

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

B) Fuzzy Model

C) Extend Boolean Model

D) Boolean Model

E) Probabilistic Model

F) Generalized Vector Model

2.2 ให้นักศึกษาแสดงวิธีคำนวณหา Ranking ของเอกสารทุกเอกสารในระบบ ตามที่ผู้ใช้ต้องการ โดยให้ความสำคัญกับลักษณะการปรากฏของคีย์เวิร์ดในแต่ละเอกสาร กล่าวคือเอกสารที่มีลักษณะการปรากฏของคีย์เวิร์ดที่เหมือนกันจะต้องมีความตรงประเด็นที่ใกล้เคียงกัน

2.3 หากระบบกำหนดให้เอกสารที่ 1 ตรงประเด็นมากกว่าเอกสารที่ 9 โมเดลที่เลือกมาให้คำตอบถูกต้องหรือไม่ ถ้าผิดต้องแก้ไขอย่างไรจึงอธิบาย

2.1 > เพื่อให้ได้คำตอบในคำถาม 2.2 ผู้ใช้สามารถเลือกใช้โมเดล Generalized Vector Model เพราะโมเดลนี้ใช้พิจารณา ำคัญ กับลักษณะการปรากฏของ Keyword ในแต่ละเอกสาร ซึ่งลักษณะเอกสารในหนึ่งลักษณะการปรากฏของ keyword ที่เหมือนกัน ก็จะตรงประเด็นที่ใกล้เคียงกัน หรือเอกสารที่มี Match กับ keyword ของผู้ใช้ก็จะใกล้เคียงกัน เนื่องจากลักษณะการปรากฏของ keyword ใกล้เคียงกัน

## 2.2

	bird	cat	dog	tiger	Max
Doc 1	3	2	2	0	3
Doc 2	0	2	1	1	2
Doc 3	2	0	1	0	2
Doc 4	0	1	0	1	1
Doc 5	0	1	1	3	3
Doc 6	0	2	0	2	2
Doc 7	1	1	1	0	1
Doc 8	1	1	1	0	1
Doc 9	0	1	1	1	1
Doc 10	0	1	0	2	2
n	4	9	7	6	

$$\text{idf}_{\text{bird}} = \log\left(\frac{10}{3}\right) = 0.5179$$

$$\text{idf}_{\text{cat}} = \log\left(\frac{10}{2}\right) = 0.6909$$

$$\text{idf}_{\text{dog}} = \log\left(\frac{10}{7}\right) = 0.4519$$

$$\text{idf}_{\text{tiger}} = \log\left(\frac{10}{6}\right) = 0.5108$$

Doc 1

$$\text{tf}_{\text{bird}} = \frac{3}{3} = 1.000$$

$$\text{tf}_{\text{cat}} = \frac{2}{3} = 0.6667$$

$$\text{tf}_{\text{dog}} = \frac{2}{3} = 0.6667$$

$$\text{tf}_{\text{tiger}} = \frac{0}{3} = 0.000$$

Doc 2

$$\text{tf}_{\text{bird}} = \frac{0}{2} = 0.000$$

$$\text{tf}_{\text{cat}} = \frac{2}{2} = 1.000$$

$$\text{tf}_{\text{dog}} = \frac{1}{2} = 0.5000$$

$$\text{tf}_{\text{tiger}} = \frac{1}{2} = 0.5000$$

Doc 3

$$\text{tf}_{\text{bird}} = \frac{2}{2} = 1.0000$$

$$\text{tf}_{\text{cat}} = \frac{0}{2} = 0.0000$$

$$\text{tf}_{\text{dog}} = \frac{1}{2} = 0.5000$$

$$\text{tf}_{\text{tiger}} = \frac{0}{2} = 0.0000$$

Doc 4

$$\text{tf}_{\text{bird}} = \frac{0}{1} = 0.0000$$

$$\text{tf}_{\text{cat}} = \frac{1}{1} = 1.0000$$

$$\text{tf}_{\text{dog}} = \frac{0}{1} = 0.0000$$

$$\text{tf}_{\text{tiger}} = \frac{1}{1} = 1.0000$$

Doc 5

$$\text{tf}_{\text{bird}} = \frac{0}{3} = 0.0000$$

$$\text{tf}_{\text{cat}} = \frac{1}{3} = 0.3333$$

$$\text{tf}_{\text{dog}} = \frac{1}{3} = 0.3333$$

$$\text{tf}_{\text{tiger}} = \frac{3}{3} = 1.0000$$

Doc 6

$$\text{tf}_{\text{bird}} = \frac{0}{2} = 0.0000$$

$$\text{tf}_{\text{cat}} = \frac{2}{2} = 1.0000$$

$$\text{tf}_{\text{dog}} = \frac{0}{2} = 0.0000$$

$$\text{tf}_{\text{tiger}} = \frac{2}{2} = 1.0000$$

Doc 7

$$\text{tf}_{\text{bird}} = \frac{1}{1} = 1.0000$$

$$\text{tf}_{\text{cat}} = \frac{1}{1} = 1.0000$$

$$\text{tf}_{\text{dog}} = \frac{1}{1} = 1.0000$$

$$\text{tf}_{\text{tiger}} = \frac{0}{1} = 0.0000$$

Doc 8

$$\text{tf}_{\text{bird}} = \frac{1}{1} = 1.0000$$

$$\text{tf}_{\text{cat}} = \frac{1}{1} = 1.0000$$

$$\text{tf}_{\text{dog}} = \frac{1}{1} = 1.0000$$

$$\text{tf}_{\text{tiger}} = \frac{0}{1} = 0.0000$$

Doc 9

$$\text{tf}_{\text{bird}} = \frac{0}{1} = 0.0000$$

$$\text{tf}_{\text{cat}} = \frac{1}{1} = 1.0000$$

$$\text{tf}_{\text{dog}} = \frac{1}{1} = 1.0000$$

$$\text{tf}_{\text{tiger}} = \frac{1}{1} = 1.0000$$

Doc 10

$$\text{tf}_{\text{bird}} = \frac{0}{2} = 0.0000$$

$$\text{tf}_{\text{cat}} = \frac{1}{2} = 0.5000$$

$$\text{tf}_{\text{dog}} = \frac{0}{2} = 0.0000$$

$$\text{tf}_{\text{tiger}} = \frac{2}{2} = 1.0000$$

חשבון המרחק = keyword function: cosine

	bird	cat	dog	tiger
Doc 1	0.3979	0.0305	0.1093	0.0000
Doc 2	0.0000	0.0458	0.0775	0.1109
Doc 3	0.3979	0.0000	0.0775	0.0000
Doc 4	0.0000	0.0458	0.0000	0.2218
Doc 5	0.0000	0.0193	0.0516	0.2218
Doc 6	0.0000	0.0498	0.0000	0.2218
Doc 7	0.3979	0.0498	0.1549	0.0000
Doc 8	0.3979	0.0498	0.1549	0.0000
Doc 9	0.0000	0.0498	0.1549	0.2218
Doc 10	0.0000	0.0229	0.0000	0.2218

$$m_1 = (1, 1, 1, 0)$$

$$m_2 = (0, 1, 1, 1)$$

$$m_3 = (0, 1, 0, 1)$$

$$m_4 = (1, 0, 1, 0)$$

$$k_1 = \frac{c_{1,1} m_1 + c_{1,4} m_4}{\sqrt{c_{1,1}^2 + c_{1,4}^2}}$$

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

$$= 0.3979 + 0.3979 + 0.3979$$

$$= 1.1937$$

$$k_1 = \frac{1.1937 m_1 + 0.3979 m_4}{\sqrt{1.4249 + 0.1583}}$$

$$c_{1,4} = w_{1,3} = 0.3979$$

$$k_1 = \frac{1.1937 m_1 + 0.3979 m_4}{\sqrt{1.5832}}$$

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

$$= 0.0709 + 0.0458 + 0.0458$$

$$= 0.1221$$

$$k_2 = \frac{c_{2,1} m_1 + c_{2,2} m_2 + c_{2,3} m_3}{\sqrt{c_{2,1}^2 + c_{2,2}^2 + c_{2,3}^2}}$$

$$c_{2,2} = w_{2,2} + w_{2,5} + w_{2,9}$$

$$= 0.0498 + 0.0193 + 0.0498$$

$$= 0.1069$$

$$k_2 = \frac{0.1221 m_1 + 0.1069 m_2 + 0.1145 m_3}{\sqrt{0.0149 + 0.0114 + 0.0131}}$$

$$c_{2,3} = w_{2,4} + w_{2,6} + w_{2,10}$$

$$= 0.0258 + 0.0458 + 0.0229$$

$$= 0.1145$$

$$k_2 = \frac{0.1221 m_1 + 0.1069 m_2 + 0.1145 m_3}{\sqrt{0.0394}}$$

$$k_3 = \frac{c_{3,1} m_1 + c_{3,2} m_2 + c_{3,4} m_4}{\sqrt{c_{3,1}^2 + c_{3,2}^2 + c_{3,4}^2}}$$

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

$$= 0.1093 + 0.1549 + 0.1549$$

$$= 0.4191$$

$$k_3 = \frac{0.4191 m_1 + 0.2840 m_2 + 0.0775 m_4}{\sqrt{0.1707 + 0.0807 + 0.0060}}$$

$$c_{3,2} = w_{3,2} + w_{3,5} + w_{3,9}$$

$$= 0.0775 + 0.0516 + 0.1549$$

$$= 0.2840$$

$$c_{3,4} = w_{3,3} = 0.0775$$

$$k_3 = \frac{0.4191 m_1 + 0.2840 m_2 + 0.0775 m_4}{\sqrt{0.2974}}$$

$$k_4 = \frac{C_{4,2} m_2 + C_{4,3} m_3}{\sqrt{C_{4,2}^2 + C_{4,3}^2}}$$

$$k_4 = \frac{0.9545 m_2 + 0.6679 m_3}{\sqrt{0.9075 + 0.4429}}$$

$$k_4 = \frac{0.9545 m_2 + 0.6679 m_3}{\sqrt{0.7703}}$$

$$\begin{aligned} C_{4,2} &= W_{4,2} + W_{4,9} + W_{4,8} \\ &= 0.1109 + 0.2218 + 0.2218 \\ &= 0.5545 \end{aligned}$$

$$\begin{aligned} C_{4,3} &= W_{4,4} + W_{4,6} + W_{4,10} \\ &= 0.2218 + 0.2218 + 0.2218 \\ &= 0.6654 \end{aligned}$$

Phonem Quera

Query = "cat cat cat dog dog tiger"

$$w_{cat,q} = \left( 0.5 + \left( \frac{0.5+3}{9} \right) \right) \times \log\left(\frac{10}{9}\right) = 0.0458$$

$$w_{dog,q} = \left( 0.5 + \left( \frac{0.5 \times 2}{3} \right) \right) \times \log\left(\frac{10}{7}\right) = 0.1291$$

$$w_{tiger,q} = \left( 0.5 + \left( \frac{0.5+1}{3} \right) \right) \times \log\left(\frac{10}{6}\right) = 0.1479$$

$$\begin{aligned} q &= 0.0458 k_2 + 0.1291 k_3 + 0.1479 k_4 \\ &= 0.0458 \left( \frac{0.1221 m_1 + 0.1069 m_2 + 0.1145 m_3}{\sqrt{0.0394}} \right) + 0.1291 \left( \frac{0.4191 m_1 + 0.2810 m_2 + 0.0775 m_3}{\sqrt{0.2934}} \right) + 0.1479 \left( \frac{0.5545 m_2 + 0.6679 m_3}{\sqrt{0.7703}} \right) \end{aligned}$$

$$q = 0.1333 m_1 + 0.1916 m_2 + 0.1400 m_3 + 0.0197 m_4$$

$$\begin{aligned} d_1 &= 0.3979 k_7 + 0.0305 k_2 + 0.1033 k_3 \\ &= 0.3979 \left( \frac{1.1937 m_1 + 0.3979 m_2}{\sqrt{1.5832}} \right) + 0.0305 \left( \frac{0.1221 m_1 + 0.1069 m_2 + 0.1145 m_3}{\sqrt{0.0394}} \right) + 0.1033 \left( \frac{0.4191 m_1 + 0.2810 m_2 + 0.0775 m_3}{\sqrt{0.2934}} \right) \\ &= 0.4804 m_1 + 0.2001 m_2 + 0.0176 m_3 + 0.0158 m_4 \end{aligned}$$

$$\begin{aligned} \text{Sim}(q, d_1) &= \frac{(0.1333)(0.4804) + (0.1916)(0.2001) + (0.1400)(0.0176) + (0.0197)(0.0158)}{\sqrt{(0.1333^2 + 0.1916^2 + 0.1400^2 + 0.0197^2) \times (0.4804^2 + 0.2001^2 + 0.0176^2 + 0.0158^2)}} \\ &= 0.7797 \end{aligned}$$

$$\begin{aligned} d_2 &= 0.0458 k_2 + 0.0775 k_3 + 0.1109 k_4 \\ &= 0.0458 \left( \frac{0.1221 m_1 + 0.1069 m_2 + 0.1145 m_3}{\sqrt{0.0394}} \right) + 0.0775 \left( \frac{0.4191 m_1 + 0.2810 m_2 + 0.0775 m_3}{\sqrt{0.2934}} \right) + 0.1109 \left( \frac{0.5545 m_2 + 0.6679 m_3}{\sqrt{0.7703}} \right) \\ &= 0.0913 m_1 + 0.1391 m_2 + 0.0118 m_4 \end{aligned}$$

$$\begin{aligned} \text{Sim}(q, d_2) &= \frac{(0.1333)(0.0913) + (0.1916)(0.1391) + (0.1400)(0.0118) + (0.0197)(0.0118)}{\sqrt{(0.1333^2 + 0.1916^2 + 0.1400^2 + 0.0197^2) \times (0.0913^2 + 0.1391^2 + 0.0118^2 + 0.0118^2)}} \\ &= 0.9939 \end{aligned}$$

$$d_3 = 0.3979k_1 + 0.0775k_3$$

$$= 0.3979 \left( \frac{1.1997m_1 + 0.3979m_2}{\sqrt{1.5832}} \right) + 0.0775 \left( \frac{0.4191m_1 + 0.2810m_2 + 0.0775m_3}{\sqrt{0.2974}} \right)$$

$$= 0.4406m_1 + 0.1692m_2 + 0.0118m_3$$

$$g_1m(g, d_3) = \frac{(0.1333)(0.4406) + (0.1916)(0.1692) + (0.0197)(0.0118)}{\sqrt{(0.1333^2 + 0.1916^2 + 0.1400^2 + 0.0197^2) \times (0.4406^2 + 0.1692^2 + 0.0118^2)}}$$

$$= 0.7093$$

$$d_4 = 0.0458k_2 + 0.2218k_4$$

$$= 0.0458 \left( \frac{0.1221m_1 + 0.1069m_2 + 0.1145m_3}{\sqrt{0.0294}} \right) + 0.2218 \left( \frac{0.5449m_1 + 0.6694m_2}{\sqrt{0.7703}} \right)$$

$$= 0.0281m_1 + 0.1666m_2 + 0.1968m_3$$

$$g_1m(g, d_4) = \frac{(0.1333)(0.0281) + (0.1916)(0.1666) + (0.1400)(0.1968)}{\sqrt{(0.1333^2 + 0.1916^2 + 0.1400^2 + 0.0197^2) \times (0.0281^2 + 0.1666^2 + 0.1968^2)}}$$

$$= 0.8932$$

$$d_5 = 0.0157k_2 + 0.0916k_3 + 0.2218k_4$$

$$= 0.0157 \left( \frac{0.1221m_1 + 0.1069m_2 + 0.1145m_3}{\sqrt{0.0294}} \right) + 0.0916 \left( \frac{0.4191m_1 + 0.2810m_2 + 0.0775m_3}{\sqrt{0.2974}} \right) + 0.2218 \left( \frac{0.5449m_1 + 0.6694m_2}{\sqrt{0.7703}} \right)$$

$$= 0.0514m_1 + 0.1791m_2 + 0.1792m_3 + 0.0079m_4$$

$$g_1m(g, d_5) = \frac{(0.1333)(0.0514) + (0.1916)(0.1791) + (0.0197)(0.0079)}{\sqrt{(0.1333^2 + 0.1916^2 + 0.1400^2 + 0.0197^2) \times (0.0514^2 + 0.1791^2 + 0.1792^2 + 0.0079^2)}}$$

$$= 0.9409$$

$$d_6 = 0.0458k_2 + 0.2218k_4$$

$$= 0.0458 \left( \frac{0.1221m_1 + 0.1069m_2 + 0.1145m_3}{\sqrt{0.0294}} \right) + 0.2218 \left( \frac{0.5449m_1 + 0.6694m_2}{\sqrt{0.7703}} \right)$$

$$= 0.0281m_1 + 0.1666m_2 + 0.1968m_3$$

$$g_1m(g, d_6) = \frac{(0.1333)(0.0281) + (0.1916)(0.1666) + (0.1400)(0.1968)}{\sqrt{(0.1333^2 + 0.1916^2 + 0.1400^2 + 0.0197^2) \times (0.0281^2 + 0.1666^2 + 0.1968^2)}}$$

$$d_7 = 0.3979k_1 + 0.0458k_2 + 0.1549k_3$$

$$= 0.3979 \left( \frac{1.1997m_1 + 0.3979m_2}{\sqrt{1.5832}} \right) + 0.0458 \left( \frac{0.1221m_1 + 0.1069m_2 + 0.1145m_3}{\sqrt{0.0294}} \right) + 0.1549 \left( \frac{0.4191m_1 + 0.2810m_2 + 0.0775m_3}{\sqrt{0.2974}} \right)$$

$$= 0.8932$$

$$g_1m(g, d_7) = \frac{(0.1333)(0.8932) + (0.1916)(0.2972) + (0.1400)(0.0264) + (0.0197)(0.0237)}{\sqrt{(0.1333^2 + 0.1916^2 + 0.1400^2 + 0.0197^2) \times (0.8932^2 + 0.2972^2 + 0.0264^2 + 0.0237^2)}}$$

$$= 0.7569$$

$$d_8 = 0.3979 k_1 + 0.0458 k_2 + 0.1549 k_3$$

$$= 0.3979 \left( \frac{1.1997 m_1 + 0.3979 m_2}{\sqrt{1.5832}} \right) + 0.0458 k_2 \left( \frac{0.1221 m_1 + 0.1069 m_2 + 0.1145 m_3}{\sqrt{0.0394}} \right) + 0.1549 \left( \frac{0.4191 m_1 + 0.2810 m_2 + 0.0775 m_3}{\sqrt{0.2924}} \right)$$

$$= 0.5318 m_1 + 0.2372 m_2 + 0.0264 m_3 + 0.0237 m_4$$

$$\text{Sim}(g, d_7) = \frac{(0.1333)(0.5318) + (0.1916)(0.2372) + (0.1400)(0.0264) + (0.0197)(0.0237)}{\sqrt{(0.1333^2 + 0.1916^2 + 0.1400^2 + 0.0197^2)} \times \sqrt{(0.5318^2 + 0.2372^2 + 0.0264^2 + 0.0237^2)}}$$

$$= 0.7569$$

$$d_9 = 0.0458 k_2 + 0.1549 k_3 + 0.2218 k_4$$

$$= 0.0458 \left( \frac{0.1221 m_1 + 0.1069 m_2 + 0.1145 m_3}{\sqrt{0.0394}} \right) + 0.1549 \left( \frac{0.4191 m_1 + 0.2810 m_2 + 0.0775 m_3}{\sqrt{0.2924}} \right) + 0.2218 \left( \frac{0.5445 m_1 + 0.6654 m_2}{\sqrt{0.7707}} \right)$$

$$= 0.1543 m_1 + 0.2554 m_2 + 0.1968 m_3 + 0.0237 m_4$$

$$q_{im}(g, d_9) = \frac{(0.1333)(0.1543) + (0.1916)(0.2554) + (0.1400)(0.1968) + (0.0197)(0.0237)}{\sqrt{(0.1333^2 + 0.1916^2 + 0.1400^2 + 0.0197^2)} \times \sqrt{(0.1543^2 + 0.2554^2 + 0.1968^2 + 0.0237^2)}}$$

$$= 0.9976$$

$$d = 0.0229 k_2 + 0.2218 k_4$$

$$= 0.0229 \left( \frac{0.1221 m_1 + 0.1069 m_2 + 0.1145 m_3}{\sqrt{0.0394}} \right) + 0.2218 \left( \frac{0.5445 m_1 + 0.6654 m_2}{\sqrt{0.7707}} \right)$$

$$= 0.0191 m_1 + 0.1543 m_2 + 0.1876 m_3$$

$$q_{im}(g, d) = \frac{(0.1333)(0.0191) + (0.1916)(0.1543) + (0.1400)(0.1876)}{\sqrt{(0.1333^2 + 0.1916^2 + 0.1400^2 + 0.0197^2)} \times \sqrt{(0.0191^2 + 0.1543^2 + 0.1876^2)}}$$

$$= 0.8717$$

$$m_1 = (1, 1, 1, 0)$$

$$m_2 = (0, 1, 1, 1)$$

$$m_3 = (0, 1, 0, 1)$$

$$m_4 = (1, 0, 1, 0)$$

	bird	cat	dog	tiger
Doc 1	3	2	2	0
Doc 2	0	2	1	1
Doc 3	2	0	1	0
Doc 4	0	1	0	1
Doc 5	0	1	1	3
Doc 6	0	2	0	2
Doc 7	1	1	1	0
Doc 8	1	1	1	0
Doc 9	0	1	1	1
Doc 10	0	1	0	2

Query = "cat cat cat dog dog tiger"

