times 1} = 0.536$$

$$S(V3, V4) = \frac{0.643}{2 \times 1} = 0.322$$

C	S1	S2	S3	S4
S1	–	0.643	0.253	0.333
S2	0.643	–	0.583	1.071
S3	0.253	0.583	–	0.643
S4	0.333	1.071	0.643	–

V1 = {Bird}

Size = 1

V2 = {Cat, Kitty}

Size = 2

V3 = {Dog, Puppy }

Size = 2

V4 = {Tiger}

Size = 1

ตาราง Normalize ความสัมพันธ์ระหว่างรากศัพท์

N	S1	S2	S3	S4
S1	-	0.322	0.127	0.333
S2	0.322	-	0.146	0.536
S3	0.127	0.146	-	0.322
S4	0.333	0.536	0.322	-

ความสัมพันธ์ระหว่างรากศัพท์

$S1 \rightarrow S4$

$S2 \rightarrow S4$

$S3 \rightarrow S4$

$S4 \rightarrow S2$

Original Query:

$$3\text{Dog} - 2\text{Cat} = 3S_3 - 2S_2$$

New Query:

$$\begin{aligned}
 3S_3 - 2S_2 &= 3(S_3 + 0.322S_4) - 2(S_2 + 0.536S_4) \\
 &= -2S_2 + 3S_3 - 0.107S_4
 \end{aligned}$$