PreExam

ส่งคำตอบที่ : IR.CE.KMITL@Gmail.com

ชื่อเมล์ : PreExam_รหัสนักศึกษา เช่น PreExam_62010109

เขียนคำตอบด้วยลายมือ ไฟล์แนบชื่อใดก็ได้

ส่งภายใน 4/10/2566 ไม่เกินเที่ยงวัน (12.00น.)

(การคัดลอกถือว่าผิดจริยธรรม)

สมมติในระบบมีเอกสาร 10 เอกสารดังนี้ (bird, cat, dog, tiger คือ **Keyword ซึ่งไม่มีความสัมพันธ์กัน**)

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: {bird,cat,bird,cat,tiger,tiger,bird}
D7: {bird,tiger,cat,dog}
D8: {dog,cat,bird}
D9: {**???,???,???**}
D10: {**???,???,???**}

*** เอกสารหมายเลข 9 และ 10 ให้น.ศ.ป้อน Keyword (bird cat dog tiger) เองตามใจชอบ โดยเอกสารทั้งสองต้องไม่เท่ากัน

ผู้ใช้ส่งคำเรียกค้น **"cat dog tiger cat"** เข้าไปในระบบ จงตอบคำถาม

1) เพื่อคำนวณหา Ranking ของเอกสารทุกเอกสารในระบบ ผู้ใช้**สามารถเลือกใช้โมเดลใดได้บ้างอย่างน้อย 2 โมเดล เพราะอะไร (เลือกเฉพาะโมเดลที่ให้มาเท่านั้น)**

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

2) จากข้อ 1 ให้นักศึกษาแสดงวิธีคำนวณหา Ranking ของเอกสารทุกเอกสารในระบบจำนวน 2 โมเดล และนำผลที่ได้จาก 2 โมเดลที่เลือกมาเปรียบเทียบกัน

(100 คะแนน)

สมมติในระบบมีเอกสาร 10 เอกสารดังนี้ (bird, cat, dog, tiger คือ **Keyword ซึ่งไม่มีความสัมพันธ์กัน**)

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: {bird,cat,bird,cat,tiger,tiger,bird}
D7: {bird,tiger,cat,dog}
D8: {dog,cat,bird}
D9: {cat,dog,tiger}
D10: {tiger,tiger,tiger}

ผู้ใช้ส่งคำเรียกค้น **"cat dog tiger cat"** เข้าไปในระบบ จงตอบคำถาม

1) เพื่อคำนวณหา Ranking ของเอกสารทุกเอกสารในระบบ ผู้ใช้**สามารถเลือกใช้โมเดลใดได้บ้างอย่างน้อย 2 โมเดล เพราะอะไร (เลือกเฉพาะโมเดลที่ให้มาเท่านั้น)**

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

2) จากข้อ 1 ให้นักศึกษาแสดงวิธีคำนวณหา Ranking ของเอกสารทุกเอกสารในระบบจำนวน 2 โมเดล และนำผลที่ได้จาก 2 โมเดลที่เลือกมาเปรียบเทียบกัน

Answer

2.1 เลือกใช้ BM25 Model และ Vector Model เนื่องจากลักษณะของ Query เป็น keyword แยกกัน ไม่มี Expression และโจทย์กำหนดให้ Keyword ไม่สัมพันธ์กัน

Vector 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: {bird,cat,bird,cat,tiger,tiger,bird}
D7: {bird,tiger,cat,dog}
D8: {dog,cat,bird}
D9: {cat,dog,tiger}
D10: {tiger,tiger,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	3	2	0	2	3
Doc7	1	1	1	1	1
Doc8	1	1	1	0	1
Doc9	0	1	1	1	1
Doc10	0	0	0	3	3
n	5	8	7	7	

Vector 1

Only Doc1

	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	3	2	0	2	3
Doc7	1	1	1	1	1
Doc8	1	1	1	0	1
Doc9	0	1	1	1	1
Doc10	0	0	0	3	3
n	5	8	7	7	

$$tf_{bird} = \frac{3}{3} = 1.000$$
$$tf_{cat} = \frac{2}{3} = 0.667$$
$$tf_{dog} = \frac{2}{3} = 0.667$$
$$tf_{tiger} = \frac{0}{3} = 0.000$$

$$idf_{bird} = \log(\frac{10}{5}) = 0.301$$
$$idf_{cat} = \log(\frac{10}{8}) = 0.097$$
$$idf_{dog} = \log(\frac{10}{7}) = 0.155$$
$$idf_{tiger} = \log(\frac{10}{7}) = 0.155$$

$$\boldsymbol{w_{bird} = 1.000 * 0.301 = 0.301}$$
$$\boldsymbol{w_{cat} = 0.667 * 0.097 = 0.065}$$
$$\boldsymbol{w_{dog} = 0.667 * 0.155 = 0.103}$$
$$\boldsymbol{w_{tiger} = 0.000 * 0.155 = 0.000}$$

Vector 1

น้ำหนักของแต่ละ **Keyword** ในแต่ละเอกสาร

	Bird	Cat	Dog	Tiger
Doc1	0.301	0.065	0.103	0.000
Doc2	0.000	0.097	0.077	0.077
Doc3	0.301	0.000	0.077	0.000
Doc4	0.000	0.097	0.000	0.155
Doc5	0.000	0.032	0.052	0.155
Doc6	0.301	0.065	0.000	0.103
Doc7	0.301	0.097	0.155	0.155
Doc8	0.301	0.097	0.155	0.000
Doc9	0.000	0.097	0.155	0.155
Doc10	0.000	0.000	0.000	0.155

Vector 2

กรณี คำนวณ bird ใน Query

	Bird	Cat	Dog	Tiger
Doc1	0.301	0.065	0.103	0.000
Doc2	0.000	0.097	0.077	0.077
Doc3	0.301	0.000	0.077	0.000
Doc4	0.000	0.097	0.000	0.155
Doc5	0.000	0.032	0.052	0.155
Doc6	0.301	0.065	0.000	0.103
Doc7	0.301	0.097	0.155	0.155
Doc8	0.301	0.097	0.155	0.000
Doc9	0.000	0.097	0.155	0.155
Doc10	0.000	0.000	0.000	0.155

Query = cat dog tiger cat

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

$$W_{bird,\,q} = \left(0.5 + \frac{0.5 * 0}{2}\right) * 0.301 = 0.151$$

$$W_{cat,\,q} = \left(0.5 + \frac{0.5 * 2}{2}\right) * 0.097 = 0.097$$

$$W_{dog,\,q} = \left(0.5 + \frac{0.5 * 1}{2}\right) * 0.155 = 0.117$$

$$W_{tiger,\,q} = \left(0.5 + \frac{0.5 * 1}{2}\right) * 0.155 = 0.117$$

$$idf_{bird} = \log(\frac{10}{5}) = 0.301$$

$$idf_{cat} = \log(\frac{10}{8}) = 0.097$$

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

$$idf_{tiger} = \log(\frac{10}{7}) = 0.155$$

Vector 3

กรณี คำนวณ bird ใน Query

Query = cat dog tiger cat

	Bird	Cat	Dog	Tiger
Doc1	0.301	0.065	0.103	0.000
Doc2	0.000	0.097	0.077	0.077
Doc3	0.301	0.000	0.077	0.000
Doc4	0.000	0.097	0.000	0.155
Doc5	0.000	0.032	0.052	0.155
Doc6	0.301	0.065	0.000	0.103
Doc7	0.301	0.097	0.155	0.155
Doc8	0.301	0.097	0.155	0.000
Doc9	0.000	0.097	0.155	0.155
Doc10	0.000	0.000	0.000	0.155
q	0.151	0.097	0.117	0.117

$$sim(d_j, q) = \frac{\sum_{j=1}^{t} w_{q_j} w_{d_{ij}}}{\sqrt{\sum_{j=1}^{t} (w_{q_j})^2 \sum_{j=1}^{t} (w_{d_{ij}})^2}}$$

$$sim(d_1, q) = \frac{0.301 * 0.151 + 0.065 * 0.097 + 0.103 * 0.117}{\sqrt{(0.1512 + 0.0972 + 0.1172 + 0.1172)(0.3012 + 0.0652 + 0.1032 + 02)}}$$

$$= 0.806$$

Vector 4

กรณี คำนวณ bird ใน Query

Query = cat dog tiger cat

	Sim
Doc1	0.806
Doc2	0.771
Doc3	0.719
Doc4	0.617
Doc5	0.671
Doc6	0.806
Doc7	0.970
Doc8	0.850
Doc9	0.780
Doc10	0.478

Rank →

	Sim
Doc7	0.970
Doc8	0.850
Doc1	0.806
Doc6	0.806
Doc9	0.780
Doc2	0.771
Doc3	0.719
Doc5	0.671
Doc4	0.617
Doc10	0.478

	Bird	Cat	Dog	Tiger
Doc1	0.301	0.065	0.103	0.000
Doc2	0.000	0.097	0.077	0.077
Doc3	0.301	0.000	0.077	0.000
Doc4	0.000	0.097	0.000	0.155
Doc5	0.000	0.032	0.052	0.155
Doc6	0.301	0.065	0.000	0.103
Doc7	0.301	0.097	0.155	0.155
Doc8	0.301	0.097	0.155	0.000
Doc9	0.000	0.097	0.155	0.155
Doc10	0.000	0.000	0.000	0.155
q	0.151	0.097	0.117	0.117

Rank → D7,D8,D6,D1,D9,D2,D3,D5,D4,D10

Vector 2

กรณี ไม่คำนวณ bird ใน Query

	Bird	Cat	Dog	Tiger
Doc1	0.301	0.065	0.103	0.000
Doc2	0.000	0.097	0.077	0.077
Doc3	0.301	0.000	0.077	0.000
Doc4	0.000	0.097	0.000	0.155
Doc5	0.000	0.032	0.052	0.155
Doc6	0.301	0.065	0.000	0.103
Doc7	0.301	0.097	0.155	0.155
Doc8	0.301	0.097	0.155	0.000
Doc9	0.000	0.097	0.155	0.155
Doc10	0.000	0.000	0.000	0.155

Query = cat dog tiger cat

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

$W_{bird,\,q} = 0$ bird เป็น 0 เพราะไม่มีในการเรียกค้น

$$W_{cat,\,q} = \left(0.5 + \frac{0.5 * 2}{2}\right) * 0.097 = 0.097$$

$$W_{dog,\,q} = \left(0.5 + \frac{0.5 * 1}{2}\right) * 0.155 = 0.117$$

$$W_{dog,\,q} = \left(0.5 + \frac{0.5 * 1}{2}\right) * 0.155 = 0.117$$

$$idf_{bird} = \log(\frac{10}{5}) = 0.301$$

$$idf_{cat} = \log(\frac{10}{8}) = 0.097$$

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

$$idf_{tiger} = \log(\frac{10}{7}) = 0.155$$

Vector 3

กรณี ไม่คำนวณ bird ใน Query

Query = cat dog tiger cat

	Bird	Cat	Dog	Tiger
Doc1	0.301	0.065	0.103	0.000
Doc2	0.000	0.097	0.077	0.077
Doc3	0.301	0.000	0.077	0.000
Doc4	0.000	0.097	0.000	0.155
Doc5	0.000	0.032	0.052	0.155
Doc6	0.301	0.065	0.000	0.103
Doc7	0.301	0.097	0.155	0.155
Doc8	0.301	0.097	0.155	0.000
Doc9	0.000	0.097	0.155	0.155
Doc10	0.000	0.000	0.000	0.155
q	0	0.097	0.117	0.117

$$sim(d_j, q) = \frac{\sum_{j=1}^{t} w_{q_j} w_{d_{ij}}}{\sqrt{\sum_{j=1}^{t} (w_{q_j})^2 \sum_{j=1}^{t} (w_{d_{ij}})^2}}$$

$$sim(d_1, q) = \frac{0.065 * 0.097 + 0.103 * 0.117}{\sqrt{(0.0972 + 0.1172 + 0.1172)(0.3012 + 0.0652 + 0.1032 + 02)}}$$

$$= 0.295$$

Vector 4

กรณี ไม่คำนวณ bird ใน Query

Query = cat dog tiger cat

	Sim
Doc1	0.295
Doc2	0.982
Doc3	0.152
Doc4	0.786
Doc5	0.854
Doc6	0.295
Doc7	0.618
Doc8	0.408
Doc9	0.993
Doc10	0.609

Rank →

	Sim
Doc9	0.993
Doc2	0.982
Doc5	0.854
Doc4	0.786
Doc7	0.618
Doc10	0.609
Doc8	0.408
Doc1	0.295
Doc6	0.295
Doc3	0.152

	Bird	Cat	Dog	Tiger
Doc1	0.301	0.065	0.103	0.000
Doc2	0.000	0.097	0.077	0.077
Doc3	0.301	0.000	0.077	0.000
Doc4	0.000	0.097	0.000	0.155
Doc5	0.000	0.032	0.052	0.155
Doc6	0.301	0.065	0.000	0.103
Doc7	0.301	0.097	0.155	0.155
Doc8	0.301	0.097	0.155	0.000
Doc9	0.000	0.097	0.155	0.155
Doc10	0.000	0.000	0.000	0.155
q	0	0.097	0.117	0.117

Rank → D9,D2,D5,D4,D7,D8,D10,D1,D6,D3

Vector 4

คำนวณ bird VS ไม่คำนวณ bird

Query = cat dog tiger cat

	Sim
Doc7	0.970
Doc8	0.850
Doc1	0.806
Doc6	0.806
Doc9	0.780
Doc2	0.771
Doc3	0.719
Doc5	0.671
Doc4	0.617
Doc10	0.478

คำนวณ bird

	Sim
Doc9	0.993
Doc2	0.982
Doc5	0.854
Doc4	0.786
Doc7	0.618
Doc10	0.609
Doc8	0.408
Doc1	0.295
Doc6	0.295
Doc3	0.152

ไม่คำนวณ bird

	Bird	Cat	Dog	Tiger
Doc1	0.301	0.065	0.103	0.000
Doc2	0.000	0.097	0.077	0.077
Doc3	0.301	0.000	0.077	0.000
Doc4	0.000	0.097	0.000	0.155
Doc5	0.000	0.032	0.052	0.155
Doc6	0.301	0.065	0.000	0.103
Doc7	0.301	0.097	0.155	0.155
Doc8	0.301	0.097	0.155	0.000
Doc9	0.000	0.097	0.155	0.155
Doc10	0.000	0.000	0.000	0.155

จะเห็นได้ว่าการไม่คำนวณ **bird** จะให้ผลลัพธ์ที่สอดคล้องต่อความต้องการของ **User** มากกว่า อาทิ เอกสาร **9** ตรงประเด็นมากกว่าเอกสาร **2** เพราะมีน้ำหนักของ **dog** และ **tiger** มากกว่า

BM25 1

Query = cat dog tiger cat

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: {bird,cat,bird,cat,tiger,tiger,bird}
D7: {bird,tiger,cat,dog}
D8: {dog,cat,bird}
D9: {cat,dog,tiger}
D10: {tiger,tiger,tiger}
```

$$sim\ (d_j, q) = \sum_{i \epsilon q} \log \frac{(r_i + 0.5)/(R - r_i + 0.5)}{(n_i - r_i + 0.5)/(N - n_i - R + r_i + 0.5)} \cdot \frac{(k_1 + 1) f_i}{k_1 \left((1 - b) + b \cdot \frac{dl}{avdl} \right) + f_i} \cdot \frac{(k_2 + 1) qf_i}{k_2 + qf_i}$$

d_j - เอกสารที่ j

R – จำนวนเอกสารที่ตรงประเด็น

N – จำนวนเอกสารทั้งหมด

r_i – จำนวนเอกสารที่ตรงประเด็นที่มี keyword i

n_i – จำนวนเอกสารทั้งหมดที่มี keyword i

f_i – ความถี่ของ keyword i ในเอกสาร j

dl – จำนวนคำของเอกสาร j

avdl – จำนวนคำเฉลี่ยของทุกเอกสาร

qf_i – ความถี่ของ keyword i ใน query

b – ค่าคงที่โดยตาม TREC จะใช้ค่า 0.75 ($0.5 < b < 0.8$)

k_1 – ค่าคงที่โดยตาม TREC จะใช้ค่า 1.25 ($1.2 < k_1 < 2$)

k_2 – ค่าคงที่โดยปกติจะอยู่ในช่วง 0 - 1000

BM25 1

Query = cat dog tiger cat

	Bird	Cat	Dog	Tiger	Length
Doc1	3	2	2	0	7
Doc2	0	2	1	1	4
Doc3	2	0	1	0	3
Doc4	0	1	0	1	2
Doc5	0	1	1	3	5
Doc6	3	2	0	2	7
Doc7	1	1	1	1	4
Doc8	1	1	1	0	3
Doc9	0	1	1	1	3
Doc10	0	0	0	3	3

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: {bird,cat,bird,cat,tiger,tiger,bird}
D7: {bird,tiger,cat,dog}
D8: {dog,cat,bird}
D9: {cat,dog,tiger}
D10: {tiger,tiger,tiger}

$$Avdl = \frac{41}{10} = 4.1$$

$N = 10$
$n_{Bird} = 5$
$n_{Cat} = 8$
$n_{Dog} = 7$
$n_{Tiger} = 7$

$R = 0$
$r_{Bird} = 0$
$r_{Cat} = 0$
$r_{Dog} = 0$
$r_{Tiger} = 0$

เนื่องจากไม่มีการกำหนดให้เอกสารใดตรงประเด็น

BM25 2

Query = cat dog tiger cat

$$idf_i = log \frac{(r_i + 0.5)/(R - r_i + 0.5)}{(n_i - r_i + 0.5)/(N - n_i - R + r_i + 0.5)}$$

$$idf_i = log \frac{N - n_i + 0.5}{(n_i + 0.5)}$$

$$idf_{bird} = log \frac{10 - 5 + 0.5}{(5 + 0.5)} = 0.0$$

$$idf_{cat} = log \frac{10 - 8 + 0.5}{(8 + 0.5)} = -0.531$$

$$idf_{dog} = log \frac{10 - 7 + 0.5}{(7 + 0.5)} = -0.331$$

$$idf_{tiger} = log \frac{10 - 7 + 0.5}{(7 + 0.5)} = -0.331$$

	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	3	2	0	2
Doc7	1	1	1	1
Doc8	1	1	1	0
Doc9	0	1	1	1
Doc10	0	0	0	3

$N = 10$

$n_{Bird} = 5$

$n_{Cat} = 8$

$n_{Dog} = 7$

$n_{Tiger} = 7$

$R = 0$

$r_{Bird} = 0$

$r_{Cat} = 0$

$r_{Dog} = 0$

$r_{Tiger} = 0$

$\boldsymbol{Avdl} = 4.1$

BM25 2

Query = cat dog tiger cat

d_j - เอกสารที่ j
R – จำนวนเอกสารที่ตรงประเด็น
N – จำนวนเอกสารทั้งหมด
r_i – จำนวนเอกสารที่ตรงประเด็นที่มี keyword i
n_i – จำนวนเอกสารทั้งหมดที่มี keyword i
f_i – ความถี่ของ keyword i ในเอกสาร j
dl – จำนวนคำของเอกสาร j
avdl – จำนวนคำเฉลี่ยของทุกเอกสาร
qf_i – ความถี่ของ keyword i ใน query
b – ค่าคงที่โดยตาม TREC จะใช้ค่า 0.75 ($0.5 < b < 0.8$)
k_1 – ค่าคงที่โดยตาม TREC จะใช้ค่า 1.25 ($1.2 < k_1 < 2$)
k_2 – ค่าคงที่โดยปกติจะอยู่ในช่วง 0 - 1000

	idf
Bird	0.000
Cat	-0.531
Dog	-0.331
Tiger	-0.331

	Bird	Cat	Dog	Tiger	Length
Doc1	3	2	2	0	7
Doc2	0	2	1	1	4
Doc3	2	0	1	0	3
Doc4	0	1	0	1	2
Doc5	0	1	1	3	5
Doc6	3	2	0	2	7
Doc7	1	1	1	1	4
Doc8	1	1	1	0	3
Doc9	0	1	1	1	3
Doc10	0	0	0	3	3

$$sim\ (d_j, q) = \sum_{i \epsilon q} \log \frac{(r_i + 0.5)/(R - r_i + 0.5)}{(n_i - r_i + 0.5)/(N - n_i - R + r_i + 0.5)} \cdot \frac{(k_1 + 1) f_i}{k_1 \left((1 - b) + b \cdot \frac{dl}{avdl}\right) + f_i} \cdot \frac{(k_2 + 1) qf_i}{k_2 + qf_i}$$

$$sim\ (d_1, q) = 0.0 * \frac{(2.25)3}{1.25\left((1 - 0.75) + 0.75 * \frac{7}{4.1}\right) + 3} * \frac{201 * 0}{200 + 0} + (-0.531) * \frac{(2.25)2}{1.25\left((1 - 0.75) + 0.75 * \frac{7}{4.1}\right) + 2} * \frac{201 * 2}{200 + 2}$$

$$+(-0.331) * \frac{(2.25)2}{1.25\left((1 - 0.75) + 0.75 * \frac{7}{4.1}\right) + 2} * \frac{201 * 1}{200 + 1} + (-0.331) * \frac{(2.25)0}{1.25\left((1 - 0.75) + 0.75 * \frac{7}{4.1}\right) + 0} * \frac{201 * 1}{200 + 1}$$

$$= -1.597$$

BM25 2

Query = cat dog tiger cat

	Sim
Doc1	-1.597
Doc2	-1.821
Doc3	-0.373
Doc4	-1.765
Doc5	-1.774
Doc6	-1.597
Doc7	-1.737
Doc8	-1.073
Doc9	-1.936
Doc10	-0.559

Rank →

BM25	Sim
Doc3	-0.373
Doc10	-0.559
Doc8	-1.073
Doc1	-1.597
Doc6	-1.597
Doc7	-1.737
Doc4	-1.765
Doc5	-1.774
Doc2	-1.821
Doc9	-1.936

	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	3	2	0	2
Doc7	1	1	1	1
Doc8	1	1	1	0
Doc9	0	1	1	1
Doc10	0	0	0	3
q	*0*	*2*	*1*	*1*

สรุป

Query = cat dog tiger cat

Vector	Sim
Doc9	0.993
Doc2	0.982
Doc5	0.854
Doc4	0.786
Doc7	0.618
Doc10	0.609
Doc8	0.408
Doc1	0.295
Doc6	0.295
Doc3	0.152

BM25	Sim
Doc3	-0.373
Doc10	-0.559
Doc8	-1.073
Doc1	-1.597
Doc6	-1.597
Doc7	-1.737
Doc4	-1.765
Doc5	-1.774
Doc2	-1.821
Doc9	-1.936

	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	3	2	0	2
Doc7	1	1	1	1
Doc8	1	1	1	0
Doc9	0	1	1	1
Doc10	0	0	0	3
q	*0*	*2*	*1*	*1*

สรุป **Vector model** มีความตรงประเด็นที่ใกล้เคียงกว่า **BM25 Model**
เนื่องจาก **BM25** มีการกำหนดเอกสารตัวอย่างที่น้อยเกินไป และไม่มีการกำหนดว่าเอกสารใดบ้างที่ตรงประเด็น