

บทที่ 4

ผลการทดลองและบทวิเคราะห์

ในบทนี้จะแสดงผลการวิเคราะห์ข้อมูลการเปรียบเทียบประสิทธิภาพของการค้นคืนเอกสาร 2 รูปแบบ คือ การค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน (Euclidean Distance) ภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering) และการค้นคืนเอกสารด้วยวิธีการวัดความคล้ายคลึงเชิงมุม (Cosine Angle) เพื่อนำมาตอบวัตถุประสงค์ของงานวิจัย

โดยจะกล่าวถึง การแสดงผลการวิเคราะห์ข้อมูลจากการกำหนดค่าตัวแปรอื่น ๆ ที่เกี่ยวข้องกับ การทดลอง คือ จำนวนกลุ่มของเอกสาร และค่ารัศมีเพื่อกำหนดกรอบความคล้ายคลึง การค้นคืน การแสดงผลการวิเคราะห์ทดลองเลือกกำหนดกรอบค่าความคล้ายคลึงที่เหมาะสม สำหรับใช้ในการค้นคืนเอกสารรูปแบบที่ 1 การวิเคราะห์ผลการทดลองและสรุปผลการเปรียบเทียบประสิทธิภาพของการค้นคืนเอกสารรูปแบบที่ 1 กับการค้นคืนเอกสารรูปแบบที่ 2 การวิเคราะห์ผลการทดลองปรับปรุงการค้นคืนเอกสารรูปแบบที่ 1 ด้วยการพิจารณาค้นคืนเฉพาะเอกสารที่มีค่าในข้อสอบถามปรากฏ และแสดงผลการวิเคราะห์ทดลองการเปรียบเทียบประสิทธิภาพกับการค้นคืนเอกสารรูปแบบที่ 2 และสรุปผลการศึกษา ดังรายละเอียดต่อไปนี้

4.1 การกำหนดค่าตัวแปรอื่น ๆ ที่เกี่ยวข้องกับการทดลอง

4.1.1 การพิจารณาเลือกจำนวนกลุ่มของเอกสารที่เหมาะสม

เนื่องจากการดำเนินการทดลองการค้นคืนเอกสารด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน (Euclidean Distance) ภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering) แบบ K-Means Clustering จำเป็นที่จะต้องมีการระบุกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล

ในเทคนิค K-Means Clustering ซึ่งใช้เป็นเทคนิคการจัดกลุ่มข้อมูลในการศึกษารั้งนี้ ต้องมีการกำหนดจำนวนกลุ่มที่ต้องการแบ่ง หรือกำหนดค่า K ไว้ล่วงหน้าสำหรับจัดกลุ่มข้อมูล ดังนั้นจึงต้องทดลองกำหนดค่า K ที่แตกต่างกันออกไป แล้วพิจารณาหาจำนวนกลุ่มหรือค่า K ที่เหมาะสมจากผลลัพธ์ของการจัดกลุ่มแต่ละครั้ง

ในงานวิจัยนี้ได้ทดลองกำหนดกลุ่มให้กับชุดเอกสารนิตยสารใหม่ซึ่งเป็นชุดเอกสารที่ใช้ในการศึกษา ซึ่งมีจำนวน 425 เอกสาร โดยใช้โปรแกรมแซสเอนเตอร์ไพส์ไมน์เนอร์ 5.1 (SAS Enterprise Miner 5.1) และเพื่อความสะดวกได้ทดลองกำหนดจำนวนกลุ่มที่ต้องการแบ่งเป็น 5, 10, 15, 20, 25 และ 30 กลุ่ม¹ ผลการทดลองที่ได้แสดงได้ดังตารางที่ 4.1

ตารางที่ 4.1 ตารางแสดงผลการจัดกลุ่มเอกสารด้วยการกำหนดจำนวนกลุ่มที่แตกต่างกัน

Number of cluster (K)	Root-Mean-Square Standard Deviation	Average Maximum Distance from Cluster Seed	Average Distance to Nearest Cluster
5	1.073	139.954	184.979
10	1.855	119.761	163.597
15	1.873	79.761	136.670
20	1.596	70.754	144.109
25	1.188	62.813	134.781
30	1.135	51.338	133.389

จากตารางที่ 4.1 ค่าสถิติต่าง ๆ ที่ได้จากการจัดกลุ่ม อธิบายได้ดังนี้

- Root-Mean-Square Standard Deviation เป็นค่าแสดงลักษณะการกระจายตัวของข้อมูลภายในกลุ่ม
 - Maximum Distance from Cluster Seed แสดงค่าระยะห่างมากที่สุดของข้อมูลในกลุ่ม จากจุดศูนย์กลางของกลุ่มนั้น ๆ
 - Distance to Nearest Cluster แสดงค่าระยะห่างระหว่างกลุ่มข้อมูลที่อยู่ใกล้กันมากที่สุด
- ค่าสถิติทั้ง 3 ค่าที่แสดงในตารางเป็น ค่าเฉลี่ย (Average) ของค่าสถิติเหล่านั้น ๆ กล่าวคือ จากตารางที่ 4.1 หากพิจารณากำหนดกลุ่มเอกสาร 15 กลุ่ม ค่าสถิติ Root-Mean-Square Standard Deviation คือ ค่าเฉลี่ยของค่า Root-Mean-Square Standard Deviation ของทั้ง 15 กลุ่ม เท่ากับ 1.873 ค่าสถิติ Maximum Distance from Cluster Seed คือ ค่าเฉลี่ยของค่า Maximum Distance from Cluster Seed ของทั้ง 15 กลุ่ม เท่ากับ 79.761 และค่าสถิติ Distance to Nearest Cluster คือค่าเฉลี่ยของค่า Distance to Nearest Cluster ของทั้ง 15 กลุ่ม เท่ากับ 136.670

¹ การเลือกกำหนดจำนวนกลุ่ม สามารถเลือกทดลองกำหนดจำนวนกลุ่มที่ต้องการแบ่งให้กับชุดเอกสาร ด้วยวิธีอื่น ๆ ที่แตกต่างกันได้ เช่น กำหนดจำนวนกลุ่มที่ต้องการแบ่งเป็น 25, 50, 100 และ 200 เป็นต้น

จากผลการทดลองที่ได้ตั้งข้างต้น การพิจารณาวิเคราะห์กำหนดจำนวนกลุ่มเอกสารที่เหมาะสมสำหรับในงานวิจัยนี้ มีแนวทางดังนี้

- Root-Mean-Square Standard Deviation เป็นค่าแสดงลักษณะการกระจายตัวของข้อมูลภายในกลุ่ม จะถูกใช้วัดความเหมือนของข้อมูลภายในกลุ่ม ซึ่งลักษณะของการแบ่งกลุ่มเอกสารที่ดีนั้น ข้อมูลภายในกลุ่ม การกระจายตัวน้อย ดังนั้นค่า Root-Mean-Square Standard Deviation ของลักษณะกลุ่มข้อมูลที่ดีควรมีค่าน้อย
- Maximum Distance from Cluster Seed เป็นค่าสถิติแสดงค่าระยะทางมากที่สุดระหว่างจุดศูนย์กลางของกลุ่มไปยังข้อมูลภายในกลุ่มนั้น ๆ ในทำนองเดียวกันกับค่า Root-Mean-Square Standard Deviation ถ้าข้อมูลภายในกลุ่มมีการกระจุกตัว ค่า Maximum Distance from Cluster Seed ควรจะมีค่าน้อย
- Distance to Nearest Cluster เป็นค่าแสดงระยะทางระหว่างกลุ่มข้อมูลที่อยู่ใกล้กัน ซึ่งลักษณะของการแบ่งกลุ่มเอกสารที่ดีนั้น ควรจะมีระยะห่างระหว่างกลุ่มเอกสารมาก เพื่อแสดงให้เห็นว่าข้อมูลมีการจัดกลุ่มอย่างชัดเจน ถ้าระยะห่างระหว่างกลุ่มข้อมูลที่อยู่ใกล้กันมีค่าน้อย แสดงว่ากลุ่มข้อมูลไม่สามารถแยกความแตกต่างระหว่างกลุ่มออกจากกันได้เท่าที่ควร

จากการนำผลสถิติทั้ง 3 ค่าที่ได้จากการทดลองกำหนดกลุ่มให้กับเอกสารเป็น 5, 10, 15, 20, 25 และ 30 กลุ่ม มาพิจารณาตามหลักเกณฑ์ที่ได้กล่าวมาแล้ว พบว่าการกำหนดจำนวนกลุ่มเป็น 15, 20 และ 25 กลุ่ม จะมีค่าสถิติทั้ง 3 ค่าที่เหมาะสม ดังแสดงในตารางที่ 4.2

ตารางที่ 4.2 ตารางแสดงจำนวนกลุ่มที่มีความเหมาะสมสำหรับการแบ่งกลุ่มเอกสารนิตยสารใหม่

Number of cluster	Root-Mean-Square Standard Deviation	Maximum Distance from Cluster Seed	Distance to Nearest Cluster
5	1.591074381	104.1032572	432.5980508
10	1.273479924	108.5099462	294.86809
15	1.294669273	59.61353283	281.8335282
20	1.243828449	87.95557696	227.6590004
25	1.155694878	63.35062315	229.0007343
30	1.129279472	66.93778436	212.310915

ในขั้นตอนต่อไป ผู้วิจัยได้นำลักษณะของกลุ่มที่เป็นผลลัพธ์จากการกำหนดจำนวนกลุ่ม เป็น 15, 20 และ 25 มาพิจารณาเพื่อตัดสินใจเลือกจำนวนกลุ่มใดเป็นจำนวนที่เหมาะสมที่สุด สำหรับชุดเอกสารนี้ โดยจะพิจารณาจากจำนวนของเอกสารในแต่ละกลุ่ม (Frequency of Cluster) หรือขนาดของกลุ่ม

จำนวนเอกสารในแต่ละกลุ่มที่เป็นผลลัพธ์จากการกำหนดจำนวนกลุ่มเป็น 15, 20 และ 25 แสดงภายในกรอบเส้นประในตารางที่ 4.3, 4.4 และ 4.5 โดยในการจัดกลุ่มเอกสารที่ดีนั้น ขนาดของกลุ่มต้องไม่ใหญ่ หรือเล็กจนเกินไป²

ตารางที่ 4.3 ตารางค่าสถิติของกลุ่มเมื่อกำหนดจำนวนกลุ่มเท่ากับ 15 กลุ่ม

Cluster->15					
SEGMENT	Frequency of Cluster	Nearest Cluster	Root-Mean-Square Standard Deviation	Maximum Distance from Cluster Seed	Distance to Nearest Cluster
1	20	7	1.145819865	193.8391724	65.34905892
2	1	7	NaN	0	439.7664571
3	1	7	NaN	0	243.9688973
4	1	7	NaN	0	701.8445065
5	1	8	NaN	0	492.7214501
6	1	7	NaN	0	225.1396671
7	377	1	0.896074446	211.6484415	65.34905892
8	1	7	NaN	0	282.6841624
9	1	1	NaN	0	385.6383892
10	1	1	NaN	0	261.3725028
11	1	7	NaN	0	339.7760149
12	14	7	1.477584942	243.2332041	76.67035371
13	2	1	1.445572694	120.1348316	201.0589315
14	1	7	NaN	0	250.405998
15	2	1	1.50829442	125.3473429	195.7574741

² อย่างไรก็ดี ขนาดของกลุ่มขึ้นอยู่กับการกระจายตัวโดยธรรมชาติของข้อมูล โดยปกติแล้วจะพิจารณาถึงความเหมาะสมในการนำไปใช้งานประกอบด้วย

ตารางที่ 4.4 ตารางค่าสถิติของกลุ่มเมื่อกำหนดจำนวนกลุ่มเท่ากับ 20 กลุ่ม

Cluster->20					
SEGMENT	Frequency of Cluster	Nearest Cluster	Root-Mean-Square Standard Deviation	Maximum Distance from Cluster Seed	Distance to Nearest Cluster
1	1	17	NaN	0	700.9831459
2	2	11	1.739937902	144.5981567	154.0973457
3	1	4	NaN	0	489.5711368
4	6	15	1.042077517	153.6716288	88.27550797
5	1	17	NaN	0	334.489114
6	2	11	1.525510758	126.7781128	206.2138629
7	9	11	0.856219246	173.9395335	58.92335225
8	1	11	NaN	0	250.527414
9	1	11	NaN	0	225.2595349
10	1	17	NaN	0	378.7389857
11	306	17	0.859424864	200.1784817	30.54972468
12	2	13	1.445572694	120.1348316	187.2351978
13	10	11	1.147476761	184.3874076	82.8432606
14	2	13	1.50197826	124.8224362	175.2220783
15	21	11	1.036747199	184.3626426	42.44116221
16	1	11	NaN	0	244.181873
17	53	11	1.079131869	191.1502565	30.54972468
18	1	15	NaN	0	269.2977339
19	1	17	NaN	0	438.3244285
20	3	13	1.448035866	155.0880511	165.4554246

ตารางที่ 4.5 ตารางค่าสถิติของกลุ่มเมื่อกำหนดจำนวนกลุ่มเท่ากับ 25 กลุ่ม

Cluster->25					
SEGMENT	Frequency of Cluster	Nearest Cluster	Root-Mean-Square Standard Deviation	Maximum Distance from Cluster Seed	Distance to Nearest Cluster
1	5	16	1.021514212	144.7296181	101.2233222
2	12	16	1.137799546	172.897735	50.25748696
3	1	18	NaN	0	492.7214501
4	12	16	1.000718152	168.2198746	60.81018438
5	1	16	NaN	0	194.7858968
6	1	4	NaN	0	371.1662993
7	1	16	NaN	0	210.6612473
8	2	24	1.50829442	125.3473429	190.4517392
9	2	16	0.94533378	78.56229925	144.8857876
10	11	16	1.00584302	177.0883858	64.58538831
11	1	16	NaN	0	200.0876921
12	1	16	NaN	0	439.8850115
13	1	16	NaN	0	244.0927471
14	1	16	NaN	0	250.4847031
15	4	16	1.191462396	151.6314183	109.4675043
16	349	2	0.867104527	208.362268	50.25748696
17	1	16	NaN	0	213.7889367
18	1	15	NaN	0	278.3977654
19	1	16	NaN	0	200.8706959
20	2	16	1.525510758	126.7781128	206.1422535
21	1	16	NaN	0	269.1756758
22	1	16	NaN	0	339.9180078
23	1	16	NaN	0	225.2568565
24	11	16	1.353367972	230.1485239	113.7461622
25	1	16	NaN	0	701.8980575

เพื่อความสะดวกในการวิเคราะห์ ผู้วิจัยจึงขอกำหนดลักษณะของกลุ่มออกเป็น 3 ประเภท คือ กลุ่มใหญ่ กลุ่มกลาง และกลุ่มเล็ก โดยวิธีการแบ่งลักษณะของกลุ่มจะยึดกลุ่มที่มีจำนวนเอกสารในกลุ่มมากที่สุดเป็นหลัก ลักษณะของกลุ่มต่าง ๆ แสดงดังนี้

- 1) กลุ่มใหญ่ หรือ กลุ่มที่มีจำนวนเอกสารในกลุ่มคิดเป็น 50 เปอร์เซ็นต์ เป็นอย่างน้อย เมื่อเทียบกับจำนวนของเอกสารในกลุ่มที่มีขนาดใหญ่กว่าเป็นลำดับถัดไป
- 2) กลุ่มเดี่ยว หรือ กลุ่มที่มีจำนวนเอกสารในกลุ่ม 1 เอกสาร
- 3) กลุ่มกลาง หรือ กลุ่มที่ไม่ได้จัดอยู่ในกลุ่มใหญ่ และกลุ่มเดี่ยว

หากจัดกลุ่มเป็น 3 ประเภทดังกล่าว จะสามารถสรุปผลที่ได้จากการกำหนดจำนวนกลุ่มเป็น 15, 20 และ 25 ได้ดังตารางที่ 4.6 และแสดงจำนวนของเอกสารที่เป็นผลลัพธ์ที่ได้จากการจัดกลุ่ม ดังตารางที่ 4.7

ตารางที่ 4.6 ตารางแสดงลักษณะของกลุ่มตามจำนวนเอกสารที่ปรากฏ

จำนวนกลุ่ม (K)	ลักษณะกลุ่มที่เป็นผลลัพธ์		
	กลุ่มใหญ่	กลุ่มกลาง	กลุ่มเดี่ยว
15	1	4	10
20	1	10	9
25	1	9	15

ตารางที่ 4.7 ตารางแสดงจำนวนเอกสารในกลุ่ม และจำนวนกลุ่มของลักษณะของกลุ่มต่าง ๆ

ค่า k	กลุ่มใหญ่		กลุ่มกลาง		กลุ่มเดี่ยว	
	เอกสารในกลุ่ม	จำนวนกลุ่ม	เอกสารในกลุ่ม	จำนวนกลุ่ม	เอกสารในกลุ่ม	จำนวนกลุ่ม
15	377	1	2	2	1	10
			14	1		
			20	1		
20	306	1	2	4	1	9
			3	1		
			6	1		

ค่า k	กลุ่มใหญ่		กลุ่มกลาง		กลุ่มเดียว	
	เอกสารในกลุ่ม	จำนวนกลุ่ม	เอกสารในกลุ่ม	จำนวนกลุ่ม	เอกสารในกลุ่ม	จำนวนกลุ่ม
			9	1		
			10	1		
			21	1		
			53	1		
25	349	1	2	3	1	15
			5	1		
			4	1		
			11	2		
			12	2		

จากตารางที่ 4.7 เมื่อพิจารณาผลลัพธ์จากจำนวนของกลุ่มประเภทต่าง ๆ สามารถแสดงการกระจายตัวของกลุ่มเอกสารในกลุ่มขนาดต่าง ๆ ได้ดังตารางที่ 4.8

ตารางที่ 4.8 ตารางแสดงร้อยละของการกระจายตัวของกลุ่มเอกสารในกลุ่มขนาดต่าง ๆ

ค่า k	ร้อยละของการกระจายตัวของกลุ่มเอกสาร		
	กลุ่มใหญ่	กลุ่มกลาง	กลุ่มเดียว
15	7%	27%	66%
20	5%	50%	45%
25	4%	36%	60%

ดังที่ได้กล่าวมาแล้วว่า ในการจัดกลุ่มเอกสารที่ดีนั้น ไม่ควรจะได้ผลลัพธ์เป็นกลุ่มที่มีขนาดใหญ่เพียงกลุ่มเดียว หรือเป็นกลุ่มเดียวจำนวนมาก ดังนั้นหากพิจารณาผลจากตารางที่ 4.6, 4.7 และ 4.8 แล้ว จะพบว่า การกำหนดจำนวนกลุ่มของเอกสาร (K) เป็น 20 จะให้ผลลัพธ์ที่มีความเหมาะสมสำหรับการจัดกลุ่มเอกสารนิตยสารไทม์ (TIME Magazine) มากกว่าการกำหนดจำนวนกลุ่มของเอกสาร (K) เป็น 15 และ 25

4.1.2 การกำหนดกรอบค่าความคล้ายคลึงด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering)

งานวิจัยนี้ศึกษาประสิทธิภาพการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน (Euclidean Distance) ซึ่งจะค้นคืนเอกสารต่อผู้ใช้ภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มเอกสาร (Clustering) เช่น ขนาดของกลุ่ม หรือ ค่ารัศมีของกลุ่มเอกสาร ที่สัมพันธ์กับข้อสอบถาม

จากการผลการทดลองจัดกลุ่มชุดเอกสารนิตยสารไทม์ (TIME Magazine) จำนวน 425 เอกสาร ด้วยการใช้เทคนิคการจัดกลุ่มข้อมูลแบบ K-means Clustering และกำหนดจำนวนกลุ่มเอกสารที่ต้องการจัดกลุ่ม 20 กลุ่ม ได้ผลการทดลองดังแสดงในตารางที่ 4.4 จากผลการทดลองพบว่าค่าสถิติ Maximum Distance from Cluster Seed จะแสดงค่าระยะทางมากที่สุดของเวกเตอร์เอกสาร ไปยังจุดศูนย์กลางของกลุ่มนั้น ๆ ซึ่งถือได้ว่าเป็นค่าที่แสดงรัศมีของกลุ่มเอกสาร ดังนั้นผู้วิจัยจึงนำค่า Maximum Distance from Cluster Seed ในแต่ละกลุ่ม มากำหนดกรอบค่าความคล้ายคลึงของการค้นคืน และผู้วิจัยจะเรียกค่า Maximum Distance from Cluster Seed ว่า "ค่ารัศมีกลุ่ม (R)"

ผู้วิจัยนำค่ารัศมีกลุ่ม (R) นี้มากำหนดกรอบค่าความคล้ายคลึงในการค้นคืน โดยมีจุดศูนย์กลางเป็นข้อสอบถาม ซึ่งเอกสารใดที่มีค่าความคล้ายคลึงกับข้อสอบถามภายในวงรัศมีกลุ่ม (R) ที่ได้กำหนดไว้ เอกสารนั้นก็จะเป็นผลลัพธ์ของการค้นคืน

4.2 ผลการทดลองประสิทธิภาพการค้นคืนเอกสาร

การศึกษานี้มีจุดประสงค์ เพื่อวัดประสิทธิภาพค่าความแม่นยำ (Precision), ค่าความระลึก (Recall) และค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของเทคนิคการค้นคืนเอกสาร 2 รูปแบบที่กำหนดในบทที่ 3 โดยสร้างเครื่องมือทดสอบการค้นคืนเอกสาร ดังนี้

เครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 คือ การค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน (Euclidean Distance) ภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering)

เครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 2 คือ การค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงมุม (Cosine Angle)

4.2.1 การวัดประสิทธิภาพการค้นคืนของเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 ด้วยการกำหนดกรอบค่าความคล้ำยคลึงที่เหมาะสม

สำหรับเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 นั้น จะนำค่ารัศมีกลุ่ม (R) ดังที่ได้กล่าวไปข้างต้น มากำหนดกรอบค่าความคล้ำยคลึงของการค้นคืน ดังนั้นผู้วิจัยได้ทดลองเปรียบเทียบประสิทธิภาพการค้นคืนด้วยการกำหนดกรอบค่าความคล้ำยคลึงที่ต่างกัน คือ กำหนดกรอบค่าความคล้ำยคลึงด้วยค่ารัศมีกลุ่ม (R) และค่ารัศมีกลุ่มหารสอง (R/2)

ซึ่งผลการทดลองค่าความแม่นยำ (Precision), ค่าความระลึก (Recall) และค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) จากการกำหนดกรอบค่าความคล้ำยคลึงด้วยค่ารัศมีกลุ่ม (R) และค่ารัศมีกลุ่มหารสอง (R/2) แสดงในภาคผนวก จ และสามารถสรุปได้ดังตารางที่ 4.9

ตารางที่ 4.9 ตารางสรุปค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 ที่กำหนดกรอบค่าความคล้ำยคลึงที่ต่างกัน

ค่าเฉลี่ยฮาร์โมนิก (Harmonic mean)	กรอบค่าความคล้ำยคลึงด้วยค่ารัศมีกลุ่ม (R)	กรอบค่าความคล้ำยคลึงด้วยค่ารัศมีกลุ่มหารสอง (R/2)
ค่าเฉลี่ย	0.0169	0.0175
ค่าเบี่ยงเบนมาตรฐาน	0.0160	0.0180

ตารางที่ 4.9 แสดงให้เห็นว่า ค่าเฉลี่ยของค่าเฉลี่ยฮาร์โมนิกที่ได้จากการทดลองเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 ด้วยการกำหนดกรอบค่าความคล้ำยคลึงค่ารัศมีกลุ่มหารสอง (R/2) มีค่ามากกว่าการกำหนดกรอบค่าความคล้ำยคลึงค่ารัศมีกลุ่ม (R)

ดังนั้น ผู้วิจัยจึงกำหนดกรอบค่าความคล้ำยคลึงในการค้นคืนของเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 ด้วยค่ารัศมีกลุ่มหารสอง (R/2) และดำเนินการทดลองเปรียบเทียบประสิทธิภาพของค่าความแม่นยำ (Precision), ค่าความระลึก (Recall) และค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) กับเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 2 ต่อไป

ผลการเปรียบเทียบประสิทธิภาพของเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 และเครื่องมือทดสอบการค้นคืนเอกสาร 2

ค่าความแม่นยำ (Precision), ค่าความระลึก (Recall) และค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ที่ได้จากการทดลองแสดงได้ดังตารางที่ 4.10

ตารางที่ 4.10 ตารางแสดงผลการทดลองค่าเฉลี่ยฮาร์โมนิก (Harmonic mean), ค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) ของเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 และรูปแบบที่ 2

ลำดับข้อ สอบถาม	ค่าเฉลี่ยฮาร์โมนิก (Harmonic mean)		ค่าความแม่นยำ (Precision)		ค่าความระลึก (Recall)	
	เครื่องมือ ทดสอบ รูปแบบที่ 1	เครื่องมือ ทดสอบ รูปแบบที่ 2	เครื่องมือ ทดสอบ รูปแบบที่ 1	เครื่องมือ ทดสอบ รูปแบบที่ 2	เครื่องมือ ทดสอบ รูปแบบที่ 1	เครื่องมือ ทดสอบ รูปแบบที่ 2
	1	0.0000	0.4118	0.0000	0.2593	0.0000
2	0.0000	0.1111	0.0000	0.0588	0.0000	1.0000
3	0.0280	0.1154	0.0143	0.0625	0.7500	0.7500
4	0.0278	0.2222	0.0142	0.1250	0.6000	1.0000
5	0.0185	0.2326	0.0095	0.1316	0.4000	1.0000
6	0.0000	0.3913	0.0000	0.2432	0.0000	1.0000
7	0.0094	0.2000	0.0047	0.1111	0.5000	1.0000
8	0.0094	0.2353	0.0047	0.1333	0.5000	1.0000
9	0.0457	0.6364	0.0237	0.5000	0.6250	0.8750
10	0.0369	0.1690	0.0190	1.0000	0.6667	0.0923
11	0.0094	0.2353	0.0047	0.1333	0.5000	1.0000
12	0.0367	0.5600	0.0190	0.3889	0.5714	1.0000
13	0.0000	0.3158	0.0000	0.1875	0.0000	1.0000
14	0.0094	0.3333	0.0047	0.2000	1.0000	1.0000
15	0.0278	0.5714	0.0142	0.4444	0.6000	0.8000
16	0.0093	0.2727	0.0047	0.1579	0.3333	1.0000
17	0.0094	0.2353	0.0047	0.1333	0.5000	1.0000
18	0.0094	0.3333	0.0047	0.2000	1.0000	1.0000
19	0.0000	0.4167	0.0000	0.2632	0.0000	1.0000
20	0.0094	0.0000	0.0047	0.0000	1.0000	0.0000

ลำดับข้อ สอบถาม	ค่าเฉลี่ยฮาร์โมนิค (Harmonic mean)		ค่าความแม่นยำ (Precision)		ค่าความระลึก (Recall)	
	เครื่องมือ ทดสอบ รูปแบบที่	เครื่องมือ ทดสอบ รูปแบบที่	เครื่องมือ ทดสอบ รูปแบบที่	เครื่องมือ ทดสอบ รูปแบบที่	เครื่องมือ ทดสอบ รูปแบบที่	เครื่องมือ ทดสอบ รูปแบบที่
	1	2	1	2	1	2
21	0.0188	0.1250	0.0095	0.0667	1.0000	1.0000
22	0.0188	0.2000	0.0095	0.1111	1.0000	1.0000
23	0.0094	0.4000	0.0047	0.2500	1.0000	1.0000
24	0.0094	0.0909	0.0047	0.0476	1.0000	1.0000
25	0.0094	0.1000	0.0047	0.0526	1.0000	1.0000
26	0.0282	0.0000	0.0142	0.0000	1.5000	0.0000
27	0.0000	0.2222	0.0000	0.1250	0.0000	1.0000
28	0.0184	0.2424	0.0094	0.1429	0.4000	0.8000
29	0.0093	0.2105	0.0047	0.1176	0.5000	1.0000
30	0.0092	0.2941	0.0047	0.1724	0.2000	1.0000
31	0.0457	0.3478	0.0236	0.2500	0.7143	0.5714
32	0.0000	0.0400	0.0000	0.0204	0.0000	1.0000
33	0.0093	1.0000	0.0047	1.0000	0.5000	1.0000
34	0.0000	0.2857	0.0000	0.1667	0.0000	1.0000
35	0.0094	0.1818	0.0047	0.1000	1.0000	1.0000
36	0.0094	0.5000	0.0047	0.3333	1.0000	1.0000
37	0.0093	0.2222	0.0047	0.1429	0.5000	0.5000
38	0.0094	0.0000	0.0047	0.0000	1.0000	0.0000
39	0.0362	0.5000	0.0189	0.3333	0.4444	1.0000
40	0.0362	0.3913	0.0189	0.2432	0.4444	1.0000
41	0.0459	0.2703	0.0236	0.1613	0.8333	0.8333
42	0.0000	0.0526	0.0000	0.0270	0.0000	1.0000
43	0.0093	0.1739	0.0047	0.0952	0.5000	1.0000

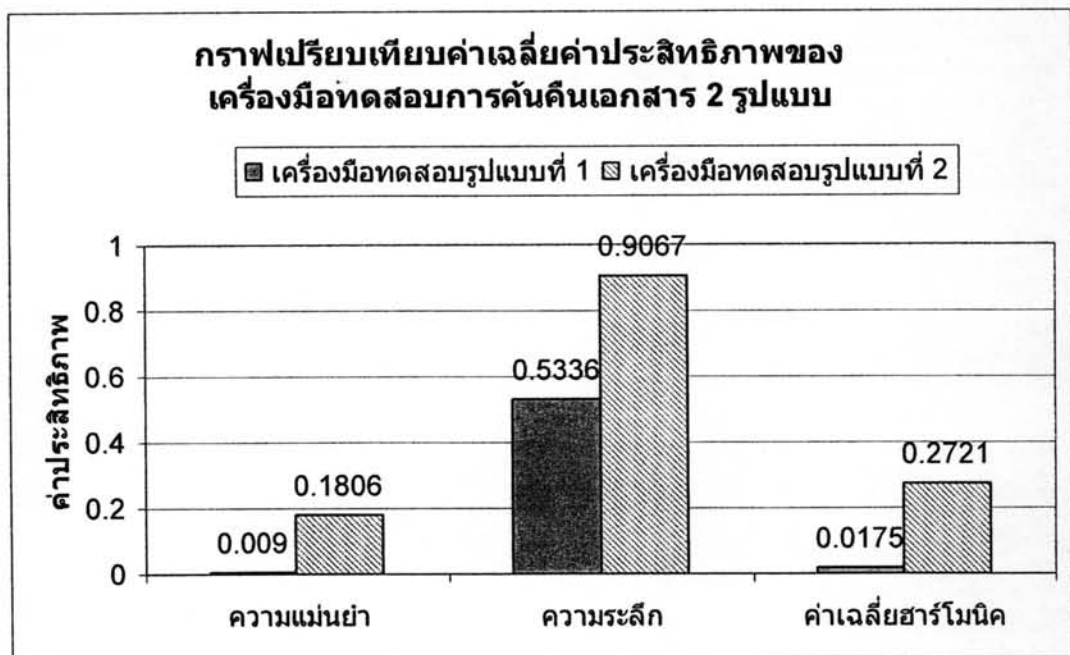
ลำดับข้อ สอบถาม	ค่าเฉลี่ยฮาร์โมนิก (Harmonic mean)		ค่าความแม่นยำ (Precision)		ค่าความระลึก (Recall)	
	เครื่องมือ ทดสอบ รูปแบบที่	เครื่องมือ ทดสอบ รูปแบบที่	เครื่องมือ ทดสอบ รูปแบบที่	เครื่องมือ ทดสอบ รูปแบบที่	เครื่องมือ ทดสอบ รูปแบบที่	เครื่องมือ ทดสอบ รูปแบบที่
	1	2	1	2	1	2
44	0.0000	0.2667	0.0000	0.1538	0.0000	1.0000
45	0.0276	0.2353	0.0142	0.1379	0.6000	0.8000
46	0.0783	0.5667	0.0425	0.4048	0.5000	0.9444
47	0.0275	0.3571	0.0142	0.2273	0.5000	0.8333
48	0.0000	0.1667	0.0000	0.0909	0.0000	1.0000
49	0.0545	0.5161	0.0283	0.3478	0.7500	1.0000
50	0.0094	0.0833	0.0047	0.0435	1.0000	1.0000
51	0.0279	0.1579	0.0142	0.0857	1.0000	1.0000
52	0.0187	0.0500	0.0094	0.0263	1.0000	0.5000
53	0.0000	0.1905	0.0000	0.1053	0.0000	1.0000
54	0.0187	0.0769	0.0094	0.0400	1.0000	1.0000
55	0.0536	0.4490	0.0283	0.2973	0.5000	0.9167
56	0.0094	0.0909	0.0047	0.0476	1.0000	1.0000
57	0.0000	0.2222	0.0000	0.1250	0.0000	1.0000
58	0.0455	0.5185	0.0236	0.3684	0.6250	0.8750
59	0.0000	0.1600	0.0000	0.0870	0.0000	1.0000
60	0.0093	0.1818	0.0047	0.1000	0.5000	1.0000
61	0.0705	0.6842	0.0377	0.5652	0.5333	0.8667
62	0.0000	0.2222	0.0000	0.1250	0.0000	1.0000
63	0.0448	0.4400	0.0236	0.2821	0.4545	1.0000
64	0.0093	0.1905	0.0047	0.1053	0.5000	1.0000
65	0.0094	0.0870	0.0047	0.0455	1.0000	1.0000
66	0.0000	0.1429	0.0000	0.0769	0.0000	1.0000

ลำดับข้อ สอบถาม	ค่าเฉลี่ยฮาร์โมนิค (Harmonic mean)		ค่าความแม่นยำ (Precision)		ค่าความระลึก (Recall)	
	เครื่องมือ ทดสอบ รูปแบบที่	เครื่องมือ ทดสอบ รูปแบบที่	เครื่องมือ ทดสอบ รูปแบบที่	เครื่องมือ ทดสอบ รูปแบบที่	เครื่องมือ ทดสอบ รูปแบบที่	เครื่องมือ ทดสอบ รูปแบบที่
	1	2	1	2	1	2
67	0.0186	0.3750	0.0094	0.2308	0.6667	1.0000
68	0.0455	0.3429	0.0236	0.2222	0.6250	0.7500
69	0.0533	0.7879	0.0283	0.6500	0.4615	1.0000
70	0.0000	0.1333	0.0000	0.0714	0.0000	1.0000
71	0.0093	0.5714	0.0047	0.5000	0.3333	0.6667
72	0.0000	0.0667	0.0000	0.0345	0.0000	1.0000
73	0.0094	0.1538	0.0047	0.0833	1.0000	1.0000
74	0.0093	0.1905	0.0047	0.1053	0.5000	1.0000
75	0.0000	0.0833	0.0000	0.0435	0.0000	1.0000
76	0.0184	0.3030	0.0094	0.1786	0.4000	1.0000
77	0.0094	0.0833	0.0047	0.0435	1.0000	1.0000
78	0.0093	0.1538	0.0047	0.0833	0.5000	1.0000
79	0.0094	0.0952	0.0047	0.0500	1.0000	1.0000
80	0.0524	0.4571	0.0283	0.4444	0.3529	0.4706
81	0.0187	0.0364	0.0094	0.0189	1.0000	0.5000
82	0.0184	0.3704	0.0094	0.2273	0.4000	1.0000
83	0.0093	0.1212	0.0047	0.0645	0.5000	1.0000

จากตารางที่ 4.10 สามารถสรุปผลการทดลองค่าเฉลี่ยฮาร์โมนิค, ค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) ของการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน (Euclidean Distance) ภายในกรอบความคล้ายคลึงด้วยคาร์คิมิกกลุ่มหารสอง ($R/2$) และวิธีการวัดความคล้ายคลึงเชิงมุม (Cosine Angle) ดังตารางที่ 4.11 และเปรียบเทียบค่าเฉลี่ยค่าประสิทธิภาพของการค้นคืนเอกสารดังรูปที่ 4.1

ตารางที่ 4.11 ตารางสรุปผลการทดลองของค่าความแม่นยำ ค่าความระลึกล และค่าเฉลี่ยฮาร์โมนิคของเครื่องมือทดสอบการค้ำคินเอกสารรูปแบบที่ 1 และรูปแบบที่ 2

		เครื่องมือทดสอบการค้ำคินเอกสาร	
		รูปแบบที่ 1	รูปแบบที่ 2
ค่าความแม่นยำ	ค่าเฉลี่ย	0.0090	0.1806
	ค่าเบี่ยงเบนมาตรฐาน	0.0095	0.1661
ค่าความระลึกล	ค่าเฉลี่ย	0.5336	0.9067
	ค่าเบี่ยงเบนมาตรฐาน	0.3752	0.2198
ค่าเฉลี่ยฮาร์โมนิค	ค่าเฉลี่ย	0.0175	0.2721
	ค่าเบี่ยงเบนมาตรฐาน	0.0180	0.1917



รูปที่ 4.1 รูปแสดงกราฟเปรียบเทียบค่าความแม่นยำ ค่าความระลึกล และค่าเฉลี่ยฮาร์โมนิค ระหว่างเครื่องมือทดสอบการค้ำคินเอกสารรูปแบบที่ 1 และรูปแบบที่ 2

ตารางที่ 4.11 แสดงให้เห็นว่า ค่าเฉลี่ยค่าความแม่นยำของเครื่องมือทดสอบการค้ำคินรูปแบบที่ 1 น้อยกว่าเครื่องมือทดสอบการค้ำคินรูปแบบที่ 2 เท่ากับ 0.1716 ค่าเฉลี่ยค่าความระลึกลของเครื่องมือทดสอบการค้ำคินรูปแบบที่ 1 น้อยกว่าเครื่องมือทดสอบการค้ำคินรูปแบบที่ 2

เท่ากับ 0.3731 และค่าเฉลี่ยของค่าเฉลี่ยฮาร์โมนิกของเครื่องมือทดสอบการค้นคืนรูปแบบที่ 1 น้อยกว่าเครื่องมือทดสอบการค้นคืนรูปแบบที่ 2 เท่ากับ 0.2546 ดังนั้น ค่าเฉลี่ยของค่าประสิทธิภาพทั้ง 3 ค่า ของเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 มีค่าน้อยกว่าค่าที่ได้จากเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 2 สามารถสรุปได้ว่าวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน (Euclidean Distance) ภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering) นั้นไม่สามารถเพิ่มประสิทธิภาพการค้นคืนเอกสารของเทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงมุม (Cosine Angle) ได้

ผู้วิจัยพบว่า ปัญหาที่อาจทำให้ค่าเฉลี่ยของค่าประสิทธิภาพของเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 ไม่ดี อาจเกิดจากการที่ค่าที่ปรากฏในเอกสารนิตยสารใหม่ มีความหลากหลาย ซึ่งส่งผลให้เวกเตอร์เอกสารมีความยาวถึง 13,813 หน่วย ในขณะที่แต่ละเอกสารมีค่าปรากฏอยู่เฉลี่ยเท่ากับ 218 คำ และข้อสอบถามมีค่าปรากฏอยู่เฉลี่ย 9 คำ ดังนั้นในการคำนวณค่าความคล้ายคลึงเชิงระยะห่างยูคลิเดียน ทำให้เอกสารที่มีค่าปรากฏอยู่น้อยมีความคล้ายคลึงกับข้อสอบถามมากกว่า ซึ่งผิดวัตถุประสงค์ของการค้นคืนเอกสาร

ดังนั้นผู้วิจัยจึงต้องการทดลองว่า ค่าเฉลี่ยของค่าประสิทธิภาพเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 จะสามารถเพิ่มได้หรือไม่ โดยกำหนดให้เครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 ให้ค้นหาเอกสารจากการคำนวณค่าความคล้ายคลึงเชิงระยะห่างยูคลิเดียนระหว่างเฉพาะเอกสารที่มีค่าในข้อสอบถามปรากฏเท่านั้น

4.2.2 การวัดประสิทธิภาพการค้นคืนของเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 เฉพาะเอกสารที่มีค่าในข้อสอบถามปรากฏเท่านั้น

จากปัญหาที่เกิดขึ้นดังกล่าวข้างต้นนั้น ผู้วิจัยต้องการศึกษาประสิทธิภาพการค้นคืนของเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 เมื่อกำหนดให้พิจารณาค้นคืนเฉพาะเอกสารที่มีค่าในข้อสอบถามนั้น ๆ ปรากฏเท่านั้น เอกสารใดที่ไม่มีค่าในข้อสอบถามปรากฏจะไม่ถูกนำมาพิจารณาค้นคืน และทดลองเปรียบเทียบประสิทธิภาพการค้นคืนด้วยการกำหนดกรอบค่าความคล้ายคลึงด้วยค่ารัศมีกลุ่ม (R) และค่ารัศมีกลุ่มหารสอง (R/2)

ซึ่งผลการทดลองค่าความแม่นยำ (Precision), ค่าความระลึก (Recall) และค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) จากการพิจารณาค้นคืนเฉพาะเอกสารที่มีค่าในข้อสอบถามปรากฏเท่านั้น โดยกำหนดกรอบค่าความคล้ายคลึงที่ต่างกัน แสดงในภาคผนวก ๑ และสามารถสรุปค่าประสิทธิภาพได้ดังตารางที่ 4.12

ตารางที่ 4.12 ตารางสรุปผลค่าเฉลี่ยฮาร์โมนิก ของเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 ภายใต้กรอบค่าความคล้ายคลึงที่ต่างกัน และด้วยการพิจารณาค้นคืนเฉพาะเอกสารที่มีค่าในข้อสอบถามปรากฏเท่านั้น

ค่าเฉลี่ยฮาร์โมนิก (Harmonic mean)	การค้นคืนเอกสารที่มีค่าในข้อสอบถามปรากฏ ด้วยค่ารัศมีกลุ่ม (R)	การค้นคืนเอกสารที่มีค่าในข้อสอบถามปรากฏ ด้วยค่ารัศมีกลุ่มหารสอง (R/2)
ค่าเฉลี่ย	0.0315	0.0377
ค่าเบี่ยงเบนมาตรฐาน	0.0280	0.0417

ตารางที่ 4.12 แสดงให้เห็นว่า ค่าเฉลี่ยของค่าเฉลี่ยฮาร์โมนิกที่ได้จากการทดลองเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 ด้วยการค้นคืนเอกสารที่มีค่าในข้อสอบถามปรากฏ และกำหนดกรอบค่าความคล้ายคลึงด้วยค่ารัศมีกลุ่มหารสอง (R/2) มีค่ามากกว่าการกำหนดกรอบค่าความคล้ายคลึงด้วยค่ารัศมีกลุ่ม (R) ดังนั้น ผู้วิจัยจึงกำหนดกรอบค่าความคล้ายคลึงในการค้นคืนของเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 ด้วยการค้นคืนเอกสารที่มีค่าในข้อสอบถามปรากฏ ด้วยค่ารัศมีกลุ่มหารสอง (R/2) และขอเรียกวิธีการนี้ว่า "เครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1* "

ผู้วิจัยดำเนินการทดลองเปรียบเทียบประสิทธิภาพของค่าความแม่นยำ (Precision), ค่าความระลึก (Recall) และค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 กับเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1* เพื่อทดลองและวิเคราะห์ว่า เครื่องมือทดสอบการค้นคืนเอกสารรูปแบบใดสามารถให้ค่าประสิทธิภาพการค้นคืนเอกสารที่ดีกว่ากัน และเหมาะสำหรับนำไปต่อบจุดประสงค์ของงานวิจัยนี้

ผลการเปรียบเทียบประสิทธิภาพของเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 และเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1*

ค่าความแม่นยำ (Precision), ค่าความระลึก (Recall) และค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ที่ได้จากการทดลองแสดงได้ดังตารางที่ 4.13

ตารางที่ 4.13 ตารางแสดงผลการทดลองค่าเฉลี่ยฮาร์โมนิก (Harmonic mean), ค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) ของเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 และรูปแบบที่ 1*

ลำดับข้อ สอบถาม	ค่าเฉลี่ยฮาร์โมนิก (Harmonic mean)		ค่าความแม่นยำ (Precision)		ค่าความระลึก (Recall)	
	เครื่องมือ ทดสอบ รูปแบบที่ 1	เครื่องมือ ทดสอบ รูปแบบที่ 1*	เครื่องมือ ทดสอบ รูปแบบที่ 1	เครื่องมือ ทดสอบ รูปแบบที่ 1*	เครื่องมือ ทดสอบ รูปแบบที่ 1	เครื่องมือ ทดสอบ รูปแบบที่ 1*
1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0280	0.0429	0.0143	0.0221	0.7500	0.7500
4	0.0278	0.0789	0.0142	0.0423	0.6000	0.6000
5	0.0185	0.0755	0.0095	0.0417	0.4000	0.4000
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
7	0.0094	0.0149	0.0047	0.0076	0.5000	0.5000
8	0.0094	0.0256	0.0047	0.0132	0.5000	0.5000
9	0.0457	0.1887	0.0237	0.1111	0.6250	0.6250
10	0.0369	0.1231	0.0190	0.0678	0.6667	0.6667
11	0.0094	0.0143	0.0047	0.0072	0.5000	0.5000
12	0.0367	0.1111	0.0190	0.0615	0.5714	0.5714
13	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
14	0.0094	0.0135	0.0047	0.0068	1.0000	1.0000
15	0.0278	0.0517	0.0142	0.0270	0.6000	0.6000
16	0.0093	0.0136	0.0047	0.0069	0.3333	0.3333
17	0.0094	0.0400	0.0047	0.0208	0.5000	0.5000
18	0.0094	0.0127	0.0047	0.0064	1.0000	1.0000
19	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
20	0.0094	0.0140	0.0047	0.0070	1.0000	1.0000

ลำดับข้อ สอบถาม	ค่าเฉลี่ยฮาร์โมนิก (Harmonic mean)		ค่าความแม่นยำ (Precision)		ค่าความระลึก (Recall)	
	เครื่องมือ ทดสอบ รูปแบบที่ 1	เครื่องมือ ทดสอบ รูปแบบที่ 1*	เครื่องมือ ทดสอบ รูปแบบที่ 1	เครื่องมือ ทดสอบ รูปแบบที่ 1*	เครื่องมือ ทดสอบ รูปแบบที่ 1	เครื่องมือ ทดสอบ รูปแบบที่ 1*
	21	0.0188	0.0305	0.0095	0.0155	1.0000
22	0.0188	0.0506	0.0095	0.0260	1.0000	1.0000
23	0.0094	0.0154	0.0047	0.0078	1.0000	1.0000
24	0.0094	0.0175	0.0047	0.0088	1.0000	1.0000
25	0.0094	0.0556	0.0047	0.0286	1.0000	1.0000
26	0.0282	0.0351	0.0142	0.0179	1.5000	1.0000
27	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
28	0.0184	0.0305	0.0094	0.0159	0.4000	0.4000
29	0.0093	0.0123	0.0047	0.0062	0.5000	0.5000
30	0.0092	0.0135	0.0047	0.0070	0.2000	0.2000
31	0.0457	0.0794	0.0236	0.0420	0.7143	0.7143
32	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
33	0.0093	0.0172	0.0047	0.0088	0.5000	0.5000
34	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
35	0.0094	0.0238	0.0047	0.0120	1.0000	1.0000
36	0.0094	0.0213	0.0047	0.0108	1.0000	1.0000
37	0.0093	0.0172	0.0047	0.0088	0.5000	0.5000
38	0.0094	0.0227	0.0047	0.0115	1.0000	1.0000
39	0.0362	0.0556	0.0189	0.0296	0.4444	0.4444
40	0.0362	0.0523	0.0189	0.0278	0.4444	0.4444
41	0.0459	0.0637	0.0236	0.0331	0.8333	0.8333
42	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
43	0.0093	0.0357	0.0047	0.0185	0.5000	0.5000

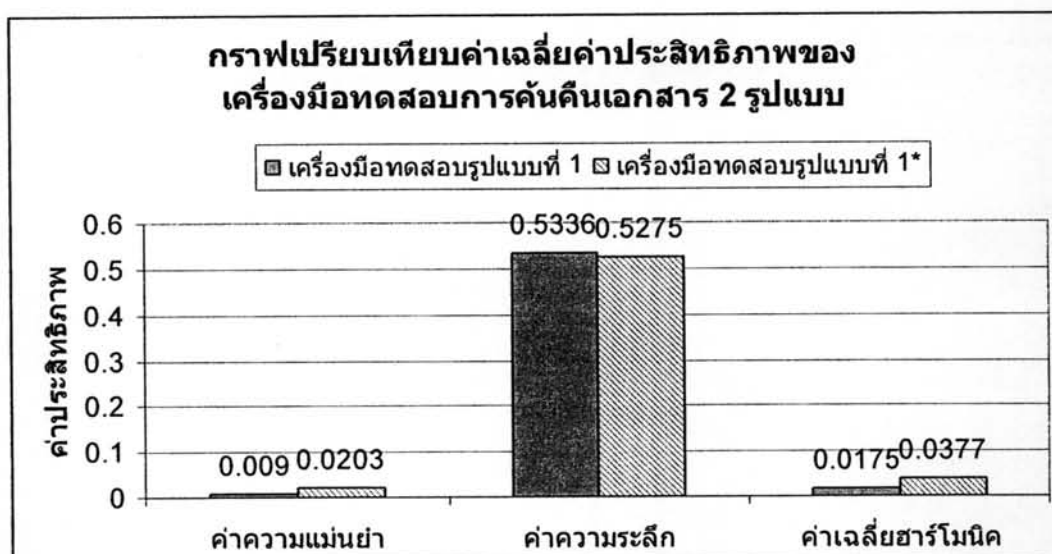
ลำดับข้อ สอบถาม	ค่าเฉลี่ยฮาร์โมนิค (Harmonic mean)		ค่าความแม่นยำ (Precision)		ค่าความระลึก (Recall)	
	เครื่องมือ ทดสอบ รูปแบบที่ 1	เครื่องมือ ทดสอบ รูปแบบที่ 1*	เครื่องมือ ทดสอบ รูปแบบที่ 1	เครื่องมือ ทดสอบ รูปแบบที่ 1*	เครื่องมือ ทดสอบ รูปแบบที่ 1	เครื่องมือ ทดสอบ รูปแบบที่ 1*
	67	0.0186	0.1333	0.0094	0.0741	0.6667
68	0.0455	0.0943	0.0236	0.0510	0.6250	0.6250
69	0.0533	0.0741	0.0283	0.0403	0.4615	0.4615
70	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
71	0.0093	0.0171	0.0047	0.0088	0.3333	0.3333
72	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
73	0.0094	0.0160	0.0047	0.0081	1.0000	1.0000
74	0.0093	0.0313	0.0047	0.0161	0.5000	0.5000
75	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
76	0.0184	0.0247	0.0094	0.0127	0.4000	0.4000
77	0.0094	0.0260	0.0047	0.0132	1.0000	1.0000
78	0.0093	0.0175	0.0047	0.0089	0.5000	0.5000
79	0.0094	0.0130	0.0047	0.0065	1.0000	1.0000
80	0.0524	0.0732	0.0283	0.0408	0.3529	0.3529
81	0.0187	0.0274	0.0094	0.0139	1.0000	1.0000
82	0.0184	0.0272	0.0094	0.0141	0.4000	0.4000
83	0.0093	0.0131	0.0047	0.0066	0.5000	0.5000

จากตารางที่ 4.13 สามารถสรุปผลการทดลองค่าเฉลี่ยฮาร์โมนิค ค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) ของการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน (Euclidean Distance) ภายในกรอบความคล้ายคลึงด้วยคาร์คมีกลุ่มหารสอง ($R/2$) และการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน ด้วยการพิจารณาค้นคืนเฉพาะเอกสารที่มีค่าใน

ข้อสอบถามปรากฏ ภายในกรอบความคล้อยคลึงด้วยค่ารัศมีกลุ่มหารสอง (R/2) ดังตารางที่ 4.14 และเปรียบเทียบค่าเฉลี่ยค่าประสิทธิภาพของการคั่นคืนเอกสารดังรูปที่ 4.2

ตารางที่ 4.14 ตารางสรุปผลการทดลองของค่าความแม่นยำ ค่าความระลึกลับ และค่าเฉลี่ยฮาร์โมนิคของเครื่องมือทดสอบการคั่นคืนเอกสารรูปแบบที่ 1 และรูปแบบที่ 1*

		เครื่องมือทดสอบการคั่นคืนเอกสาร	
		รูปแบบที่ 1	รูปแบบที่ 1*
ค่าความแม่นยำ	ค่าเฉลี่ย	0.0090	0.0203
	ค่าเบี่ยงเบนมาตรฐาน	0.0095	0.0235
ค่าความระลึกลับ	ค่าเฉลี่ย	0.5336	0.5275
	ค่าเบี่ยงเบนมาตรฐาน	0.3752	0.3633
ค่าเฉลี่ยฮาร์โมนิค	ค่าเฉลี่ย	0.0175	0.0377
	ค่าเบี่ยงเบนมาตรฐาน	0.0180	0.0417



รูปที่ 4.2 รูปแสดงกราฟเปรียบเทียบ ค่าความแม่นยำ ค่าความระลึกลับ และค่าเฉลี่ยฮาร์โมนิค ระหว่างเครื่องมือทดสอบการคั่นคืนเอกสารรูปแบบที่ 1 และรูปแบบที่ 1*

ตารางที่ 4.14 แสดงให้เห็นว่า ค่าเฉลี่ยค่าความแม่นยำของเครื่องมือทดสอบการคั่นคืนรูปแบบที่ 1 น้อยกว่าเครื่องมือทดสอบการคั่นคืนรูปแบบที่ 1* เท่ากับ 0.0113 ค่าเฉลี่ยค่าความ

ระลึกรของเครื่องมือทดสอบการค้นคืนรูปแบบที่ 1 มากกว่าเครื่องมือทดสอบการค้นคืนรูปแบบที่ 1* เท่ากับ 0.0061 และค่าเฉลี่ยของค่าเฉลี่ยฮาร์โมนิกของเครื่องมือทดสอบการค้นคืนรูปแบบที่ 1 น้อยกว่าเครื่องมือทดสอบการค้นคืนรูปแบบที่ 1* เท่ากับ 0.0202 เพื่อทดสอบว่า เครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1* มีประสิทธิภาพแตกต่างจากเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 อย่างมีนัยสำคัญ ผู้วิจัยได้ดำเนินการทดลองทางสถิติ (รายละเอียดดังแสดงในภาคผนวก ข) แล้วพบว่า เครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1* มีค่าประสิทธิภาพความแม่นยำ และค่าเฉลี่ยฮาร์โมนิกมากกว่าเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1 อย่างมีนัยสำคัญทางสถิติ ในส่วนของค่าความระลึกร ผลการทดสอบทางสถิติ แสดงให้เห็นว่า เครื่องมือทดสอบการค้นคืนเอกสารทั้ง 2 รูปแบบ มีประสิทธิภาพไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

ดังนั้น ผู้วิจัยจึงเลือกใช้การค้นคืนเอกสารด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิดเดียน ด้วยวิธีพิจารณาค้นคืนเฉพาะเอกสารที่มีค่าในข้อสอบถามปรากฏ ภายใต้กรอบค่าความคล้ายคลึงในการค้นคืนด้วยคาร์ตมีกลุ่มหารสอง (R/2) (เครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1*) ในการเปรียบเทียบประสิทธิภาพค่าความแม่นยำ (Precision), ค่าความระลึกร (Recall) และค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) กับเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 2 เพื่อตอบวัตถุประสงค์ของงานวิจัยต่อไป

4.2.3 การเปรียบเทียบประสิทธิภาพของเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1* และเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 2

ผลการเปรียบเทียบค่าความแม่นยำ (Precision), ค่าความระลึกร (Recall) และค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1* และเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 2 ที่ได้จากการทดลองแสดงในตารางที่ 4.15

ตารางที่ 4.15 ตารางแสดงผลการทดลองค่าเฉลี่ยฮาร์โมนิก (Harmonic mean), ค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) ของเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1* และรูปแบบที่ 2

ลำดับข้อ สอบถาม	ค่าเฉลี่ยฮาร์โมนิก (Harmonic mean)		ค่าความแม่นยำ (Precision)		ค่าความระลึก (Recall)	
	เครื่องมือ ทดสอบ รูปแบบที่ 1*	เครื่องมือ ทดสอบ รูปแบบที่ 2	เครื่องมือ ทดสอบ รูปแบบที่ 1*	เครื่องมือ ทดสอบ รูปแบบที่ 2	เครื่องมือ ทดสอบ รูปแบบที่ 1*	เครื่องมือ ทดสอบ รูปแบบที่ 2
	1	0.0000	0.4118	0.0000	0.2593	0.0000
2	0.0000	0.1111	0.0000	0.0588	0.0000	1.0000
3	0.0429	0.1154	0.0221	0.0625	0.7500	0.7500
4	0.0789	0.2222	0.0423	0.1250	0.6000	1.0000
5	0.0755	0.2326	0.0417	0.1316	0.4000	1.0000
6	0.0000	0.3913	0.0000	0.2432	0.0000	1.0000
7	0.0149	0.2000	0.0076	0.1111	0.5000	1.0000
8	0.0256	0.2353	0.0132	0.1333	0.5000	1.0000
9	0.1887	0.6364	0.1111	0.5000	0.6250	0.8750
10	0.1231	0.5217	0.0678	0.3529	0.6667	1.0000
11	0.0143	0.2353	0.0072	0.1333	0.5000	1.0000
12	0.1111	0.5600	0.0615	0.3889	0.5714	1.0000
13	0.0000	0.3158	0.0000	0.1875	0.0000	1.0000
14	0.0135	0.3333	0.0068	0.2000	1.0000	1.0000
15	0.0517	0.5714	0.0270	0.4444	0.6000	0.8000
16	0.0136	0.2727	0.0069	0.1579	0.3333	1.0000
17	0.0400	0.2353	0.0208	0.1333	0.5000	1.0000
18	0.0127	0.3333	0.0064	0.2000	1.0000	1.0000
19	0.0000	0.4167	0.0000	0.2632	0.0000	1.0000
20	0.0140	0.0000	0.0070	0.0000	1.0000	0.0000

ลำดับข้อ สอบถาม	ค่าเฉลี่ยฮาร์โมนิก (Harmonic mean)		ค่าความแม่นยำ (Precision)		ค่าความระลึก (Recall)	
	เครื่องมือ ทดสอบ รูปแบบที่ 1*	เครื่องมือ ทดสอบ รูปแบบที่ 2	เครื่องมือ ทดสอบ รูปแบบที่ 1*	เครื่องมือ ทดสอบ รูปแบบที่ 2	เครื่องมือ ทดสอบ รูปแบบที่ 1*	เครื่องมือ ทดสอบ รูปแบบที่ 2
	21	0.0305	0.1250	0.0155	0.0667	1.0000
22	0.0506	0.2000	0.0260	0.1111	1.0000	1.0000
23	0.0154	0.4000	0.0078	0.2500	1.0000	1.0000
24	0.0175	0.0909	0.0088	0.0476	1.0000	1.0000
25	0.0556	0.1000	0.0286	0.0526	1.0000	1.0000
26	0.0351	0.0000	0.0179	0.0000	1.0000	0.0000
27	0.0000	0.2222	0.0000	0.1250	0.0000	1.0000
28	0.0305	0.2424	0.0159	0.1429	0.4000	0.8000
29	0.0123	0.2105	0.0062	0.1176	0.5000	1.0000
30	0.0135	0.2941	0.0070	0.1724	0.2000	1.0000
31	0.0794	0.3478	0.0420	0.2500	0.7143	0.5714
32	0.0000	0.0400	0.0000	0.0204	0.0000	1.0000
33	0.0172	1.0000	0.0088	1.0000	0.5000	1.0000
34	0.0000	0.2857	0.0000	0.1667	0.0000	1.0000
35	0.0238	0.1818	0.0120	0.1000	1.0000	1.0000
36	0.0213	0.5000	0.0108	0.3333	1.0000	1.0000
37	0.0172	0.2222	0.0088	0.1429	0.5000	0.5000
38	0.0227	0.0000	0.0115	0.0000	1.0000	0.0000
39	0.0556	0.5000	0.0296	0.3333	0.4444	1.0000
40	0.0523	0.3913	0.0278	0.2432	0.4444	1.0000
41	0.0637	0.2703	0.0331	0.1613	0.8333	0.8333
42	0.0000	0.0526	0.0000	0.0270	0.0000	1.0000
43	0.0357	0.1739	0.0185	0.0952	0.5000	1.0000

ลำดับข้อ สอบถาม	ค่าเฉลี่ยฮาร์โมนิก (Harmonic mean)		ค่าความแม่นยำ (Precision)		ค่าความระลึก (Recall)	
	เครื่องมือ ทดสอบ รูปแบบที่ 1*	เครื่องมือ ทดสอบ รูปแบบที่ 2	เครื่องมือ ทดสอบ รูปแบบที่ 1*	เครื่องมือ ทดสอบ รูปแบบที่ 2	เครื่องมือ ทดสอบ รูปแบบที่ 1*	เครื่องมือ ทดสอบ รูปแบบที่ 2
	44	0.0000	0.2667	0.0000	0.1538	0.0000
45	0.0526	0.2353	0.0275	0.1379	0.6000	0.8000
46	0.1268	0.5667	0.0726	0.4048	0.5000	0.9444
47	0.0923	0.3571	0.0508	0.2273	0.5000	0.8333
48	0.0000	0.1667	0.0000	0.0909	0.0000	1.0000
49	0.1846	0.5161	0.1053	0.3478	0.7500	1.0000
50	0.0130	0.0833	0.0065	0.0435	1.0000	1.0000
51	0.0822	0.1579	0.0429	0.0857	1.0000	1.0000
52	0.0244	0.0500	0.0123	0.0263	1.0000	0.5000
53	0.0000	0.1905	0.0000	0.1053	0.0000	1.0000
54	0.0270	0.0769	0.0137	0.0400	1.0000	1.0000
55	0.0800	0.4490	0.0435	0.2973	0.5000	0.9167
56	0.0154	0.0909	0.0078	0.0476	1.0000	1.0000
57	0.0000	0.2222	0.0000	0.1250	0.0000	1.0000
58	0.0862	0.5185	0.0463	0.3684	0.6250	0.8750
59	0.0000	0.1600	0.0000	0.0870	0.0000	1.0000
60	0.0541	0.1818	0.0286	0.1000	0.5000	1.0000
61	0.1168	0.6842	0.0656	0.5652	0.5333	0.8667
62	0.0000	0.2222	0.0000	0.1250	0.0000	1.0000
63	0.0526	0.4400	0.0279	0.2821	0.4545	1.0000
64	0.0097	0.1905	0.0049	0.1053	0.5000	1.0000
65	0.0556	0.0870	0.0286	0.0455	1.0000	1.0000
66	0.0000	0.1429	0.0000	0.0769	0.0000	1.0000

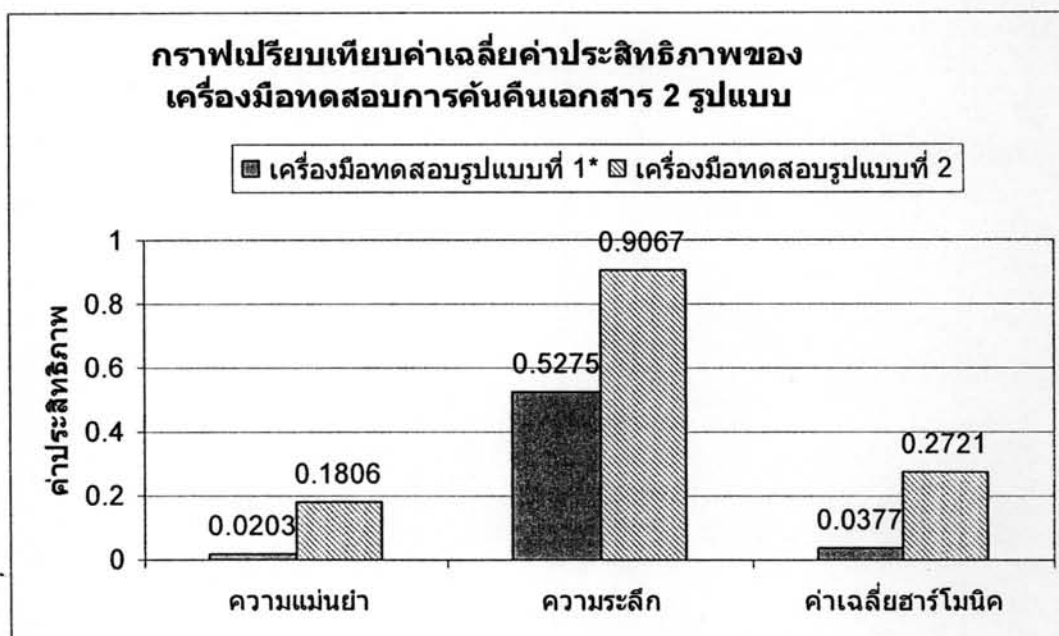
ลำดับข้อ สอบถาม	ค่าเฉลี่ยฮาร์โมนิค (Harmonic mean)		ค่าความแม่นยำ (Precision)		ค่าความระลึก (Recall)	
	เครื่องมือ ทดสอบ รูปแบบที่ 1*	เครื่องมือ ทดสอบ รูปแบบที่ 2	เครื่องมือ ทดสอบ รูปแบบที่ 1*	เครื่องมือ ทดสอบ รูปแบบที่ 2	เครื่องมือ ทดสอบ รูปแบบที่ 1*	เครื่องมือ ทดสอบ รูปแบบที่ 2
	67	0.1333	0.3750	0.0741	0.2308	0.6667
68	0.0943	0.3429	0.0510	0.2222	0.6250	0.7500
69	0.0741	0.7879	0.0403	0.6500	0.4615	1.0000
70	0.0000	0.1333	0.0000	0.0714	0.0000	1.0000
71	0.0171	0.5714	0.0088	0.5000	0.3333	0.6667
72	0.0000	0.0667	0.0000	0.0345	0.0000	1.0000
73	0.0160	0.1538	0.0081	0.0833	1.0000	1.0000
74	0.0313	0.1905	0.0161	0.1053	0.5000	1.0000
75	0.0000	0.0833	0.0000	0.0435	0.0000	1.0000
76	0.0247	0.3030	0.0127	0.1786	0.4000	1.0000
77	0.0260	0.0833	0.0132	0.0435	1.0000	1.0000
78	0.0175	0.1538	0.0089	0.0833	0.5000	1.0000
79	0.0130	0.0952	0.0065	0.0500	1.0000	1.0000
80	0.0732	0.4571	0.0408	0.4444	0.3529	0.4706
81	0.0274	0.0364	0.0139	0.0189	1.0000	0.5000
82	0.0272	0.3704	0.0141	0.2273	0.4000	1.0000
83	0.0131	0.1212	0.0066	0.0645	0.5000	1.0000

จากตารางที่ 4.15 สามารถสรุปผลการทดลองค่าเฉลี่ยฮาร์โมนิค, ค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) ของการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน (Euclidean Distance) ภายในกรอบความคล้ายคลึงด้วยค่ารัศมีกลุ่มหารสอง ($R/2$) ด้วยการพิจารณาค้นคืนเฉพาะเอกสารที่มีค่าในข้อ

สอบถามปรากฏ และวิธีการวัดความคล้ายคลึงเชิงมุม (Cosine Angle) ดังตารางที่ 4.16 และเปรียบเทียบค่าเฉลี่ยค่าประสิทธิภาพของการคั่นคั้นเอกสารดังรูปที่ 4.3

ตารางที่ 4.16 ตารางสรุปผลการทดลองของค่าความแม่นยำ ค่าความระลึกลับ และค่าเฉลี่ยฮาร์โมนิกของเครื่องมือทดสอบการคั่นคั้นเอกสารรูปแบบที่ 1* และรูปแบบที่ 2

		เครื่องมือทดสอบการคั่นคั้นเอกสาร	
		รูปแบบที่ 1*	รูปแบบที่ 2
ค่าความแม่นยำ	ค่าเฉลี่ย	0.0203	0.1806
	ค่าเบี่ยงเบนมาตรฐาน	0.0235	0.1661
ค่าความระลึกลับ	ค่าเฉลี่ย	0.5275	0.9067
	ค่าเบี่ยงเบนมาตรฐาน	0.3633	0.2198
ค่าเฉลี่ยฮาร์โมนิก	ค่าเฉลี่ย	0.0377	0.2721
	ค่าเบี่ยงเบนมาตรฐาน	0.0417	0.1917



รูปที่ 4.3 รูปแสดงกราฟเปรียบเทียบ ค่าความแม่นยำ ค่าความระลึกลับ และค่าเฉลี่ยฮาร์โมนิก ระหว่างเครื่องมือทดสอบการคั่นคั้นเอกสารรูปแบบที่ 1* และรูปแบบที่ 2

ตารางที่ 4.16 แสดงให้เห็นว่า ค่าเฉลี่ยค่าความแม่นยำของเครื่องมือทดสอบการค้นคืนรูปแบบที่ 1* น้อยกว่าเครื่องมือทดสอบการค้นคืนรูปแบบที่ 2 เท่ากับ 0.1603 ค่าเฉลี่ยค่าความระลึกของเครื่องมือทดสอบการค้นคืนรูปแบบที่ 1* น้อยกว่าเครื่องมือทดสอบการค้นคืนรูปแบบที่ 2 เท่ากับ 0.3792 และค่าเฉลี่ยของค่าเฉลี่ยฮาร์โมนิกของเครื่องมือทดสอบการค้นคืนรูปแบบที่ 1* น้อยกว่าเครื่องมือทดสอบการค้นคืนรูปแบบที่ 2 เท่ากับ 0.2344 ดังนั้น ค่าเฉลี่ยของค่าประสิทธิภาพทั้ง 3 ค่า ของเครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 1* โดยเทคนิควิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน (Euclidean Distance) ภายในกรอบความคล้ายคลึงค่ารัศมีกลุ่มหารสอง ($R/2$) ด้วยการพิจารณาค้นคืนเฉพาะเอกสารที่มีค่าในข้อสอบถามปรากฏ มีค่าเฉลี่ยของค่าประสิทธิภาพทั้ง 3 ค่าน้อยกว่า เครื่องมือทดสอบการค้นคืนเอกสารรูปแบบที่ 2

เพื่อสรุปว่า ประสิทธิภาพของเครื่องมือทดสอบการค้นคืนเอกสารทั้ง 2 แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ผู้วิจัยจึงได้แสดงผลการวิเคราะห์ทางสถิติในหัวข้อถัดไป

4.3 การวิเคราะห์ผลการทดลองทางสถิติ

งานวิจัยนี้ต้องการเปรียบเทียบประสิทธิภาพของการค้นคืนเอกสาร (Document Retrieval) ด้วยวิธีการวัดความคล้ายคลึงระหว่างเอกสารและข้อสอบถามที่ต่างกัน การตรวจสอบเงื่อนไขพื้นฐานขั้นแรกต้องตรวจสอบการแจกแจงของค่าประสิทธิภาพที่ได้มาว่า มีการแจกแจงปกติหรือไม่ เพื่อเลือกทางเลือกในการทดสอบสมมติฐานว่าจะใช้วิธีการทดสอบสมมติฐานแบบอิงพารามิเตอร์ (Parametric Test) หรือแบบไม่อิงกับพารามิเตอร์ (Non Parametric Test) (กัลยา วานิชย์บัญชา, 2546) ถ้าผลการทดสอบพบว่าค่าประสิทธิภาพมีการแจกแจงแบบปกติ จึงจะใช้สถิติทดสอบแบบที่ (t-test) เพื่อทดสอบสมมติฐานที่ตั้งไว้ แต่ถ้าผลทดสอบพบว่ามีการแจกแจงประชากรไม่เป็นแบบปกติ ผู้วิจัยต้องใช้วิธีการทดสอบสมมติฐานแบบไม่ใช้พารามิเตอร์ (Non Parametric Test)

ในงานวิจัยนี้ผู้วิจัยสนใจตัวแปร คือค่าประสิทธิภาพของเทคนิคการค้นคืนเอกสารทั้ง 2 รูปแบบ ที่วัดด้วยค่าความแม่นยำ (Precision), ค่าความระลึก (Recall) และค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของการค้นคืนเอกสารที่ทดสอบโดยใช้ข้อสอบถามจำนวน 83 ข้อสอบถาม ดังนั้นผู้วิจัยจึงดำเนินการทดสอบทางสถิติ ดังต่อไปนี้

4.3.1 การวิเคราะห์ข้อมูลค่าความแม่นยำ (Precision)

การวิเคราะห์การแจกแจงข้อมูล

ผู้วิจัยจะตรวจสอบว่าค่าความแม่นยำ (Precision) ของการค้นคืนเอกสารมีการแจกแจงแบบปกติหรือไม่ จากค่าความแม่นยำ (Precision) 83 ค่าที่ได้ในการค้นคืนเอกสารแต่ละรูปแบบ โดยตั้งสมมติฐานของการทดสอบสำหรับทดสอบค่าตัวแปรค่าความแม่นยำ (Precision) ของการค้นคืนเอกสารแต่ละรูปแบบ ดังนี้

1) ทดสอบการแจกแจงค่าความแม่นยำของการค้นคืนเอกสารรูปแบบที่ 1*

H_0 : ข้อมูลค่าความแม่นยำของการค้นคืนเอกสารรูปแบบที่ 1* มีการแจกแจงแบบปกติ

H_1 : ข้อมูลค่าความแม่นยำของการค้นคืนเอกสารรูปแบบที่ 1* ไม่ได้มีการแจกแจงแบบปกติ

2) ทดสอบการแจกแจงค่าความแม่นยำของการค้นคืนเอกสารรูปแบบที่ 2

H_0 : ข้อมูลค่าความแม่นยำของการค้นคืนเอกสารรูปแบบที่ 2 มีการแจกแจงแบบปกติ

H_1 : ข้อมูลค่าความแม่นยำของการค้นคืนเอกสารรูปแบบที่ 2 ไม่ได้มีการแจกแจงแบบปกติ

การตรวจสอบการแจกแจงของข้อมูลเชิงปริมาณมีสถิติทดสอบที่ใช้คือ Kolmogorov-Smirnev เมื่อขนาดตัวอย่างมากกว่า 50 หน่วย และของ Shapiro-Wilk เมื่อขนาดตัวอย่างอย่างน้อยน้อยกว่า 50 หน่วย (กัลยา วานิชย์บัญชา, 2546) สำหรับในงานวิจัยนี้ตัวอย่างในแต่ละกลุ่มมีขนาดมากกว่า 50 หน่วย จึงใช้วิธีตรวจสอบการแจกแจงข้อมูลเทคนิคของ Kolmogorov-Smirnov โดยจะยอมรับสมมติฐาน H_0 ถ้าค่า Sig. (Significance) มีค่ามากกว่าค่านัยสำคัญ α ซึ่งงานวิจัยนี้กำหนดให้เท่ากับ 0.05 ดังตารางต่อไปนี้

ตารางที่ 4.17 ตารางแสดงค่าสถิติทดสอบการแจกแจงปกติ (Normality Test) ของค่าความแม่นยำ

	เครื่องมือทดสอบการ ค้นคืนเอกสาร	Kolmogorov-Smirnov		
		Statistic	df	Sig.
ค่าความแม่นยำ	1*	0.197	83	0.000
	2	0.168	83	0.000

ผลการทดสอบในตารางที่ 4.17 ข้างต้นพบว่าค่าสถิติค่า Sig. ของตัวแปรของการค้นคืนเอกสารทั้ง 2 รูปแบบเป็นดังนี้

1) การค้นคืนเอกสารรูปแบบที่ 1* มีค่า Sig. เท่ากับ 0.000 ซึ่งมีค่าน้อยกว่าค่า $\alpha = 0.05$ ดังนั้นจึงปฏิเสธสมมติฐาน H_0

2) การค้นคืนเอกสารรูปแบบที่ 2 มีค่า Sig. เท่ากับ 0.000 ซึ่งมีค่าน้อยกว่าค่า $\alpha = 0.05$ ดังนั้นจึงปฏิเสธสมมติฐาน H_0

ดังนั้นสรุปได้ว่าค่าความแม่นยำ (Precision) ของการค้นคืนเอกสารทั้ง 2 รูปแบบนั้น ไม่ได้มีการแจกแจงแบบปกติ ผู้วิจัยจึงใช้การทดสอบสมมติฐานแบบไม่อิงพารามิเตอร์ (Non Parametric Test) โดยเลือกใช้การทดสอบแบบแมนวิทนีย์ (Mann-Whitney Test)

การวิเคราะห์เปรียบเทียบค่าความแม่นยำ (Precision) ของการค้นคืนเอกสารทั้ง 2 รูปแบบ

จากการวิเคราะห์การแจกแจงของข้อมูลเห็นได้ว่า การแจกแจงของตัวแปรค่าความแม่นยำ (Precision) ของการค้นคืนเอกสารทั้ง 2 รูปแบบนั้น ไม่ได้มีการแจกแจงแบบปกติ ดังนั้นผู้วิจัยจึงทดสอบสมมติฐานด้วยการทดสอบแบบแมนวิทนีย์ (Mann-Whitney Test) โดยใช้โปรแกรม SPSS กำหนดให้

μ_1 คือ ค่าเฉลี่ยค่าความแม่นยำของการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียนด้วยการค้นคืนเอกสารต่อผู้ใช้ภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering) หรือการค้นคืนเอกสารรูปแบบที่ 1*

μ_2 คือ ค่าเฉลี่ยค่าความแม่นยำของการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ ด้วยวิธีการวัดความคล้ายคลึงเชิงมุมหรือการค้นคืนเอกสารรูปแบบที่ 2

วิเคราะห์เปรียบเทียบค่าความแม่นยำของการค้นคืนเอกสารทั้ง 2 รูปแบบว่ามีความแตกต่างกันหรือไม่ โดยตั้งสมมติฐานดังนี้

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_1: \mu_1 - \mu_2 \neq 0$$

ผลการทดลองที่ได้ เมื่อใช้สถิติทดสอบแบบแมนวิทนี (Mann-Whitney Test) สามารถแสดงค่าสถิติทดสอบได้ดังตารางที่ 4.18 และจากสมมติฐานข้างต้นจะปฏิเสธสมมติฐาน H_0 เมื่อค่า Asymp. Sig. (2-tailed) น้อยกว่าระดับค่านัยสำคัญ α (กัลยา วาณิชย์บัญชา, 2548)

ตารางที่ 4.18 ตารางแสดงค่าสถิติทดสอบค่าความแม่นยำของการค้นคืนเอกสาร

	ค่าความแม่นยำของการค้นคืนเอกสาร
Mann-Whitney U	491.000
Z	-9.550
Asymp. Sig. (2-tailed)	0.000

จากตารางที่ 4.18 จะเห็นได้ว่าค่า Asymp. Sig. (2-tailed) มีค่าเท่ากับ 0.000 ซึ่งมีค่าน้อยกว่าค่านัยสำคัญ $\alpha = 0.05$ ดังนั้น จึงปฏิเสธสมมติฐาน H_0 กล่าวคือ ค่าเฉลี่ยค่าความแม่นยำของการค้นคืนเอกสารทั้ง 2 รูปแบบแตกต่างกันอย่างมีนัยสำคัญ

สรุปการวิเคราะห์ข้อมูลประสิทธิภาพของค่าความแม่นยำ (Precision)

ค่าความแม่นยำ (Precision) ของการค้นคืนเอกสารรูปแบบที่ 1* แตกต่างกับรูปแบบที่ 2 อย่างมีนัยสำคัญ และเมื่อพิจารณาจากค่าเฉลี่ย (Mean) ของค่าความแม่นยำจากตารางที่ 4.16 จะเห็นได้ว่าค่าเฉลี่ยค่าความแม่นยำของการค้นคืนเอกสารรูปแบบที่ 1* น้อยกว่าการค้นคืนเอกสารรูปแบบที่ 2 ดังนั้น ค่าความแม่นยำของการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน ด้วยการค้นคืนเอกสารต่อผู้ใช้ภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering) มีค่าน้อยกว่าค่าความแม่นยำของการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงมุม

4.3.2 การวิเคราะห์ข้อมูลค่าความระลึก (Recall)

การวิเคราะห์การแจกแจงข้อมูล

ผู้วิจัยจะตรวจสอบว่าค่าความระลึก (Recall) ของการค้นคืนเอกสารมีการแจกแจงแบบปกติหรือไม่ จากค่าความระลึก (Recall) 83 ค่าที่ได้ในการค้นคืนเอกสารแต่ละรูปแบบ โดยตั้งสมมติฐานของการทดสอบสำหรับทดสอบค่าตัวแปรค่าความระลึก (Recall) ของการค้นคืนเอกสารแต่ละรูปแบบ ดังนี้

1) ทดสอบการแจกแจงค่าความระลึกของการค้นคืนเอกสารรูปแบบที่ 1*

H_0 : ข้อมูลค่าความระลึกของการค้นคืนเอกสารรูปแบบที่ 1* มีการแจกแจงแบบปกติ

H_1 : ข้อมูลค่าความระลึกของการค้นคืนเอกสารรูปแบบที่ 1* ไม่ได้มีการแจกแจงแบบปกติ

2) ทดสอบการแจกแจงค่าความระลึกของการค้นคืนเอกสารรูปแบบที่ 2

H_0 : ข้อมูลค่าความระลึกของการค้นคืนเอกสารรูปแบบที่ 2 มีการแจกแจงแบบปกติ

H_1 : ข้อมูลค่าความระลึกของการค้นคืนเอกสารรูปแบบที่ 2 ไม่ได้มีการแจกแจงแบบปกติ

การตรวจสอบการแจกแจงของข้อมูลเชิงปริมาณมีสถิติทดสอบที่ใช้คือ Kolmogorov-Smirnov เมื่อขนาดตัวอย่างมากกว่า 50 หน่วย และของ Shapiro-Wilk เมื่อขนาดตัวอย่างน้อยกว่า 50 หน่วย (กัลยา วานิชย์บัญชา, 2546) สำหรับในงานวิจัยนี้ตัวอย่างในแต่ละกลุ่มมีขนาดมากกว่า 50 หน่วย จึงใช้วิธีตรวจสอบการแจกแจงข้อมูลเทคนิคของ Kolmogorov-Smirnov โดยจะยอมรับสมมติฐาน H_0 ถ้าค่า Sig. (Significance) มีค่ามากกว่าค่านัยสำคัญ α ซึ่งงานวิจัยนี้กำหนดให้เท่ากับ 0.05 ดังตารางต่อไปนี้

ตารางที่ 4.19 ตารางแสดงค่าสถิติทดสอบการแจกแจงปกติ (Normality Test) ของค่าความระลึก

	เครื่องมือทดสอบการค้นคืนเอกสาร	Kolmogorov-Smirnov		
		Statistic	df	Sig.
ค่าความระลึก	1*	0.168	83	0.000
	2	0.411	83	0.000

ผลการทดสอบในตารางที่ 4.19 ข้างต้นพบว่าค่าสถิติค่า Sig. ของตัวแปรของการค้นคืนเอกสารทั้ง 2 รูปแบบเป็นดังนี้

1) การค้นคืนเอกสารรูปแบบที่ 1* มีค่า Sig. เท่ากับ 0.000 ซึ่งมีค่าน้อยกว่าค่า $\alpha = 0.05$ ดังนั้นจึงปฏิเสธสมมติฐาน H_0

2) การค้นคืนเอกสารรูปแบบที่ 2 มีค่า Sig. เท่ากับ 0.000 ซึ่งมีค่าน้อยกว่าค่า $\alpha = 0.05$ ดังนั้นจึงปฏิเสธสมมติฐาน H_0

ดังนั้นสรุปได้ว่าค่าความระลึก (Recall) ของการค้นคืนเอกสารทั้ง 2 รูปแบบนั้น ไม่ได้มีการแจกแจงแบบปกติ ผู้วิจัยจึงใช้การทดสอบสมมติฐานแบบไม่อิงพารามิเตอร์ (Non Parametric Test) โดยเลือกใช้การทดสอบแบบแมนวิทนี (Mann-Whitney Test)

การวิเคราะห์เปรียบเทียบค่าความระลึก (Recall) ของการค้นคืนเอกสารทั้ง 2 รูปแบบ

จากการวิเคราะห์การแจกแจงของข้อมูลเห็นได้ว่า การแจกแจงของตัวแปรค่าความระลึก (Recall) ของการค้นคืนเอกสารทั้ง 2 รูปแบบนั้น ไม่ได้มีการแจกแจงแบบปกติ ดังนั้นผู้วิจัยจึงทดสอบสมมติฐานด้วยการทดสอบแบบแมนวิทนี (Mann-Whitney Test)

กำหนดให้

μ_3 คือ ค่าเฉลี่ยค่าความระลึกของการค้นคืนเอกสารที่ใช้เทคนิคบริภูมิเวกเตอร์ ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียนด้วยการค้นคืนเอกสารต่อผู้ใช้ภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering) หรือการค้นคืนเอกสารรูปแบบที่ 1*

μ_4 คือ ค่าเฉลี่ยค่าความระลึกของการค้นคืนเอกสารที่ใช้เทคนิคบริภูมิเวกเตอร์ ด้วยวิธีการวัดความคล้ายคลึงเชิงมุมหรือการค้นคืนเอกสารรูปแบบที่ 2

วิเคราะห์เปรียบเทียบค่าความระลึกของการค้นคืนเอกสารทั้ง 2 รูปแบบว่ามีความแตกต่างกันหรือไม่ โดยตั้งสมมติฐานดังนี้

$$H_0 : \mu_3 - \mu_4 = 0$$

$$H_1 : \mu_3 - \mu_4 \neq 0$$

ผลการทดลองที่ได้ เมื่อใช้สถิติทดสอบแบบแมนวิทนี (Mann-Whitney Test) สามารถแสดงค่าสถิติทดสอบได้ดังตารางที่ 4.20 และจากสมมติฐานข้างต้นจะปฏิเสธสมมติฐาน H_0 เมื่อค่า Asymp. Sig. (2-tailed) น้อยกว่าระดับค่านัยสำคัญ α (กัลยา วานิชย์บัญชา, 2548)

ตารางที่ 4.20 ตารางแสดงค่าสถิติทดสอบค่าความระลึกของการค้นคืนเอกสาร

	ค่าความระลึกของการค้นคืนเอกสาร
Mann-Whitney U	1420.000
Z	-7.024
Asymp. Sig. (2-tailed)	0.000

จากตารางที่ 4.26 จะเห็นได้ว่าค่า Asymp. Sig. (2-tailed) มีค่าเท่ากับ 0.000 ซึ่งมีค่าน้อยกว่าค่านัยสำคัญ $\alpha = 0.05$ ดังนั้น จึงปฏิเสธสมมติฐาน H_0 กล่าวคือ ค่าเฉลี่ยค่าความระลึกของการค้นคืนเอกสารทั้ง 2 รูปแบบแตกต่างกันอย่างมีนัยสำคัญ

สรุปการวิเคราะห์ข้อมูลประสิทธิภาพของค่าความระลึก (Recall)

ค่าความระลึก (Recall) ของการค้นคืนเอกสารรูปแบบที่ 1* แตกต่างกับรูปแบบที่ 2 อย่างมีนัยสำคัญ และเมื่อพิจารณาจากค่าเฉลี่ย (Mean) ของค่าความระลึกจากตารางที่ 4.16 จะเห็นได้ว่าค่าเฉลี่ยค่าความระลึกของการค้นคืนเอกสารรูปแบบที่ 1* น้อยกว่าการค้นคืนเอกสารรูปแบบที่ 2 ดังนั้น ค่าความระลึกของการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน ด้วยการค้นคืนเอกสารต่อผู้ใช้ภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering) มีค่าน้อยกว่าค่าความระลึกของการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงมุม

4.3.3 การวิเคราะห์ข้อมูลค่าเฉลี่ยฮาร์โมนิก (Harmonic mean)

การวิเคราะห์การแจกแจงข้อมูล

ผู้วิจัยจะตรวจสอบว่าค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของการค้นคืนเอกสารมีการแจกแจงแบบปกติหรือไม่ จากค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) 83 ค่าที่ได้ในการค้นคืนเอกสารแต่ละรูปแบบ โดยตั้งสมมติฐานของการทดสอบสำหรับทดสอบค่าตัวแปรค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของการค้นคืนเอกสารแต่ละรูปแบบ ดังนี้

1) ทดสอบการแจกแจงค่าเฉลี่ยฮาร์โมนิกการคั่นคั้นเอกสารรูปแบบที่ 1*

H_0 : ข้อมูลค่าเฉลี่ยฮาร์โมนิกของการคั่นคั้นเอกสารรูปแบบที่ 1* มีการแจกแจงแบบปกติ

H_1 : ข้อมูลค่าเฉลี่ยฮาร์โมนิกของการคั่นคั้นเอกสารรูปแบบที่ 1* ไม่ได้มีการแจกแจงแบบปกติ

2) ทดสอบการแจกแจงค่าเฉลี่ยฮาร์โมนิกการคั่นคั้นเอกสารรูปแบบที่ 2

H_0 : ข้อมูลค่าเฉลี่ยฮาร์โมนิกของการคั่นคั้นเอกสารรูปแบบที่ 2 มีการแจกแจงแบบปกติ

H_1 : ข้อมูลค่าเฉลี่ยฮาร์โมนิกของการคั่นคั้นเอกสารรูปแบบที่ 2 ไม่ได้มีการแจกแจงแบบปกติ

การตรวจสอบการแจกแจงของข้อมูลเชิงปริมาณมีสถิติทดสอบที่ใช้คือ Kolmogorov-Smirnov เมื่อขนาดตัวอย่างมากกว่า 50 หน่วย และของ Shapiro-Wilk เมื่อขนาดตัวอย่างน้อยกว่า 50 หน่วย (กัลยา วานิชย์บัญชา, 2546) สำหรับในงานวิจัยนี้ตัวอย่างในแต่ละกลุ่มมีขนาดมากกว่า 50 หน่วย จึงใช้วิธีตรวจสอบการแจกแจงข้อมูลเทคนิคของ Kolmogorov-Smirnov โดยจะยอมรับสมมติฐาน H_0 ถ้าค่า Sig. (Significance) มีค่ามากกว่าค่านัยสำคัญ α ซึ่งงานวิจัยนี้กำหนดให้เท่ากับ 0.05 ดังตารางต่อไปนี้

ตารางที่ 4.21 ตารางแสดงค่าสถิติทดสอบการแจกแจงปกติ (Normality Test) ของค่าเฉลี่ยฮาร์โมนิก

	เครื่องมือทดสอบการคั่นคั้นเอกสาร	Kolmogorov-Smirnov		
		Statistic	df	Sig.
ค่าเฉลี่ยฮาร์โมนิก	1*	0.188	83	0.000
	2	0.142	83	0.000

ผลการทดสอบในตารางที่ 4.21 ช่างต้นพบว่าค่าสถิติค่า Sig. ของตัวแปรของการคั่นคั้นเอกสารทั้ง 2 รูปแบบเป็นดังนี้

1) การคั่นคั้นเอกสารรูปแบบที่ 1* มีค่า Sig. เท่ากับ 0.000 ซึ่งมีค่าน้อยกว่าค่า $\alpha = 0.05$ ดังนั้นจึงปฏิเสธสมมติฐาน H_0

2) การค้นคืนเอกสารรูปแบบที่ 2 มีค่า Sig. เท่ากับ 0.000 ซึ่งมีค่าน้อยกว่าค่า $\alpha = 0.05$ ดังนั้นจึงปฏิเสธสมมติฐาน H_0

ดังนั้นสรุปได้ว่าค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของการค้นคืนเอกสารทั้ง 2 รูปแบบนั้น ไม่ได้มีการแจกแจงแบบปกติ ผู้วิจัยจึงใช้การทดสอบสมมติฐานแบบไม่อิงพารามิเตอร์ (Non Parametric Test) โดยเลือกใช้การทดสอบแบบแมนวิทนี (Mann-Whitney Test)

การวิเคราะห์เปรียบเทียบค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของการค้นคืนเอกสารทั้ง 2 รูปแบบ

จากการวิเคราะห์การแจกแจงของข้อมูลเห็นได้ว่า การแจกแจงของตัวแปรค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของการค้นคืนเอกสารทั้ง 2 รูปแบบนั้น ไม่ได้มีการแจกแจงแบบปกติ ดังนั้นผู้วิจัยจึงทดสอบสมมติฐานด้วยการทดสอบแบบแมนวิทนี (Mann-Whitney Test)

กำหนดให้

μ_5 คือ ค่าเฉลี่ยของค่าเฉลี่ยฮาร์โมนิกในการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิดีเนียนด้วยการค้นคืนเอกสารต่อผู้ใช้ภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering) หรือการค้นคืนเอกสารรูปแบบที่ 1*

μ_6 คือ ค่าเฉลี่ยของค่าเฉลี่ยฮาร์โมนิกในการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงมุมหรือการค้นคืนเอกสารรูปแบบที่ 2

วิเคราะห์เปรียบเทียบค่าเฉลี่ยฮาร์โมนิกของการค้นคืนเอกสารทั้ง 2 รูปแบบว่ามีความแตกต่างกันหรือไม่ โดยตั้งสมมติฐานดังนี้

$$H_0 : \mu_5 - \mu_6 = 0$$

$$H_1 : \mu_5 - \mu_6 \neq 0$$

ผลการทดลองที่ได้ เมื่อใช้สถิติทดสอบแบบแมนวิทนี (Mann-Whitney Test) สามารถแสดงค่าสถิติทดสอบได้ดังตารางที่ 4.22 และจากสมมติฐานข้างต้นจะปฏิเสธสมมติฐาน H_0 เมื่อค่า Asymp. Sig. (2-tailed) น้อยกว่าระดับค่านัยสำคัญ α (กัลยา วานิชย์บัญชา, 2548)

ตารางที่ 4.22 ตารางแสดงค่าสถิติทดสอบค่าเฉลี่ยฮาร์โมนิกของการค้นคืนเอกสาร

	ค่าเฉลี่ยฮาร์โมนิกของการค้นคืนเอกสาร
Mann-Whitney U	472.000
Z	-9.612
Asymp. Sig. (2-tailed)	0.000

จากตารางที่ 4.22 จะเห็นได้ว่าค่า Asymp. Sig. (2-tailed) มีค่าเท่ากับ 0.000 ซึ่งมีค่าน้อยกว่าค่านัยสำคัญ $\alpha = 0.05$ ดังนั้น จึงปฏิเสธสมมติฐาน H_0 กล่าวคือ ค่าเฉลี่ยของค่าเฉลี่ยฮาร์โมนิกของการค้นคืนเอกสารทั้ง 2 รูปแบบแตกต่างกันอย่างมีนัยสำคัญ

สรุปการวิเคราะห์ข้อมูลประสิทธิภาพของค่าเฉลี่ยฮาร์โมนิก (Harmonic mean)

ค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของการค้นคืนเอกสารรูปแบบที่ 1* แตกต่างกับรูปแบบที่ 2 อย่างมีนัยสำคัญ และเมื่อพิจารณาจากค่าเฉลี่ย (Mean) ของค่าเฉลี่ยฮาร์โมนิกจากตารางที่ 4.16 จะเห็นได้ว่าค่าเฉลี่ยของค่าเฉลี่ยฮาร์โมนิกของการค้นคืนเอกสารรูปแบบที่ 1* น้อยกว่าการค้นคืนเอกสารรูปแบบที่ 2 ดังนั้น ค่าเฉลี่ยฮาร์โมนิกของการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน ด้วยการค้นคืนเอกสารต่อผู้ใช้ภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering) มีค่าน้อยกว่าค่าเฉลี่ยฮาร์โมนิกของการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงมุม

4.4 สรุปผลการทดลอง

จากการวิเคราะห์ผลการทดลองการวัดประสิทธิภาพการค้นคืนเอกสาร ด้วยค่าความแม่นยำ (Precision), ค่าความระลึก (Recall) และค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) สามารถสรุปได้ว่า ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของการค้นคืนเอกสารรูปแบบที่ 1* มีค่าน้อยกว่าการค้นคืนเอกสารรูปแบบที่ 2 แสดงว่า การค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียนด้วยการค้นคืนเอกสารต่อผู้ใช้ภายในกรอบค่าความคล้ายคลึงที่กำหนดด้วยผลลัพธ์ที่ได้จากเทคนิคการจัดกลุ่มข้อมูล (Clustering) มีประสิทธิภาพด้อยกว่าการค้นคืนเอกสารที่ใช้เทคนิคปริภูมิเวกเตอร์ด้วยวิธีการวัดความคล้ายคลึงเชิงมุม

4.5 อภิปรายผลการทดลอง

งานวิจัยนี้ต้องการศึกษาเปรียบเทียบประสิทธิภาพการค้นคืนเอกสารโดยใช้เทคนิคปริภูมิเวกเตอร์ (Vector Space Model) ด้วยการวัดความคล้ายคลึงเชิงมุม (Cosine Angle) และการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน (Euclidean Distance) ด้วยการค้นคืนเอกสารต่อผู้ใช้ภายในกรอบค่าความคล้ายคลึงที่กำหนดโดยการประยุกต์ใช้เทคนิคการจัดกลุ่มข้อมูล (Clustering) จากผลการทดลองและวิเคราะห์ข้อมูลสรุปได้ว่า การค้นคืนเอกสารรูปแบบที่ 1* มีประสิทธิภาพดีกว่าการค้นคืนเอกสารรูปแบบที่ 2 ทั้งค่าความแม่นยำ, ความระลึก และค่าเฉลี่ยฮาร์โมนิค นั่นคือ การค้นคืนเอกสารโดยใช้เทคนิคปริภูมิเวกเตอร์ (Vector Space Model) ด้วยการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน (Euclidean Distance) ไม่สามารถเพิ่มประสิทธิภาพการค้นคืนเอกสารได้

ผู้วิจัยจึงตั้งข้อสังเกตว่า อาจจะมีสาเหตุมาจากวิธีการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน ซึ่งการวัดค่าความคล้ายคลึงเชิงระยะห่างยูคลิเดียน (Euclidean distance) นั้นเป็นวิธีวัดระยะทาง (Distance) ระหว่างข้อมูล ด้วยการคำนวณวัดระยะทางระหว่าง 2 จุดให้เป็นรากที่สองของผลรวมของผลต่างของข้อมูลในแต่ละมิติยกกำลังสอง (ดังสมการที่ 2.8) เมื่อนำมาใช้ในกระบวนการค้นคืนเอกสาร ผลการคำนวณความคล้ายคลึงระหว่างเวกเตอร์เอกสารและเวกเตอร์ข้อสอบถาม จะขึ้นอยู่กับ ค่าน้ำหนักของค่าในแต่ละมิติของเวกเตอร์เอกสารและข้อสอบถาม ส่วนการวัดความคล้ายคลึงเชิงมุม (Cosine Angle) (ดังสมการที่ 2.7) ผลการคำนวณความคล้ายคลึงระหว่างเวกเตอร์เอกสารและเวกเตอร์ข้อสอบถามจะขึ้นอยู่กับ ค่าน้ำหนักของค่าในตำแหน่งมิติของเวกเตอร์ข้อสอบถามเท่านั้น ดังตารางที่ 4.23 ซึ่งเป็นตารางแสดงตัวอย่างการคำนวณค่าความคล้ายคลึงหรือความเหมือนระหว่างเวกเตอร์เอกสารและเวกเตอร์ข้อสอบถามในแต่ละมิติ ด้วยวิธีวัดความคล้ายคลึงเชิงมุมและเชิงระยะห่างยูคลิเดียน โดยกำหนดให้ ตัวอย่างของเวกเตอร์เอกสารและเวกเตอร์ข้อสอบถามมีขนาดมิติเท่ากับ 4

ตารางที่ 4.23 ตารางแสดงตัวอย่างการคำนวณค่าความเหมือนระหว่างเวกเตอร์เอกสารและ
ข้อสอบถามในแต่ละมิติ

มิติ	ค่าน้ำหนักของคำใน มิติเวกเตอร์ข้อสอบถาม	ค่าน้ำหนักของคำใน มิติเวกเตอร์เอกสาร	ค่าความเหมือน ระยะห่างเชิงมุม	ค่าความเหมือน ระยะห่างยูคลิเดียน
1	1	0	0	1
2	0	3	0	9
3	2	1	0.28	1
4	0	0	0	0

* หมายเหตุ ค่าน้ำหนักของคำในตำแหน่งมิติเท่ากับ 0 คือ เอกสารหรือข้อสอบถามไม่มีคำใน
ตำแหน่งมิตินั้นปรากฏ

จากตารางที่ 4.23 หากคำนวณตามสมการที่ 2.7 ค่าความเหมือนระยะห่างเชิงมุมระหว่าง
ตัวอย่างเวกเตอร์เอกสารและเวกเตอร์ข้อสอบถามมีค่าเท่ากับ 0.28 และคำนวณตามสมการที่ 2.8
ค่าความเหมือนระยะห่างยูคลิเดียนระหว่างตัวอย่างเวกเตอร์เอกสารและเวกเตอร์ข้อสอบถามมีค่า
เท่ากับ 3.32 เมื่อพิจารณาเฉพาะค่าความเหมือนในแต่ละมิติ จะสามารถสรุปเงื่อนไขการคำนวณ
ได้เป็น 4 เงื่อนไขด้วยกัน คือ

- 1) เมื่อในมิติเวกเตอร์ข้อสอบถามมีค่าน้ำหนักของคำ แต่ในมิตินั้นเวกเตอร์เอกสารไม่มี
ค่าน้ำหนักของคำ (ตัวอย่างเช่น มิติที่ 1 ในตารางที่ 4.23) ผลการคำนวณค่าความเหมือนระยะห่าง
เชิงมุมจะมีค่าเท่ากับ 0 คือ ไม่มีค่าความเหมือน (ตามสมการที่ 2.7) ในขณะที่ผลการคำนวณค่า
ความเหมือนระยะห่างยูคลิเดียน จะมีค่าความเหมือนไม่เท่ากับ 0 โดยมีค่าความเหมือนตามค่า
น้ำหนักของคำในข้อสอบถาม (ตามสมการที่ 2.8)
- 2) เมื่อในมิติเวกเตอร์ข้อสอบถามไม่มีมีค่าน้ำหนักของคำ แต่ในมิตินั้นเวกเตอร์เอกสารมี
ค่าน้ำหนักของคำ (ตัวอย่างเช่น มิติที่ 2 ในตารางที่ 4.23) ผลการคำนวณค่าความเหมือนระยะห่าง
เชิงมุมจะมีค่าเท่ากับ 0 ในขณะที่ผลการคำนวณค่าความเหมือนระยะห่างยูคลิเดียน จะมีค่าความ
เหมือนไม่เท่ากับ 0 โดยคิดค่าความเหมือนตามค่าน้ำหนักของคำในเอกสาร
- 3) เมื่อในมิติเวกเตอร์ข้อสอบถามและเวกเตอร์เอกสารมีค่าน้ำหนักของคำในมิติทั้ง 2 คำ
(ตัวอย่างเช่น มิติที่ 3 ในตารางที่ 4.23) ผลการคำนวณค่าความเหมือน ทั้งจากการคำนวณค่า
ความเหมือนระยะห่างเชิงมุมและการคำนวณค่าความเหมือนระยะห่างยูคลิเดียน ต่างมีค่าความ
เหมือนของมิตินั้น

4) เมื่อในมิติเวกเตอร์ข้อสอบถามและเวกเตอร์เอกสารไม่มีค่าน้ำหนักของคำในมิติปรากฏอยู่ (ตัวอย่างเช่น มิติที่ 4 ในตารางที่ 4.23) ผลการคำนวณค่าความเหมือนทั้ง 2 วิธีในมิตินั้นจะมีค่าเท่ากับ 0 (ไม่มีค่าความเหมือน)

จากเงื่อนไขการคำนวณข้างต้น สามารถสรุปได้ว่า วิธีระยะห่างเชิงมุมจะคิดค่าความคล้ายคลึงระหว่างเอกสารและข้อสอบถาม ก็ต่อเมื่อค่าน้ำหนักของคำในมิติของเวกเตอร์ข้อสอบถามและเวกเตอร์เอกสารมีค่าทั้งคู่ แสดงว่าวิธีระยะห่างเชิงมุมจะคิดค่าความคล้ายคลึงระหว่างเอกสารและข้อสอบถามเฉพาะตำแหน่งมิติของคำที่มีปรากฏทั้งในข้อสอบถามและเอกสารเท่านั้น แต่วิธีระยะห่างเชิงยูคลิเดียนจะมีค่าความคล้ายคลึงระหว่างเอกสารและข้อสอบถามทุกกรณียกเว้นกรณีที่ไม่มีค่าน้ำหนักของคำในมิติทั้งเวกเตอร์ข้อสอบถามและเวกเตอร์เอกสาร หรือวิธีระยะห่างเชิงยูคลิเดียนจะคิดค่าความคล้ายคลึงระหว่างเอกสารและข้อสอบถามที่ไม่จำกัดเฉพาะคำที่ตรงกันระหว่างข้อสอบถามและเอกสารเท่านั้น

ข้อสังเกตต่อไปของผู้วิจัย คือลักษณะของชุดเอกสารที่ใช้ในการทดลอง เพราะเอกสารนิตยสารไทม์ (Time) จากฐานข้อมูลนิตยสารไทม์ (TIME Collection) เป็นเรื่องราวข่าวสารทั่วไปจึงมีเนื้อหาสาระที่หลากหลายไม่เป็นหมวดหมู่ ทำให้เมื่อนำมากำหนดเป็นเวกเตอร์ คำที่ปรากฏในระบบทั้งหมดจึงเป็นคำที่อาจจะไม่ได้ปรากฏอยู่ในเอกสารส่วนใหญ่ จากการศึกษาพบว่าจำนวนคำที่ปรากฏในระบบทั้งหมดเท่ากับ 13,813 คำ จำนวนคำที่ปรากฏในแต่ละเอกสารเฉลี่ยเท่ากับ 218 คำ จึงทำให้เวกเตอร์เอกสารมีค่าในตำแหน่งมิติที่ไม่มีค่าปรากฏเท่ากับ 0 เป็นจำนวนมาก นอกจากนี้จำนวนคำที่ปรากฏในแต่ละข้อสอบถามคิดเฉลี่ยได้เท่ากับ 9 คำ ซึ่งทำให้ตำแหน่งมิติเวกเตอร์ข้อสอบถามที่ไม่มีค่าปรากฏมีค่าเท่ากับ 0 เป็นจำนวนมากเช่นเดียวกัน

ดังนั้นเมื่อนำเวกเตอร์เอกสารไปจัดกลุ่มเอกสารด้วยเทคนิค K-means Clustering บนระยะห่างยูคลิเดียน จะแยกความแตกต่างระหว่างแต่ละเอกสารได้ยาก และอาจเป็นไปได้ว่า กลุ่มที่จัดได้เป็นกลุ่มใหญ่ เป็นกลุ่มที่มีค่าในตำแหน่งมิติเท่ากับ 0 เป็นส่วนใหญ่ หรือเป็นกลุ่มของเอกสารที่มีค่าปรากฏน้อย (ผลการทดลองแสดงในภาคผนวก จ)

เมื่อนำเวกเตอร์ข้อสอบถามทั้ง 83 ข้อ มาพิจารณาว่าอยู่ในกลุ่มใดจากผลการจัดกลุ่มของเวกเตอร์เอกสารที่ได้กำหนดไว้แล้วนั้น ผู้วิจัยตั้งข้อสังเกตว่า เมื่อนำเวกเตอร์ข้อสอบถามซึ่งมีค่าในตำแหน่งมิติเป็น 0 ส่วนใหญ่ มาคำนวณหาระยะห่างยูคลิเดียนกับเวกเตอร์เอกสาร เอกสารที่มีความคล้ายคลึงกับข้อสอบถามจะกลายเป็นเอกสารที่มีค่าในตำแหน่งมิติส่วนใหญ่เป็น 0 ไม่ใช่เอกสารที่มีค่าตรงกับในข้อสอบถาม เป็นไปได้ว่า ข้อสอบถามจะถูกจัดอยู่ในกลุ่มใหญ่ และจากผลการทดลองพบว่า เวกเตอร์ข้อสอบถามโดยส่วนใหญ่จะถูกจัดให้อยู่ในกลุ่มใหญ่เสมอ (ผลการ

ทดลองแสดงในภาคผนวก ฉ) จึงแสดงว่าเวกเตอร์ข้อสอบถามทั้ง 83 ข้อ จะถูกจัดกลุ่มให้อยู่ในกลุ่มของเอกสารที่มีค่าปรากฏน้อย และอาจจะไม่มีค่าปรากฏตรงกันของเวกเตอร์เอกสารและข้อสอบถาม ดังนั้นนำค่ารัศมีของกลุ่มใหญ่นั้นมากำหนดกรอบค่าความคล้ายคลึง ทำให้กรอบค่าความคล้ายคลึงของการค้นคืนเอกสารกว้าง เป็นสาเหตุให้ระบบค้นคืนเอกสารออกมาเป็นจำนวนมาก ส่งผลให้ประสิทธิภาพค่าความแม่นยำในการค้นคืนต่ำ และค่าความระลึกลงในการค้นคืนสูง

ผลการทดลองเปรียบเทียบประสิทธิภาพการค้นคืนเอกสารรูปแบบที่ 1* และ 2 แสดงในตารางที่ 4.16 พบว่าประสิทธิภาพค่าความแม่นยำของการค้นคืนเอกสารรูปแบบที่ 1* ต่ำกว่าการค้นคืนรูปแบบที่ 2 เท่ากับ 88.76% ค่าความระลึกลงของการค้นคืนเอกสารรูปแบบที่ 1* ต่ำกว่าการค้นคืนเอกสารรูปแบบที่ 2 เท่ากับ 41.82% ซึ่งหมายความว่า เอกสารที่ไม่ได้เป็นคำตอบจะถูกค้นคืนออกมาเป็นจำนวนมาก และเอกสารที่เป็นคำตอบจะถูกค้นคืนออกมาเป็นจำนวนน้อย หรือไม่ถูกค้นคืน ซึ่งผิดวัตถุประสงค์ของกระบวนการค้นคืนเอกสาร ที่ระบุว่าระบบค้นคืนเอกสารจะต้องประกอบไปด้วยเอกสารที่ผู้ใช้ต้องการในปริมาณที่มากที่สุด และเอกสารที่ผู้ใช้ไม่ต้องการในปริมาณที่น้อยที่สุด

ผู้วิจัยจึงสรุปว่า การค้นคืนเอกสารโดยใช้เทคนิคปริภูมิเวกเตอร์ (Vector Space Model) ด้วยการวัดความคล้ายคลึงเชิงระยะห่างยูคลิเดียน (Euclidean Distance) ไม่สามารถเพิ่มประสิทธิภาพได้เท่ากับการค้นคืนเอกสารโดยใช้เทคนิคปริภูมิเวกเตอร์ ด้วยวิธีวัดความคล้ายคลึงเชิงมุม (Cosine Angle) ได้ ดังนั้นผู้วิจัยจึงต้องการศึกษาเพิ่มเติมว่า ถ้านำเทคนิคการค้นคืนเอกสารภายในกรอบค่าความคล้ายคลึงที่กำหนดโดยการประยุกต์ใช้เทคนิคการจัดกลุ่มข้อมูล (Clustering) ด้วยการวัดความคล้ายคลึงเชิงมุม (Cosine Angle) จะสามารถเพิ่มประสิทธิภาพการค้นคืนเอกสารได้หรือไม่