

Boolean Model

หน้า

1

บทที่

2

Boolean query expression 3

ข้อ 2. สมมติในระบบมีเอกสารทั้งหมด 10 เอกสารดังนี้ (bird, cat, dog, tiger คือ Keyword โดย Keyword เหล่านี้ไม่มี ความสัมพันธ์กัน)

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 "I don't love cat and dog. But I love tiger as well as bird." เข้าไปในระบบ จงตอบคำถาม

2.1 เพื่อให้ได้คำตอบในคำถาม 2.2 ผู้ใช้สามารถเลือกใช้โมเดลใดได้บ้างเพราะอะไร (เลือกได้เฉพาะตัวเลือกที่ให้มา)

A) Vector Model

B) Boolean Model

Ans เลือกใช้ Boolean Model เพราะ query มาเป็นประโยคที่มีความหมาย สามารถเทียบได้กับ Boolean expression และใช้ค้นหาพจน์ keyword ที่มีคอนสัมพันธ์กัน

2.2 ให้นักศึกษาแสดงวิธีคำนวณหา Ranking ของเอกสารทุกเอกสารในระบบ ตามที่ผู้ใช้งานต้องการ

จากเอกสาร 10 เอกสารที่กำหนดให้

Query : $\neg(\text{cat and dog}) \text{ and } (\text{tiger or bird})$

กำหนดหา Ranking

\wedge and
 \vee or
 \neg not

	$\neg(\text{cat and dog})$	(tiger or bird)	Relevance
D ₁	0	1	0
D ₂	0	1	0
D ₃	0	1	0
D ₄	0	1	0
D ₅	0	1	0
D ₆	0	1	0
D ₇	0	1	0
D ₈	0	1	0
D ₉	0	1	0
D ₁₀	1	1	1

$$\therefore \neg(\text{cat and dog}) \text{ and } (\text{tiger or bird}) = \{D_{10}\} \wedge \{D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8, D_9, D_{10}\} \\ = \{D_{10}\}$$

Ans Ranking ของเอกสารทุกเอกสารในระบบตามผู้ใช้งานคือ D₁₀

Trick

1. แปลง query

2. ทำตารางตาม query

!"#\$%&

2.1 ใช้!!""#\$%&'("\$)\$# เนื่องจากลักษณะของ!*+\$,-'เป็นแบบ!!""#\$%&'และโจทย์กำหนดให้!.\$-/"',)'ไม่สัมพันธ์กัน

Query = “I don’t love cat and dog. But I love tiger as well as bird”
 ไม่รักแมวและสุนัข'แสดงว่าไม่ต้องการเอกสารที่มีแมวหรือมีสุนัข
 รักเสือพอๆกับรักนก'แสดงว่าต้องการเอกสารที่มีนกหรือมีเสือ

Query = (Bird OR Tiger) AND NOT(Cat OR Dog)

1 2
Query = (Bird OR Tiger) AND NOT(Cat OR Dog)

	Bird	Cat	Dog	Tiger	1	2	AND
D1	1	1	1	0	1	0	0
D2	0	1	1	1	1	0	0
D3	1	0	1	0	1	0	0
D4	0	1	0	1	1	0	0
D5	0	1	1	1	1	0	0
D6	1	1	0	1	1	0	0
D7	1	1	1	1	1	0	0
D8	1	1	1	0	1	0	0
D9	0	1	1	1	1	0	0
D10	0	0	0	1	1	1	1

เอกสารที่ตรงประเด็นคือ **D10**

minimum Boolean = 9d non-keyword function

Boolean query extensions

D1 = (0.8A, 0.5B, 0.6C)

D2 = (0.4A, 0.4B, 0.1C, 0.8D)

Query: (0.5A and 0.2B) Or (not D Or 0.3C)

Ex. search not (0.5A and 0.2B)
 $1 - (0.5 \times 0.8), 1 - (0.2 \times 0.6)$

AND 100%
OR 100%
NOT 100% - 100%

	A and B	Not D	Relevance
D1	$\min(\overset{\text{Doc}}{0.5 \times 0.8}, \overset{\text{Doc}}{0.2 \times 0.5})$ = 0.10 <i>query</i>	$\max(\underset{\substack{\text{not D} \\ 1-0}}{1}, \overset{\text{query}}{0.3 \times 0.6}) = 1$	$\max(0.1, 1) = 1$
D2	$\min(\overset{\text{Doc}}{0.5 \times 0.4}, \overset{\text{Doc}}{0.2 \times 0.4})$ = 0.08 <i>query</i>	$\max(\underset{1-0.8}{0.2}, \overset{\text{query}}{0.3 \times 0.1}) = 0.2$	$\max(0.08, 0.2) = 0.2$