

!"#\$%&' สมมติบริษัทแห่งหนึ่งมีพนักงานทั้งหมด!"#!คน!\$ชื่อ!%&' &(8)&*+!!ซึ่งมีข้อมูลตามตาราง
ด้านล่าง

Name	Address	Room#
A	Bangkok	101
B	Nonthaburi	201
C	Lopburi	202
D	Bangkok	102
E	Bangkok	103
F	Bangkok	104
G	Lopburi	203
H	Chiangmai	204
I	Nonthaburi	302
J	Bangkok	301

ส่งคำตอบที่ : **IR.CE.KMITL@Gmail.com**

ชื่อเมล : **Quiz3_รหัสนักศึกษา** เช่น

Quiz3_62010109_62010198

ไฟล์แนบชื่อใดก็ได้

ส่งภายใน **8/08/2566** ไม่เกินเที่ยงคืน

% เป็นโปรแกรมเมอร์มือหนึ่งของบริษัทและ!%!เป็นพนักงานที่กว้างขวางมีเพื่อนสนิทหลายคน!คือ! (*)+),)- !ในโรงงาน!% ได้สร้าง
ระบบเรียกค้นขึ้น!ซึ่งภายในประกอบด้วย!,!อัลกอริธึม!ต่อมา!% ต้องการทราบว่าอัลกอริธึมใดเหมาะสมกว่ากันเมื่อเรียกค้นพนักงานที่มีที่อยู่ใน!

' -./010 ซึ่งให้ผลลัพธ์ดังนี้

อัลกอริธึมที่!"!ให้ผลลัพธ์ตามลำดับคือ./*)()+,)0)!

อัลกอริธึมที่ ,!ให้ผลลัพธ์ตามลำดับคือ.-)0)/1*)+)(

จากผลลัพธ์ที่ได้นี้ให้นักศึกษาคำนวณหาว่าอัลกอริธึมใดเหมาะสมกว่ากัน!ด้วยตัวเลขเท่าใดบ้าง!\$ทุกวิธีที่สามารถประเมินได้+! 234.

คะแนน5.

/678\$&

ข้อมูลโดยตรงประเด็นทั้งหมด!2!ข้อมูลคือ

A, D, E, F, J

1	A	D	B	E	I	F	H	AVG
Precision	1.00	1.00	0.67	0.75	0.60	0.67	0.57	0.85
Recall	0.20	0.40	0.40	0.60	0.60	0.80	0.80	-
F	0.33	0.57	0.50	0.67	0.60	0.73	0.67	0.57
E(β=2)	0.24	0.45	0.43	0.63	0.60	0.77	0.74	0.52

$$Coverage = \frac{R_k}{U}$$

$$Novelty = \frac{R_u}{R_u + R_k}$$

R_k = Relevant Docs known to the user which were retrieved

R_u = Relevant Docs previously unknown to the user which were retrieved

U = Relevant Docs known to the user

A สนิทกับ **B, D, E, I, J** จึงทราบว่า **A, D, E, J** ตรงประเด็น

$$Coverage = \frac{R_k}{U} = \frac{\text{รู้ว่าตรงประเด็นแล้วออกมา}(A, D, E)}{\text{รู้ว่าตรงประเด็น}} = \frac{3}{4} = 0.75$$

$$Novelty = \frac{R_u}{R_u + R_k} = \frac{\text{ไม่รู้ว่าตรงประเด็นแล้วออกมา}(F)}{\text{ไม่รู้ว่าตรงประเด็นแล้วออกมา}(F) + \text{รู้ว่าตรงประเด็นแล้วออกมา}(A, D, E)} = \frac{1}{1 + 3} = 0.25$$

$$F = \frac{2PR}{P + R} = \frac{2}{\frac{1}{R} + \frac{1}{P}}$$

$$E = \frac{(1 + \beta^2)PR}{\beta^2 P + R} = \frac{(1 + \beta^2)}{\frac{\beta^2}{R} + \frac{1}{P}}$$

2	J	F	A	C	D	E	B	AVG
Precision	1.00	1.00	1.00	0.75	0.80	0.83	0.71	0.93
Recall	0.20	0.40	0.60	0.60	0.80	1.00	1.00	-
F	0.33	0.57	0.75	0.67	0.80	0.91	0.83	0.67
E($\beta=2$)	0.24	0.45	0.65	0.63	0.80	0.96	0.93	0.62

$$Coverage = \frac{R_k}{U}$$

$$Novelty = \frac{R_u}{R_u + R_k}$$

R_k = Relevant Docs known to the user which were retrieved

R_u = Relevant Docs previously unknown to the user which were retrieved

U = Relevant Docs known to the user

A สนิทกับ **B, D, E, I, J** จึงทราบว่า **A, D, E, J** ตรงประเด็น

$$Coverage = \frac{R_k}{U} = \frac{\text{รู้ว่าตรงประเด็นแล้วออกมา}(A, D, E, J)}{\text{รู้ว่าตรงประเด็น}} = \frac{4}{4} = 1.00$$

$$Novelty = \frac{R_u}{R_u + R_k} = \frac{\text{ไม่รู้ว่าตรงประเด็นแล้วออกมา}(F)}{\text{ไม่รู้ว่าตรงประเด็นแล้วออกมา}(F) + \text{รู้ว่าตรงประเด็นแล้วออกมา}(A, D, E, J)} = \frac{1}{1 + 4} = 0.20$$

$$F = \frac{2PR}{P + R} = \frac{2}{\frac{1}{R} + \frac{1}{P}}$$

$$E = \frac{(1 + \beta^2)PR}{\beta^2 P + R} = \frac{(1 + \beta^2)}{\frac{\beta^2}{R} + \frac{1}{P}}$$

อัลกอริธึมที่2 มีความตรงประเด็นมากกว่าอัลกอริธึมที่ 1 ด้วยเครื่องมือวัดทุกเครื่องมือ