



มหาวิทยาลัยรามคำแหง

ข้อสอบวิชา COS4351 ระบบค้นคืนสารสนเทศ ชุดที่ 2

ชื่อ.....รหัส.....

คำสั่ง ข้อสอบมีทั้งหมด 10 ข้อ รวม 100 คะแนน ทำในสมุดคำตอบทั้งหมด

ข้อที่ 1. กำหนดเอกสารมีค่าสำคัญดังนี้

DOC1 = { ANT , CAT , DOG, ELEPHANT , MONKEY, HORSE, ZOO}

DOC2 = { BIRD , DOG , ELEPHANT , FOX ,ZOO}

จงหาความคล้ายคลึงกันของเอกสารโดยวิธี

1.1 Simple matching

1.2 Dice's Coefficient

1.3 Jaccard's Coefficient

1.4 Overlap Coefficient

1.5 จงสร้างโครโมโซมต้นกำเนิดตามเอกสารที่กำหนดโดยวิธีจีเนติกอัลกอริทึม

ข้อที่ 2. จงอธิบายคำศัพท์ต่อไปนี้มาพอเข้าใจ

2.1 clustering

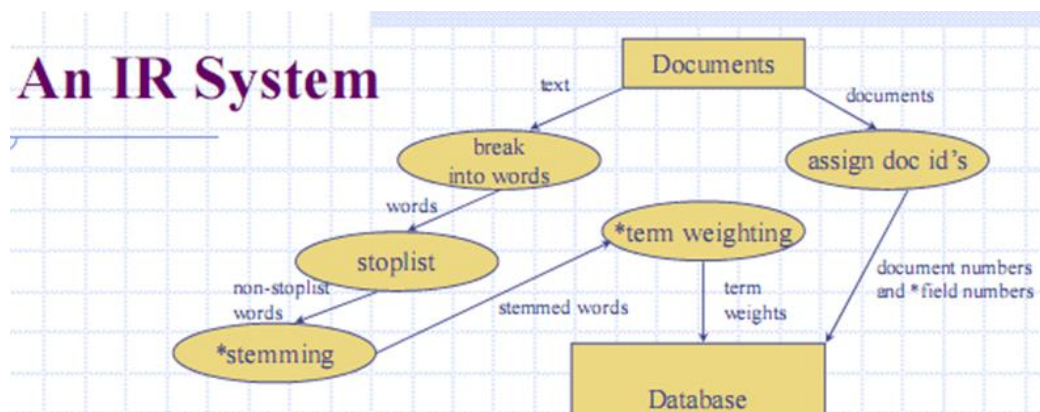
2.2 Feedback

2.3 Stopword

2.4 Thesaurus

2.5 Fallout

ข้อที่ 3. พิจารณารูปของระบบ IR ต่อไปนี้



จงอธิบายถึงขั้นตอนการทำงานของระบบ IR ข้างต้นให้เข้าใจ

ข้อที่ 4. กำหนดความสามารถของระบบ IR ระบบ1 และระบบ 2 ดังนี้

ความสามารถ	ระบบ 1	ระบบ 2
HIT	3000	2000
WASTED	4000	1000
MISSED	6000	4000
PASSED	3000	2000

4.1 จงหาค่าของ Recall, Precision ของระบบทั้งสอง

4.2 ระบบใดมีประสิทธิภาพดีกว่าจงให้เหตุผลพร้อมอธิบายให้เข้าใจ

ข้อที่ 5. กำหนดให้ $\{A, B, C\}$ เป็นเซตของสัญลักษณ์ของคีย์ R1 ,R2, R3, R4, R5,R6

เป็นเรคอร์ดที่เก็บบันทึกในโครงสร้างนี้ สมมติว่าคีย์แต่ละเรคอร์ดมีขนาดไม่เท่ากันดังนี้

R1	มีคีย์ว่า	ABC
R2	มีคีย์ว่า	ABAB
R3	มีคีย์ว่า	BBC
R4	มีคีย์ว่า	BCC
R5	มีคีย์ว่า	CA
R6	มีคีย์ว่า	CBC

จงสร้าง Double Chained Tree จากข้อมูลข้างต้น(วาดรูป) (5 คะแนน)

ข้อที่ 6. กำหนดให้ BA ,FI, MI ,NA , RM แทนเมืองในประเทศอิตาลี จงกระทำการ

ของ Hierarchical clustering ในการหาระยะทางระหว่างเมืองต่างๆ โดยระยะทางมีหน่วยเป็นกิโลเมตร โดยใช้อัลกอริทึม single-linkage พร้อมวาดรูปต้นไม้

	BA	FI	MI	NA	RM
BA	0	200	300	50	400
FI	200	0	120	600	800
MI	300	120	0	90	100
NA	50	600	90	0	70
RM	400	800	100	70	0

ข้อที่ 7. Information Retrieval แตกต่างจาก ระบบสารสนเทศอื่นๆอย่างไร

จงอธิบายถึงส่วนประกอบของ IR ในส่วนของ INPUT, PROCESS และ OUTPUT

ข้อที่ 8.(15 คะแนน) สมมติว่า เอกสาร D1, D2, D3, D4 ,D5 มีคีย์เวิร์ดดังนี้

D1 = (K1 ,K2, K3, K5)

D2 = (K2, K3, K5)

D3 = (K2, K4, K5)

D4 = (K1, K3, K5, K6)

D5 = {K3 , K4, K5, K6}

8.1 เมื่อนำมาเก็บในโครงสร้างของ Inverted file จะเก็บข้อมูลในลักษณะใด

8.2 ถ้ามีการค้นหาโดยใช้ Boolean Search ดังนี้

(K1 and K3) or (K3 and K5) or (K3 and K6) ผลเป็นอย่างไร

8.3 ถ้ากำหนดค่าทุกคำมีน้ำหนักเท่ากันเท่ากับ 1 และระบบกำหนดค่า RSV เท่ากับ 2

สมมติว่าผู้ใช้ป้อนคำค้นหาดังนี้ K1 K2 K3 ผลการทำงานเป็นอย่างไร

ข้อที่ 9. กำหนดนิยามของคลัสเตอร์ในรูปแบบของกราฟ ที่ได้จากการวัดความคล้ายคลึง

กันดังนี้ (object : 1,2,3,4,5,6)

Similarity matrix	1					
	2	.8				
	3	.6	.5			
	4	.8	.6	.6		
	5	.3	.8	.3	.8	
	6	.6	.5	.2	.6	.2
		1	2	3	4	5

9.1 จงแสดงกราฟถ้ากำหนดค่า Threshold เท่ากับ 0.75

9.2 จงแสดงกราฟถ้ากำหนดค่า Threshold เท่ากับ 0.80

ข้อที่ 10. การทำงานของเสิร์จเอนจิน(Search Engine)

10.1 โดยทั่วไปประกอบด้วย 3 ส่วนหลักอะไรบ้างจงอธิบายให้เข้าใจ

10.2 Search Engine ชื่อ Yahoo แตกต่างจาก Mamma อย่างไร