

ข้อ 2 สมมติในระบบมีเอกสารทั้งหมด 10 เอกสารดังนี้ (bird, cat, dog, tiger คือ Keyword)

- D1: {bird, cat, bird, cat, dog, dog, bird}
- D2: {cat, tiger, cat, dog}
- D3: {dog, bird, bird}
- D4: {cat, tiger}
- D5: {tiger, tiger, dog, tiger, cat}
- D6: {cat, cat, tiger, tiger}
- D7: {bird, cat, dog}
- D8: {dog, cat, bird}
- D9: {cat, dog, tiger}
- D10: {tiger, cat, tiger}

ผู้ใช้ส่งคำเรียกค้น "cat cat cat dog dog tiger" เข้าไปในระบบ จงตอบคำถาม

2.1 เพื่อให้ได้คำตอบในคำถาม 2.2 ผู้ใช้สามารถเลือกใช้โมเดลใดได้บ้างเพราะอะไร (เลือกได้เฉพาะตัวเลือกที่ให้มา)

- | | |
|-------------------------|-----------------------------|
| A) BM25 Model | D) Boolean Model |
| B) Fuzzy Model | E) Probabilistic Model |
| C) Extend Boolean Model | F) Generalized Vector Model |

2.2 ให้นักศึกษาแสดงวิธีคำนวณหา Ranking ของเอกสารทุกเอกสารในระบบ ตามที่ผู้ใช้ต้องการ โดยให้ความสำคัญกับลักษณะการปรากฏของคีย์เวิร์ดในแต่ละเอกสาร กล่าวคือเอกสารที่มีลักษณะการปรากฏของคีย์เวิร์ดที่เหมือนกันจะต้องมีความตรงประเด็นที่ใกล้เคียงกัน

2.3 หากระบบกำหนดให้เอกสารที่ 1 ตรงประเด็นมากกว่าเอกสารที่ 9 โมเดลที่เลือกมาให้คำตอบถูกต้องหรือไม่ ถ้าผิดต้องแก้ไขอย่างไรจงอธิบาย(35 คะแนน)

Answer

2.1 ใช้ Generalized Vector Model เนื่องจากโจทย์กำหนดเอกสารที่มีลักษณะการปรากฏของคีย์เวิร์ดที่เหมือนกันจะต้องมีความตรงประเด็นที่ใกล้เคียงกัน ดังนั้นรูปแบบการปรากฏของ Keyword จึงมีผลต่อความตรงประเด็น และโจทย์ไม่ได้กำหนดให้ Keyword ไม่สัมพันธ์กันหรือไม่ แสดงว่า Keyword อาจมีความสัมพันธ์กันได้

ขั้นตอนที่ 1

เอกสาร 10 เอกสารมีการแจกแจง Keyword ดังนี้

D1: {bird, cat, bird, cat, dog, dog, bird}

D2: {cat, tiger, cat, dog}

D3: {dog, bird, bird}

D4: {cat, tiger}

D5: {tiger, tiger, dog, tiger, cat}

D6: {cat, cat, tiger, tiger}

D7: {bird, cat, dog}

D8: {dog, cat, bird}

D9: {cat, dog, tiger}

D10: {tiger, cat, tiger}

	Bird	Cat	Dog	Tiger	Max
Doc1	3	2	2	0	3
Doc2	0	2	1	1	2
Doc3	2	0	1	0	2
Doc4	0	1	0	1	1
Doc5	0	1	1	3	3
Doc6	0	2	0	2	2
Doc7	1	1	1	0	1
Doc8	1	1	1	0	1
Doc9	0	1	1	1	1
Doc10	0	1	0	2	2
n	4	9	7	6	

ขั้นตอนที่ 1

Doc1 Only

$$tf_{bird} = \frac{3}{3} = 1.000$$

$$tf_{cat} = \frac{2}{3} = 0.667$$

$$tf_{dog} = \frac{2}{3} = 0.667$$

$$tf_{tiger} = \frac{0}{3} = 0.000$$

$$idf_{bird} = \log\left(\frac{10}{4}\right) = 0.398$$

$$idf_{cat} = \log\left(\frac{10}{9}\right) = 0.046$$

$$idf_{dog} = \log\left(\frac{10}{7}\right) = 0.155$$

$$idf_{tiger} = \log\left(\frac{10}{6}\right) = 0.222$$

	Bird	Cat	Dog	Tiger	Max
Doc1	3	2	2	0	3
Doc2	0	2	1	1	2
Doc3	2	0	1	0	2
Doc4	0	1	0	1	1
Doc5	0	1	1	3	3
Doc6	0	2	0	2	2
Doc7	1	1	1	0	1
Doc8	1	1	1	0	1
Doc9	0	1	1	1	1
Doc10	0	1	0	2	2
n	4	9	7	6	

$$w_{bird} = 1.000 * 0.398 = 0.398$$

$$w_{cat} = 0.667 * 0.046 = 0.031$$

$$w_{dog} = 0.667 * 0.155 = 0.103$$

$$w_{tiger} = 0.000 * 0.222 = 0.000$$

ขั้นตอนที่ 1

น้ำหนักของแต่ละ **Keyword** ในแต่ละเอกสาร

	Bird	Cat	Dog	Tiger
Doc1	0.398	0.031	0.103	0.000
Doc2	0.000	0.046	0.077	0.111
Doc3	0.398	0.000	0.077	0.000
Doc4	0.000	0.046	0.000	0.222
Doc5	0.000	0.015	0.052	0.222
Doc6	0.000	0.046	0.000	0.222
Doc7	0.398	0.046	0.155	0.000
Doc8	0.398	0.046	0.155	0.000
Doc9	0.000	0.046	0.155	0.222
Doc10	0.000	0.023	0.000	0.222

ขั้นตอนที่ 1

Query = "cat cat cat dog dog tiger"

$$W_{i,q} = (0.5 + \frac{0.5 * freq_{i,q}}{Max(freq_{i,q})}) * \log(\frac{N}{n_i})$$

$$W_{Cat,q} = (0.5 + \frac{0.5 * 3}{3}) * \log(\frac{10}{9}) = 0.523$$

$$W_{Dog,q} = (0.5 + \frac{0.5 * 2}{3}) * \log(\frac{10}{7}) = 0.552$$

$$W_{Tiger,q} = (0.5 + \frac{0.5 * 1}{3}) * \log(\frac{10}{6}) = 0.537$$

	Bird	Cat	Dog	Tiger
Doc1	0.398	0.031	0.103	0.000
Doc2	0.000	0.046	0.077	0.111
Doc3	0.398	0.000	0.077	0.000
Doc4	0.000	0.046	0.000	0.222
Doc5	0.000	0.015	0.052	0.222
Doc6	0.000	0.046	0.000	0.222
Doc7	0.398	0.046	0.155	0.000
Doc8	0.398	0.046	0.155	0.000
Doc9	0.000	0.046	0.155	0.222
Doc10	0.000	0.023	0.000	0.222
Q	0	0.523	0.552	0.537

ขั้นตอนที่ 2

Query = "cat cat cat dog dog tiger"

	Bird	Cat	Dog	Tiger	
Doc1	0.398	0.031	0.103	0.000	m1
Doc2	0.000	0.046	0.077	0.111	m2
Doc3	0.398	0.000	0.077	0.000	m3
Doc4	0.000	0.046	0.000	0.222	m4
Doc5	0.000	0.015	0.052	0.222	m2
Doc6	0.000	0.046	0.000	0.222	m4
Doc7	0.398	0.046	0.155	0.000	m1
Doc8	0.398	0.046	0.155	0.000	m1
Doc9	0.000	0.046	0.155	0.222	m2
Doc10	0.000	0.023	0.000	0.222	m4
Q	0	0.523	0.552	0.537	

$m1=(1, 1, 1, 0)$

$m2=(0, 1, 1, 1)$

$m3=(1, 0, 1, 0)$

$m4=(0, 1, 0, 1)$

ขั้นตอนที่ 2

$$k_i = \frac{\sum_{\forall r, g_i(m_r)} c_{i,r} m_r}{\sqrt{\sum_{\forall r, g_i(m_r)} c_{i,r}^2}}$$

$$k_1 = \frac{c_{1,1}m_1 + c_{1,3}m_3}{\sqrt{c_{1,1}^2 + c_{1,3}^2}}$$

$$c_{1,1} = w_{1,1} + w_{1,7} + w_{1,8} = 0.398 + 0.398 + 0.398 = 1.194$$

$$c_{1,3} = w_{1,3} = 0.398$$

$$k_1 = \frac{1.194m_1 + 0.398m_3}{\sqrt{0.194^2 + 0.398^2}}$$

$$k_1 = 0.949m_1 + 0.316m_3$$

	Bird	Cat	Dog	Tiger	
Doc1	0.398	0.031	0.103	0.000	m1
Doc2	0.000	0.046	0.077	0.111	m2
Doc3	0.398	0.000	0.077	0.000	m3
Doc4	0.000	0.046	0.000	0.222	m4
Doc5	0.000	0.015	0.052	0.222	m2
Doc6	0.000	0.046	0.000	0.222	m4
Doc7	0.398	0.046	0.155	0.000	m1
Doc8	0.398	0.046	0.155	0.000	m1
Doc9	0.000	0.046	0.155	0.222	m2
Doc10	0.000	0.023	0.000	0.222	m4
Q	0	0.523	0.552	0.537	

m1=(1, 1, 1, 0)

m2=(0, 1, 1, 1)

m3=(1, 0, 1, 0)

m4=(0, 1, 0, 1)

ขั้นตอนที่ 2

	Bird	Cat	Dog	Tiger	
Doc1	0.398	0.031	0.103	0.000	m1
Doc2	0.000	0.046	0.077	0.111	m2
Doc3	0.398	0.000	0.077	0.000	m3
Doc4	0.000	0.046	0.000	0.222	m4
Doc5	0.000	0.015	0.052	0.222	m2
Doc6	0.000	0.046	0.000	0.222	m4
Doc7	0.398	0.046	0.155	0.000	m1
Doc8	0.398	0.046	0.155	0.000	m1
Doc9	0.000	0.046	0.155	0.222	m2
Doc10	0.000	0.023	0.000	0.222	m4
Q	0	0.523	0.552	0.537	

m1=(1, 1, 1, 0)

m2=(0, 1, 1, 1)

m3=(1, 0, 1, 0)

m4=(0, 1, 0, 1)

$$k_2 = \frac{c_{2,1}m_1 + c_{2,2}m_2 + c_{2,4}m_4}{\sqrt{c_{2,1}^2 + c_{2,2}^2 + c_{2,4}^2}}$$

$$c_{2,1} = w_{2,1} + w_{2,7} + w_{2,8} = 0.031 + 0.046 + 0.046 = 0.122$$

$$c_{2,2} = w_{2,2} + w_{2,5} + w_{2,9} = 0.046 + 0.015 + 0.046 = 0.107$$

$$c_{2,4} = w_{2,4} + w_{2,6} + w_{2,10} = 0.046 + 0.046 + 0.023 = 0.114$$

$$k_2 = \frac{0.122m_1 + 0.107m_2 + 0.114m_4}{\sqrt{0.122^2 + 0.107^2 + 0.114^2}}$$

$$k_2 = 0.615m_1 + 0.538m_2 + 0.576m_4$$

ขั้นตอนที่ 2

$$k_3 = \frac{c_{3,1}m_1 + c_{3,2}m_2 + c_{3,3}m_3}{\sqrt{c_{3,1}^2 + c_{3,2}^2 + c_{3,3}^2}}$$

$$c_{3,1} = w_{3,1} + w_{3,7} + w_{3,8} = 0.103 + 0.155 + 0.155 = 0.413$$

$$c_{3,2} = w_{3,2} + w_{3,5} + w_{3,9} = 0.077 + 0.052 + 0.155 = 0.284$$

$$c_{3,3} = w_{3,3} = 0.077$$

$$k_3 = \frac{0.413m_1 + 0.284m_2 + 0.077m_3}{\sqrt{0.413^2 + 0.284^2 + 0.077^2}}$$

$$k_3 = 0.814m_1 + 0.560m_2 + 0.153m_3$$

	Bird	Cat	Dog	Tiger	
Doc1	0.398	0.031	0.103	0.000	m1
Doc2	0.000	0.046	0.077	0.111	m2
Doc3	0.398	0.000	0.077	0.000	m3
Doc4	0.000	0.046	0.000	0.222	m4
Doc5	0.000	0.015	0.052	0.222	m2
Doc6	0.000	0.046	0.000	0.222	m4
Doc7	0.398	0.046	0.155	0.000	m1
Doc8	0.398	0.046	0.155	0.000	m1
Doc9	0.000	0.046	0.155	0.222	m2
Doc10	0.000	0.023	0.000	0.222	m4
Q	0	0.523	0.552	0.537	

m1=(1, 1, 1, 0)

m2=(0, 1, 1, 1)

m3=(1, 0, 1, 0)

m4=(0, 1, 0, 1)

ขั้นตอนที่ 2

$$k_4 = \frac{c_{4,2}m_2 + c_{4,4}m_4}{\sqrt{c_{4,2}^2 + c_{4,4}^2}}$$

$$\begin{aligned} c_{4,2} &= w_{4,2} + w_{4,5} + w_{4,9} = 0.111 + 0.222 + 0.222 \\ &= 0.555 \\ c_{4,4} &= w_{4,4} + w_{4,6} + w_{4,10} = 0.222 + 0.222 + 0.222 \\ &= 0.666 \end{aligned}$$

$$k_4 = \frac{0.555m_2 + 0.666m_4}{\sqrt{0.555^2 + 0.666^2}}$$

$$k_4 = 0.640m_2 + 0.768m_4$$

	Bird	Cat	Dog	Tiger	
Doc1	0.398	0.031	0.103	0.000	m1
Doc2	0.000	0.046	0.077	0.111	m2
Doc3	0.398	0.000	0.077	0.000	m3
Doc4	0.000	0.046	0.000	0.222	m4
Doc5	0.000	0.015	0.052	0.222	m2
Doc6	0.000	0.046	0.000	0.222	m4
Doc7	0.398	0.046	0.155	0.000	m1
Doc8	0.398	0.046	0.155	0.000	m1
Doc9	0.000	0.046	0.155	0.222	m2
Doc10	0.000	0.023	0.000	0.222	m4
Q	0	0.523	0.552	0.537	

m1=(1, 1, 1, 0)

m2=(0, 1, 1, 1)

m3=(1, 0, 1, 0)

m4=(0, 1, 0, 1)

ขั้นตอนที่ 2

	Bird	Cat	Dog	Tiger	
Doc1	0.398	0.031	0.103	0.000	m1
Doc2	0.000	0.046	0.077	0.111	m2
Doc3	0.398	0.000	0.077	0.000	m3
Doc4	0.000	0.046	0.000	0.222	m4
Doc5	0.000	0.015	0.052	0.222	m2
Doc6	0.000	0.046	0.000	0.222	m4
Doc7	0.398	0.046	0.155	0.000	m1
Doc8	0.398	0.046	0.155	0.000	m1
Doc9	0.000	0.046	0.155	0.222	m2
Doc10	0.000	0.023	0.000	0.222	m4
Q	0	0.523	0.552	0.537	

$$m1=(1, 1, 1, 0)$$

$$m2=(0, 1, 1, 1)$$

$$m3=(1, 0, 1, 0)$$

$$m4=(0, 1, 0, 1)$$

$$q = 0.523k_2 + 0.552k_3 + 0.537k_4$$

$$q = 0.523 * (0.615m_1 + 0.538m_2 + 0.576m_4) + 0.552 * (0.814m_1 + 0.560m_2 + 0.153m_3) + 0.537 * (0.640m_2 + 0.768m_4)$$

$$q = 0.771m_1 + 0.934m_2 + 0.084m_3 + 0.714m_4$$

ขั้นตอนที่ 3

$$q = 0.771m_1 + 0.934m_2 + 0.084m_3 + 0.714m_4$$

$$d_1 = 0.398k_1 + 0.031k_2 + 0.103k_3$$

$$d_1 = 0.398 * (0.949m_1 + 0.316m_3) + 0.031 * (0.615m_1 + 0.538m_2 + 0.576m_4) \\ + 0.103 * (0.814m_1 + 0.560m_2 + 0.153m_3)$$

$$d_1 = 0.480m_1 + 0.074m_2 + 0.142m_3$$

	Bird	Cat	Dog	Tiger
Doc1	0.398	0.031	0.103	0.000
Doc2	0.000	0.046	0.077	0.111
Doc3	0.398	0.000	0.077	0.000
Doc4	0.000	0.046	0.000	0.222
Doc5	0.000	0.015	0.052	0.222
Doc6	0.000	0.046	0.000	0.222
Doc7	0.398	0.046	0.155	0.000
Doc8	0.398	0.046	0.155	0.000
Doc9	0.000	0.046	0.155	0.222
Doc10	0.000	0.023	0.000	0.222
Q	0	0.523	0.552	0.537

ขั้นตอนที่ 3

$$q = 0.771m_1 + 0.934m_2 + 0.084m_3 + 0.714m_4$$

$$\text{sim}(q, d_j) = \frac{\sum_r S_{d,r} \cdot S_{q,r}}{\sqrt{\sum_r S_{d,r}^2 \cdot \sum_r S_{q,r}^2}}$$

	m1	m2	m3	m4
Doc1	0.480	0.074	0.142	0.018
Doc2	0.091	0.139	0.012	0.112
Doc3	0.441	0.043	0.138	0.000
Doc4	0.028	0.167	0.000	0.197
Doc5	0.051	0.179	0.008	0.179
Doc6	0.028	0.167	0.000	0.197
Doc7	0.532	0.111	0.149	0.026
Doc8	0.532	0.111	0.149	0.026
Doc9	0.154	0.253	0.024	0.197
Doc10	0.014	0.154	0.000	0.184
Q	0.771	0.934	0.084	0.714

$$\begin{aligned}\text{sim}(q, d_1) &= \frac{0.480 * 0.771 + 0.074 * 0.934 + 0.142 * 0.084 + 0.018 * 0.714}{\sqrt{(0.480^2 + 0.074^2 + 0.142^2 + 0.018^2) * (0.771^2 + 0.934^2 + 0.084^2 + 0.714^2)}} \\ &= 0.974\end{aligned}$$

ขั้นตอนที่ 3

Doc	Sim	
Doc1	0.651	m1
Doc2	0.994	m2
Doc3	0.600	m3
Doc4	0.870	m4
Doc5	0.921	m2
Doc6	0.870	m4
Doc7	0.686	m1
Doc8	0.686	m1
Doc9	0.991	m2
Doc10	0.845	m4

Ranking	Sim	
Doc2	0.994	m2
Doc9	0.991	m2
Doc5	0.921	m2
Doc4	0.870	m4
Doc6	0.870	m4
Doc10	0.845	m4
Doc7	0.686	m1
Doc8	0.686	m1
Doc1	0.651	m1
Doc3	0.600	m3

	Bird	Cat	Dog	Tiger
Doc1	3	2	2	0
Doc2	0	2	1	1
Doc3	2	0	1	0
Doc4	0	1	0	1
Doc5	0	1	1	3
Doc6	0	2	0	2
Doc7	1	1	1	0
Doc8	1	1	1	0
Doc9	0	1	1	1
Doc10	0	1	0	2

Query = "cat cat cat dog dog tiger"

Rank → Doc2,Doc9,Doc5,Doc4,Doc6,Doc10,Doc7,Doc8,Doc1,Doc3

ขั้นตอนที่ 3

Ranking	Sim
Doc2	0.994 m2
Doc9	0.991 m2
Doc5	0.921 m2
Doc4	0.870 m4
Doc6	0.870 m4
Doc10	0.845 m4
Doc7	0.686 m1
Doc8	0.686 m1
Doc1	0.651 m1
Doc3	0.600 m3

	Bird	Cat	Dog	Tiger
Doc1	3	2	2	0
Doc2	0	2	1	1
Doc3	2	0	1	0
Doc4	0	1	0	1
Doc5	0	1	1	3
Doc6	0	2	0	2
Doc7	1	1	1	0
Doc8	1	1	1	0
Doc9	0	1	1	1
Doc10	0	1	0	2

Query = "cat cat cat dog dog tiger"

2.3 หากระบบกำหนดให้เอกสารที่ 1 ตรงประเด็นมากกว่าเอกสารที่ 9 โมเดลที่เลือกมาให้คำตอบถูกต้องหรือไม่ ถ้าผิดต้องแก้ไขอย่างไรจงอธิบาย

จากผลลัพธ์ที่ได้เอกสาร 9 ตรงประเด็นมากกว่าเอกสารที่ 1 จึงไม่เป็นไปตามที่โจทย์กำหนด เนื่องจากโจทย์ต้องการ Cat Dog Tiger แต่ไม่ต้องการ Bird ในเอกสารที่ 1 มี Bird ที่โจทย์ไม่ต้องการ และมี Cat และ Dog แต่ไม่มี Tiger จึงมีความตรงประเด็นที่น้อยกว่าเอกสารที่ 9 ที่มีทุก Keyword ที่สอดคล้องกับคำเรียกค้น หากต้องการให้ผลลัพธ์เป็นไปตามโจทย์ต้องการให้เพิ่มคีย์เวิร์ด Bird เข้าไปในคำเรียกค้น

ข้อ 3. เมื่อใช้งานคำเรียกค้น $Q = 7\text{Dog}-3\text{Cat}+\text{Bird}$ 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: "My dog wife is being stomach, A cat tell her boy
friend because its want to have son child cat"

โดย **Dog, Cat, Tiger, Bird** คือ Keyword ของระบบ

หากต้องการปรับคำเรียกค้น (Query) ให้มีผลลัพธ์มีความตรงประเด็นมากยิ่งขึ้น โดยโจทย์กำหนดว่าห้ามใช้ Associate Cluster และ Scalar Cluster นักศึกษาควรใช้โมเดลใดเพราะอะไร จงแสดงวิธีคำนวณในโมเดลที่เลือกใช้ (23 คะแนน)

Answer

เลือกใช้ Metric Cluster เนื่องจากต้องการปรับเปลี่ยน Query ดั้งเดิม โดยใช้เอกสารที่ถูกส่งออกมาและพิจารณาเฉพาะระยะห่างของ Keyword และโจทย์ไม่ต้องการให้ใช้วิธีอื่น

4.2

หาความสัมพันธ์ระหว่าง Keyword โดยพิจารณาระยะห่าง

Doc1

Keyword \rightarrow {Cat, Dog, Tiger}

หาระยะห่าง

Cat, Dog = 3 Dog, Tiger = 4

Cat, Tiger = 1

Doc2

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

หาระยะห่าง

Bird, Cat = 2 Cat, Dog = 5 Dog, Tiger = 14

Bird, Dog = 8 Cat, Tiger = 8

Bird, Tiger = 6

4.2

หาความสัมพันธ์ระหว่าง Keyword โดยพิจารณาระยะห่าง

Doc4

Keyword \rightarrow {Bird, Cat, Dog}

หาระยะห่าง

Bird, Cat = 9 Cat, Dog = 4

Bird, Dog = 5

Doc5

Keyword \rightarrow {Cat, Dog}

หาระยะห่าง

Cat, Dog = 6

เลือกกระยะห่างที่สั้นที่สุด 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

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

C	Bird	Cat	Dog	Tiger
Bird	-	0.500	0.200	0.333
Cat	0.500	-	0.333	1.000
Dog	0.200	0.333	-	0.500
Tiger	0.333	1.000	0.500	-

ความสัมพันธ์ระหว่างรากศัพท์ Stem

V1 = {Bird} Size = 1

V2 = {Cat} Size = 1

V3 = {Dog } Size = 1

V4 = {Tiger} Size = 1

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

เนื่องจากขนาดของรากศัพท์ทุกรากศัพท์มีขนาดเท่ากับ 1 ดังนั้นเมื่อคำนวณหาผลรวมของระยะห่างระหว่างรากศัพท์จึงมีค่าเท่ากับตารางเดิม

C	Bird	Cat	Dog	Tiger
Bird	-	0.500	0.200	0.333
Cat	0.500	-	0.333	1.000
Dog	0.200	0.333	-	0.500
Tiger	0.333	1.000	0.500	-

ความสัมพันธ์ระหว่าง รากศัพท์ Stem

C	Bird	Cat	Dog	Tiger
Bird	–	0.500	0.200	0.333
Cat	0.500	–	0.333	1.000
Dog	0.200	0.333	–	0.500
Tiger	0.333	1.000	0.500	–

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} Size = 1

V3 = {Dog} Size = 1

V4 = {Tiger} Size = 1

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

V1 = {Bird} Size = 1
 V2 = {Cat} Size = 1
 V3 = {Dog} Size = 1
 V4 = {Tiger} Size = 1

เนื่องจากขนาดของรากศัพท์ทุกรากศัพท์มีขนาดเท่ากับ 1 ดังนั้นเมื่อ Normalize แล้วตาราง S จึงมีค่าเท่ากับตาราง C

N	Bird	Cat	Dog	Tiger
Bird	–	0.500	0.200	0.333
Cat	0.500	–	0.333	1.000
Dog	0.200	0.333	–	0.500
Tiger	0.333	1.000	0.500	–

ตาราง Normalize ความสัมพันธ์ระหว่างรากศัพท์

N	Bird	Cat	Dog	Tiger
Bird	-	0.500	0.200	0.333
Cat	0.500	-	0.333	1.000
Dog	0.200	0.333	-	0.500
Tiger	0.333	1.000	0.500	-

ความสัมพันธ์ระหว่างรากศัพท์

Bird→Cat

Cat→Tiger

Dog→Tiger

Tiger→Cat

Original Query: **7Dog-3Cat+Bird**

New Query:

$$\begin{aligned}7\text{Dog}-3\text{Cat}+\text{Bird} &= 7(\text{Dog}+0.5\text{Tiger})-3(\text{Cat}+\text{Tiger})+(\text{Bird}+0.5\text{Cat}) \\ &= 7\text{Dog} + 3.5\text{Tiger} - 3\text{Cat} - 3\text{Tiger} + \text{Bird} + 0.5\text{Cat} \\ &= \text{Bird} - 2.5\text{Cat} + 7\text{Dog} + 0.5\text{Tiger}\end{aligned}$$

ข้อ 4. เอกสารหนึ่งมีข้อความดังนี้

เสือดำ เป็นชื่อสามัญเรียกโดยรวมของสัตว์กินเนื้อประเภทเสือและแมว (Felidae) ที่มีลักษณะลำตัวรวมถึงลวดลายเป็นสีดำนวลตลอดทั้งลำตัว ซึ่งเกิดขึ้นได้ในเสือหลายชนิด

จากการเป็นเสือดำ เกิดจากความผิดปกติในเม็ดสีที่เรียกว่าเมลานินซึม ส่งผลให้เสือที่เกิดมานั้นเป็นสีดำนวลตลอดทั้งลำตัว โดยที่ยังมีลายหรือลายจุดคงอยู่ แต่จะสังเกตเห็นได้ยาก จะเห็นได้ชัดเจนเมื่ออยู่ในแสงแดด

เสือดำในเสือดาว มักพบได้มากในป่าดิบชื้นในทวีปเอเชีย เช่น อินเดีย, เนปาล, เอเชียตะวันออกเฉียงใต้ โดยเฉพาะในเบงกอลหรือชวา

4.1 เพื่อให้ได้คำตอบในข้อ 4.2 นักศึกษาควรใช้ Model ใดเพราะอะไร (เลือกได้เฉพาะตัวเลือกที่ให้มา)

A) Knuth-Morris-Pratt (KMP) B) Breadth-first search (BFS) C) Depth-first search (DFS) D) Boyer Moore

4.2 หากต้องการค้นคำว่า "เสือ"

จงแสดงการค้นหาคำว่า "เสือ" ในเอกสารด้านบนอย่างละเอียด โดยแสดงผลตำแหน่งที่ปรากฏคำที่เรียกค้นนี้ (17 คะแนน)

Answer

4.1 เลือกใช้ Boyer Moor หรือ Knuth-Morris-Pratt (KMP) ได้ทั้งสองวิธี ซึ่งเป็นกระบวนการค้นหาข้อมูลตามลำดับ (Sequential Search)

4.2

เสือดำ เป็นชื่อสามัญเรียกโดยรวมของสัตว์กินเนื้อประเภทเสือและแมว (Felidae) ที่มีลักษณะลำตัวรวมถึงลวดลายเป็นสี
ดำตลอดทั้งลำตัว ซึ่งเกิดขึ้นได้ในเสือหลายชนิด จากการเป็นเสือดำ เกิดจากความผิดปกติในเม็ดสีที่เรียกว่าเมลานินซึม ส่งผล
ให้เสือที่เกิดมานั้นเป็นสีดำตลอดทั้งลำตัว โดยที่ยังมีลายหรือลายจุดคงอยู่ แต่จะสังเกตเห็นได้ยาก จะเห็นได้ชัดเจนเมื่ออยู่ใน
แสงแดด เสือดำนี้นับว่าหายากมากพบได้มากในป่าดิบชื้นในทวีปเอเชีย เช่น อินเดีย, เนปาล, เอเชียตะวันออกเฉียงใต้
โดยเฉพาะในเบงกอลหรือชวา

$$\text{Value} = \text{length} - \text{index} - 1$$

$$\text{เ} = 3 - 0 - 1 = 2$$

$$\text{สี} = 3 - 1 - 1 = 1$$

$$\text{อ} = 3$$

Boyer Moore

Letter	เ	สี	อ	*
Value	2	1	3	3

4.2

1

2

3

4

6

7

5

9

10

11

เสื้อ คือพบคำเรียกค้นในเอกสาร

4.2

Item	Current Position	Match	Next Move Position	Move By	Found At Position
0	1	Y	4		1
1	25		28	๑	
2	36		39	๑	
3	43	Y	46	๑	43
4	112		113	๓๕	
	113	Y	116	๑	113
5	133	Y	136	๑	133
6	180	Y	183	๑	180
7	273		275	๒	
	275	Y	278	๑	275
8	281		282	๓๕	
	282	Y	285	๑	282
9	324		327	๑	
10	332		335	๑	
11	356		359	๑	

Letter	๒	๓	๑	*
Value	2	1	3	3