

ผนวก ญ

การเปรียบเทียบการประมวลผลข้อคำถามสำหรับการกระจายแมทริกซ์แบบเซมิดีสครีต ระหว่างสมการของการวิจัยกับสมการของ Kolda และคณะ

เนื่องจากงานวิจัยนี้ได้ปรับปรุงสมการของ Kolda และคณะ (Kolda and O'Leary, 1998, p. 322-346) ตามสมการที่ 3.9 เป็นสมการที่ 3.8 เพื่อให้การกระทำกับแมทริกซ์ในการประมวลผลข้อคำถามลดลง ส่งผลให้ใช้เวลาในการประมวลผลข้อคำถามน้อยลง อย่างไรก็ตาม เพื่อพิสูจน์ว่า สมการที่ 3.8 สามารถให้ผลการค้นคืนสารสนเทศเช่นเดียวกับสมการที่ 3.9 ของ Kolda และคณะ จึงได้พิสูจน์ด้วยวิธีการทางคณิตศาสตร์และการทดลองดังต่อไปนี้

ญ 1 การพิสูจน์ด้วยวิธีการทางคณิตศาสตร์

การพิสูจน์ใช้ทฤษฎีทางพีชคณิตของแมทริกซ์ (Meyer, 2000) เกี่ยวกับคุณสมบัติของแมทริกซ์สลับเปลี่ยน (Transpose matrix) และแมทริกซ์เฉียง (Diagonal matrix) ดังนี้

$$\begin{aligned} s &= (D_k^\alpha X_k^T q)^T (D_k^{1-\alpha} Y_k^T) & (3.9) \\ &= (q^T (D_k^\alpha X_k^T)^T) (D_k^{1-\alpha} Y_k^T) & \because (AB)^T = B^T A^T \\ &= (q^T (X^T)^T (D_k^\alpha)^T) (D_k^{1-\alpha} Y_k^T) \dots \because (AB)^T = B^T A^T \\ &= (q^T X (D_k^\alpha)^T) (D_k^{1-\alpha} Y_k^T) & \because (X^T)^T = X \end{aligned}$$

เนื่องจากแมทริกซ์ D_k^α เป็นแมทริกซ์เฉียงจึงเป็นแมทริกซ์สมมาตร ดังนั้นทำให้ $(D_k^\alpha)^T = D_k^\alpha$ (Meyer, 2000, p. 85) เพราะฉะนั้น

$$s = (q^T X D_k^\alpha) (D_k^{1-\alpha} Y_k^T) \quad (3.8)$$

สรุปได้ว่า จากการพิสูจน์ทางคณิตศาสตร์ สมการที่ใช้ในงานวิจัยตามสมการที่ 3.8 ได้ผลลัพธ์การคำนวณเท่ากับสมการที่ 3.9 ของ Kolda และคณะ

ญ 2 การพิสูจน์ด้วยวิธีการทดลอง

การพิสูจน์ด้วยวิธีการทดลองดำเนินการโดยใช้ชุดข้อมูล ชุดข้อความ ฮาร์ดแวร์และซอฟต์แวร์ ตลอดจนขั้นตอนของการทดลอง เช่นเดียวที่ใช้ในงานวิจัยนี้ โดยใช้สมการที่ 3.8 ของงานวิจัย และสมการที่ 3.9 ของ Kolda และคณะในการประมวลผลข้อความ โดยใช้ปัจจัยแยกแมทริกซ์ (Splitting factor) = 0.5 และ ค่าขีดแบ่งความเหมือน (Threshold value) = 1.0

ผลการทดลองพบว่า เมื่อเปรียบเทียบผลการค้นคืนสารสนเทศโดยการประมวลผลข้อความระหว่างสมการของงานวิจัยนี้ ตามสมการ 3.8 กับสมการของ Kolda และ คณะ (Kolda and O'Leary, 1998, p. 322-346) ตามสมการ 3.9 ปรากฏว่า ผลการค้นคืนสารสนเทศในด้านค่าเฉลี่ยของความแม่นยำ (Precision) และการเรียกกลับ (Recall) มีค่าเท่ากัน ดังแสดงในตารางที่ ญ.1 ส่วนเวลาเฉลี่ยที่ใช้ในการประมวลผลข้อความโดยสมการที่ 3.8 น้อยกว่าเวลาเฉลี่ยที่ใช้ในการประมวลผลข้อความเฉลี่ยโดยสมการของ Kolda และ คณะ (Kolda and O'Leary, 1998, p. 322-346) ตามสมการ 3.9 อย่างมีนัยสำคัญทางสถิติ (Paired-Sample T-Test ที่ระดับความเชื่อมั่น 95%) ดังแสดงตารางที่ ญ.2 และ ภาพที่ ญ.1 ถึงภาพที่ ญ.4

ตารางที่ ๑.๑

ผลการค้นคืนสารสนเทศของสมการของการวิจัยกับสมการของ Kolda และคณะ

ชุดเอกสารสาขา	Rank (k)	ผลการค้นคืนสารสนเทศ			
		SDD		SVD	
		Precision (%)	Recall (%)	Precision (%)	Recall (%)
คอมพิวเตอร์	5	5.00	1.27	4.00	0.11
	11	5.00	1.27	3.89	0.38
	28	8.33	1.94	13.30	2.89
	126	32.78	4.69	36.86	14.02
	429	44.20	12.70	49.96	27.97
ประวัติศาสตร์	3	0.00	0.00	2.29	0.50
	11	0.00	0.00	3.06	1.77
	35	3.33	0.12	11.42	5.81
	141	14.17	1.58	24.95	14.41
	421	31.47	14.60	33.60	30.39
บรรณารักษศาสตร์	3	0.00	0.00	5.37	4.01
	8	0.00	0.00	12.68	6.43
	18	3.33	0.66	22.29	8.32
	79	16.14	2.82	26.08	15.80
	478	40.53	18.70	43.32	36.01
รัฐศาสตร์	2	0.00	0.00	0.02	1.67
	10	5.56	0.79	4.27	2.46
	34	8.89	0.84	14.10	4.27
	149	26.86	5.21	43.90	18.56
	505	53.29	14.03	66.53	30.12

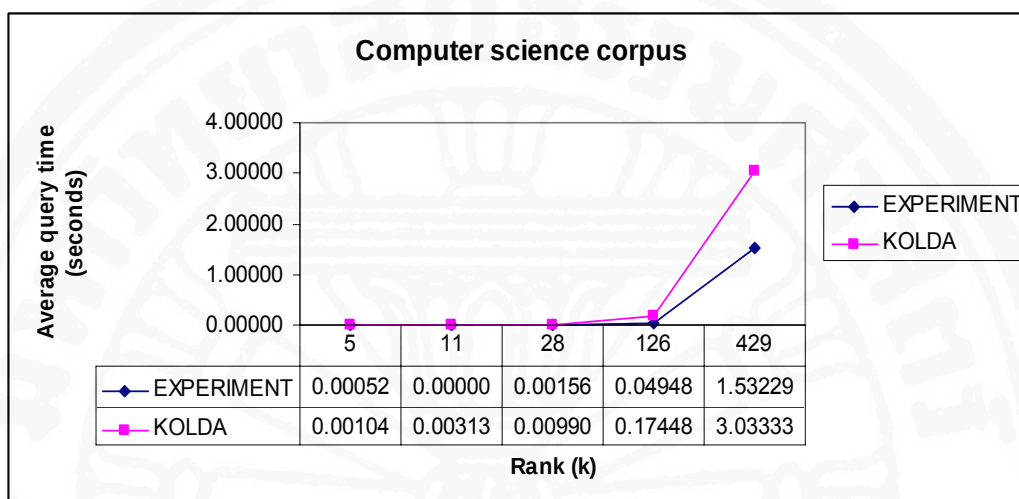
ตารางที่ ๒.๒

เปรียบเทียบเวลาที่ใช้ในการประมวลผลข้อความ (Query time)

ชุดเอกสารสาขา	Rank (k)	เวลาที่ใช้(วินาที)	
		สมการที่ใช้ในงานวิจัย	Kolda และคณะ
วิทยาการคอมพิวเตอร์	5	0.00052	0.00104
	11	0.00000	0.00313
	28	0.00156	0.00990
	126	0.04948	0.17448
	429	1.53229	3.03333
ประวัติศาสตร์	3	0.00052	0.00104
	11	0.00000	0.00260
	35	0.00313	0.02187
	141	0.06875	0.28854
	421	1.44844	3.18542
บรรณารักษศาสตร์	3	0.00000	0.00052
	8	0.00052	0.00156
	18	0.00052	0.00521
	79	0.01719	0.08021
	478	2.08802	3.90052
รัฐศาสตร์	2	0.00000	0.00052
	10	0.00000	0.00208
	34	0.00313	0.01563
	149	0.08438	0.27865
	505	2.47812	4.47719

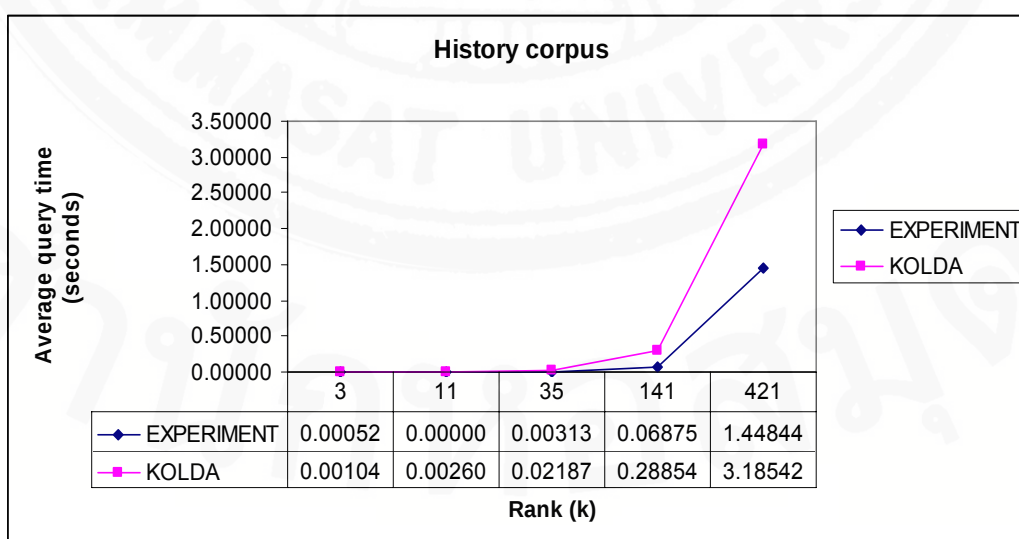
ภาพที่ ญ.1

เปรียบเทียบเวลาที่ใช้ในการประมวลผลข้อคำถามเฉลี่ยกับสมการของ Kolda และคณะ
ของชุดเอกสารสาขาวิทยาการคอมพิวเตอร์



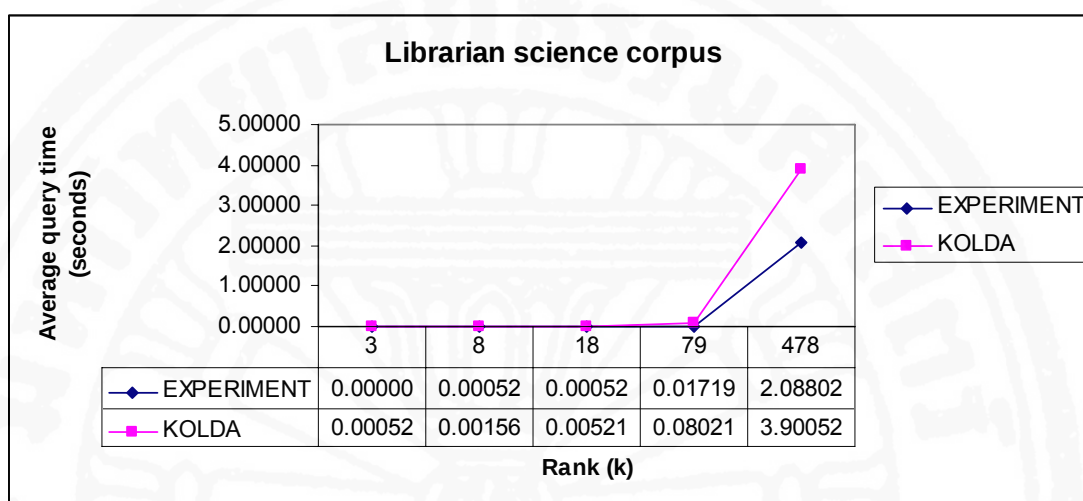
ภาพที่ ญ.2

เปรียบเทียบเวลาที่ใช้ในการประมวลผลข้อคำถามเฉลี่ยกับสมการของ Kolda และคณะ
ของชุดเอกสารสาขาประวัติศาสตร์



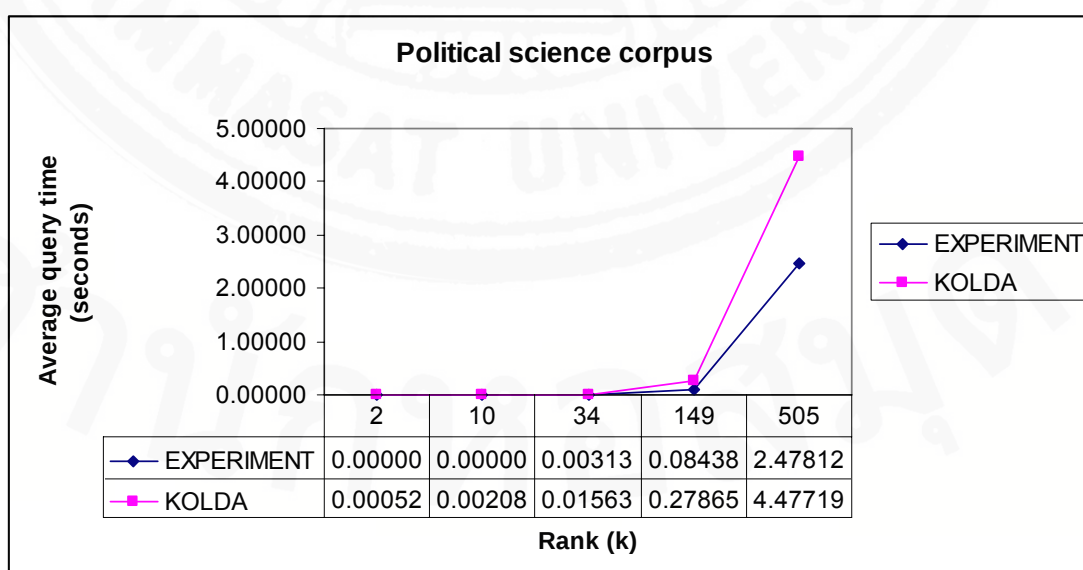
ภาพที่ ๓.๓

เปรียบเทียบเวลาที่ใช้ในการประมวลผลข้อคำถามเฉลี่ยกับสมการของ Kolda และคณะ
ของชุดเอกสารสาขาบรรณารักษศาสตร์



ภาพที่ ๓.๔

เปรียบเทียบเวลาที่ใช้ในการประมวลผลข้อคำถามเฉลี่ยกับสมการของ Kolda และคณะ
ของชุดเอกสารสาขารัฐศาสตร์



กล่าวโดยสรุป การประมวลผลข้อความสำหรับการกระจายแมทริกซ์ด้วยวิธีเฮมิ
ดีสคริตโดยสมการที่ 3.8 ที่ใช้ในงานวิจัย เปรียบเทียบกับสมการที่ 3.9 ของ Kolda และคณะ
ให้ผลการค้นคืนสารสนเทศเท่ากันทั้งการพิสูจน์ด้วยวิธีคณิตศาสตร์และการทดลอง ส่วนเวลาเฉลี่ย
ที่ใช้ในการประมวลผลข้อความของสมการที่ 3.8 ที่ใช้ในงานวิจัย ใช้เวลาน้อยกว่าสมการที่ 3.9
ของ Kolda และคณะอย่างมีนัยสำคัญทางสถิติ เนื่องจากสมการที่ 3.8 มีการกระทำกับแมทริกซ์
น้อยกว่าสมการที่ 3.9 ดูรายละเอียดการวิเคราะห์การกระทำกับแมทริกซ์ในบทที่ 3

