

Quiz 2_02

ข้อ 2. สมมติในระบบมีเอกสาร 5 เอกสารดังนี้ (คำที่อยู่ในปีกกา คือ Keyword ซึ่ง Keyword มีความสัมพันธ์กัน)

D1 : {bird, cat, bird, cat, tiger, kitty, fish, fish}

D2 : {cat, dog, tiger, zoo}

D3 : {house, kitchen, bird, bird, cat, cat, kitty}

D4 : {dog, rubber, house, dog}

D5 : {tiger, forest, fish}

โดยมี Query : "ต้องการเอกสารที่มี cat หรือ kitty แต่ไม่ต้องการเอกสารที่มี dog "

2.1 เพื่อให้ได้คำตอบในคำถาม 2.2 จงให้เหตุผลว่า "ควรเลือกใช้โมเดลใดเพราะอะไร"

A) Boolean Model B) Fuzzy Model C) Extend Boolean Model D) Vector Model

2.2 ให้นักศึกษาแสดงวิธีคำนวณหา Ranking ของเอกสารแต่ละเอกสาร เพื่อเรียงลำดับเอกสารที่จะแสดงผลให้กับผู้เรียกค้น (27 คะแนน)

ข้อสอบ 1/2558

(ดัดแปลง)

Quiz 2_02

Answer

2.1 เลือกใช้ Fuzzy Model เนื่องจาก keyword มีความสัมพันธ์กัน และคำเรียกค้นเป็นแบบ Boolean

2.2 ความสัมพันธ์ระหว่าง Keyword

bird → cat, fish, house, kitchen, kitty, tiger
cat → bird, dog, fish, house, kitchen, kitty, tiger, zoo
dog → cat, house, rubber, tiger, zoo
fish → bird, cat, forest, kitty, tiger
forest → fish, tiger
house → bird, cat, dog, kitchen, kitty, rubber
kitchen → bird, cat, house, kitty
kitty → bird, cat, fish, house, kitchen, tiger
rubber → dog, house
tiger → bird, cat, dog, fish, forest, kitty, zoo
zoo → cat, dog, tiger

Quiz 2_02

Answer

คำนวณความสับสน

$$c_{i,j} = \frac{n_{i,j}}{n_i + n_j - n_{i,j}}$$

bird,cat	=	2/(2+3-2) = 0.67	cat,dog	=	1/(3+2-1) = 0.25
bird,fish	=	1/(2+2-1) = 0.33	cat,fish	=	1/(3+2-1) = 0.25
bird,house	=	1/(2+2-1) = 0.33	cat,house	=	1/(3+2-1) = 0.25
bird,kitchen	=	1/(2+1-1) = 0.50	cat,kitchen	=	1/(3+1-1) = 0.33
bird,kitty	=	2/(2+2-2) = 1.00	cat,kitty	=	2/(3+2-2) = 0.67
bird,tiger	=	1/(2+3-1) = 0.25	cat,tiger	=	2/(3+3-2) = 0.50
			cat,zoo	=	1/(3+1-1) = 0.33

Quiz 2_02

dog,house	= $1/(2+2-1) = 0.33$	fish,forest	= $1/(2+1-1) = 0.50$
dog,rubber	= $1/(2+1-1) = 0.50$	fish,kitty	= $1/(2+2-1) = 0.33$
dog,tiger	= $1/(2+3-1) = 0.25$	fish,tiger	= $2/(2+3-2) = 0.67$
dog,zoo	= $1/(2+1-1) = 0.50$		
		forest,tiger	= $1/(1+3-1) = 0.33$

house,kitchen	= $1/(2+1-1) = 0.50$
house,kitty	= $1/(2+2-1) = 0.33$
house,rubber	= $1/(2+1-1) = 0.50$

kitchen,kitty	= $1/(1+2-1) = 0.50$
---------------	----------------------

kitty,tiger	= $1/(2+3-1) = 0.25$
-------------	----------------------

tiger,zoo	= $1/(3+1-1) = 0.33$
-----------	----------------------

Quiz 2_02

	bird	cat	dog	fish	forest	house	kitchen	kitty	rubber	tiger	zoo
bird	1.00	0.67	0.00	0.33	0.00	0.33	0.50	1.00	0.00	0.25	0.00
cat	0.67	1.00	0.25	0.25	0.00	0.25	0.33	0.67	0.00	0.50	0.33
dog	0.00	0.25	1.00	0.00	0.00	0.33	0.00	0.00	0.50	0.25	0.50
fish	0.33	0.25	0.00	1.00	0.50	0.00	0.00	0.33	0.00	0.67	0.00
forest	0.00	0.00	0.00	0.50	1.00	0.00	0.00	0.00	0.00	0.33	0.00
house	0.33	0.25	0.33	0.00	0.00	1.00	0.50	0.33	0.50	0.00	0.00
kitchen	0.50	0.33	0.00	0.00	0.00	0.50	1.00	0.50	0.00	0.00	0.00
kitty	1.00	0.67	0.00	0.33	0.00	0.33	0.50	1.00	0.00	0.25	0.00
rubber	0.00	0.00	0.50	0.00	0.00	0.50	0.00	0.00	1.00	0.00	0.00
tiger	0.25	0.50	0.25	0.67	0.33	0.00	0.00	0.50	0.00	1.00	0.33
zoo	0.00	0.33	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.33	1.00

Quiz 2_02

D1: {bird, cat, bird, cat, tiger, kitty, fish, fish}

D2: {cat, dog, tiger, zoo}

D3: {house, kitchen, bird, bird, cat, cat, kitty}

D4: {dog, rubber, house, dog}

D5: {tiger, forest, fish}

	Cat	Kitty	Dog
D1	1	1	0.25
D2	1	0.67	1
D3	1	1	0.33
D4	0.25	0.33	1
D5	0.50	0.33	0.25

Quiz 2_02

Query : "ต้องการเอกสารที่มี cat หรือ kitty แต่ไม่ต้องการเอกสารที่มี dog "

Query = (cat OR kitty) AND not dog

$$\mu_{\bar{A}}(x) = (1 - \mu_A(x))$$

$$\mu_{A \cup B}(x) = 1 - (1 - \mu_A(x)) \cdot (1 - \mu_B(x))$$

$$\mu_{A \cap B}(x) = \mu_A(x) \cdot \mu_B(x)$$

	Cat	Kitty	Dog
D1	1	1	0.25
D2	1	0.67	1
D3	1	1	0.33
D4	0.25	0.33	1
D5	0.50	0.33	0.25

$$\therefore \text{Ranking} = \left(1 - (1 - \mu_{cat}) \cdot (1 - \mu_{kitty})\right) \cdot (1 - \mu_{dog})$$

$$\text{Doc}_1 = (1 - (1 - 1) \cdot (1 - 1)) \cdot (1 - 0.25) = 1 \cdot 0.75 = 0.75$$

$$\text{Doc}_2 = (1 - (1 - 1) \cdot (1 - 0.67)) \cdot (1 - 1) = 1 \cdot 0 = 0.00$$

$$\text{Doc}_3 = (1 - (1 - 1) \cdot (1 - 1)) \cdot (1 - 0.33) = 1 \cdot 0.67 = 0.67$$

$$\text{Doc}_4 = (1 - (1 - 0.25) \cdot (1 - 0.33)) \cdot (1 - 1) = 0.49 \cdot 0 = 0.00$$

$$\text{Doc}_5 = (1 - (1 - 0.50) \cdot (1 - 0.33)) \cdot (1 - 0.25) = 0.67 \cdot 0.75 = 0.50$$

$$\text{Ranking} = \text{Doc}_1, \text{Doc}_3, \text{Doc}_5, \text{Doc}_2, \text{Doc}_4$$