



บทที่ 5

การทดสอบประสิทธิภาพการอัดข้อมูล

จากการแนวทางการปรับเปลี่ยนอัลกอริทึมการอัดข้อมูล 2 วิธีที่กล่าวถึงในบทที่ 4 ได้นำมาพัฒนาเป็น โปรแกรมอัดข้อมูล โปรแกรมขยายข้อมูล โดยที่ขั้นตอนแรกจะเป็นการพัฒนาโปรแกรมตามหลักการเดิมจากนั้นจึงพัฒนาต่อเป็น โปรแกรมที่ปรับเปลี่ยน นำโปรแกรมที่พัฒนาขึ้นทั้งหมดไปทดสอบกับข้อมูลที่จัดเตรียมไว้

เพื่อความสะดวกในการอ้างอิงวิธีการจึงได้กำหนดชื่อย่อแต่ละวิธีไว้ดังนี้

- วิธีการของอัลกอริทึมฮัฟแมน มีชื่อย่อว่า วิธีการ HUFF
- วิธีการของอัลกอริทึมแอลแซดดับบลิว มีชื่อย่อว่า วิธีการ LZW
- วิธีการของอัลกอริทึมฮัฟแมนที่ปรับเปลี่ยน มีชื่อย่อว่า วิธีการ THUFF
- วิธีการของอัลกอริทึมแอลแซดดับบลิวที่ปรับเปลี่ยน มีชื่อย่อว่า วิธีการ TLZW

5.1 ข้อมูลที่ใช้ในการทดสอบ

ข้อมูลที่ใช้ในการทดสอบจะแบ่งออกเป็น 2 กลุ่มคือ กลุ่มข้อความภาษาไทยล้วนและกลุ่มข้อความภาษาไทยทั่ว ๆ ไปที่อาจจะมีลักษณะอื่นรวมอยู่ด้วย เช่น ข้อมูลภาษาอังกฤษ หรือ การตีตารางในข้อความ เป็นต้น

ข้อความภาษาไทยล้วนประกอบด้วยข้อมูลที่ใช้ในการทดสอบ 8 แห่ง เป็นแฟ้มข้อมูลที่จัดเตรียมขึ้นเองโดยนำข้อมูลมาจากหนังสือนิตยสาร นวนิยาย หนังสือพิมพ์ เป็นต้น สำหรับข้อมูลภาษาไทยทั่ว ๆ ไปประกอบด้วยข้อมูลที่ใช้ในการทดสอบ 16 แห่ง เป็นข้อมูลที่นำมาจากงานพิมพ์ที่มีอยู่แล้ว เช่น งานวิจัย วิทยานิพนธ์ รายงานต่าง ๆ เป็นต้น

5.2 ขั้นตอนการทดสอบ

ข้อมูลทั้ง 2 กลุ่มจะทำการทดสอบตามขั้นตอนดังนี้

5.2.1 การทดสอบระหว่างวิธีการเดิมกับวิธีการที่ปรับเปลี่ยน เมื่อวิธีการที่ปรับเปลี่ยนใช้คำไทย 255 คำ ในขั้นตอนนี้จะเป็นการเปรียบเทียบระหว่างวิธีการ HUFF กับวิธีการ THUFF และวิธีการ LZW กับวิธีการ TLZW

5.2.2 การทดสอบระหว่างวิธีการที่ปรับเปลี่ยนที่ใช้คำไทยจำนวนคำไม่เท่ากัน เป็นการเปรียบเทียบระหว่าง วิธีการ THUFF ที่ใช้คำไทย 255 คำ กับวิธีการ THUFF ที่ใช้คำไทย 511 คำ และวิธีการ TLZW ที่ใช้คำไทย 255 คำ กับวิธีการ TLZW ที่ใช้คำไทย 511 คำ

5.3 การทดสอบข้อความภาษาไทยล้วน

5.3.1 การทดสอบตามขั้นตอนที่ 5.2.1

ได้ผลการทดสอบประสิทธิภาพการอัด แสดงในตารางที่ 5.1 และแสดงเป็นกราฟดังรูปที่ 5.1 ส่วนผลการทดสอบเวลาที่ใช้ในการอัดข้อมูลและการขยายข้อมูลแสดงในตารางที่ 5.2 และแสดงเป็นกราฟดังรูปที่ 5.2 รูปที่ 5.3 ตามลำดับ

5.3.2 การทดสอบตามขั้นตอนที่ 5.2.2

ได้ผลการทดสอบประสิทธิภาพการอัด แสดงในตารางที่ 5.3 ส่วนผลการทดสอบเวลาที่ใช้ในการอัดข้อมูลและการขยายข้อมูลแสดงในตารางที่ 5.4

วิธีการ THUFF ให้ประสิทธิภาพการอัดข้อมูลสูงกว่าวิธีการ HUFF และใช้เวลาในการอัดข้อมูลมากขึ้นกว่าเดิม แต่การขยายข้อมูลใช้เวลาใกล้เคียงกับวิธีการ HUFF สำหรับวิธีการ TLZW ให้ประสิทธิภาพการอัดข้อมูลสูงกว่าวิธีการ LZW เวลาที่ใช้ในการอัดและการขยายข้อมูลจะใกล้เคียงกับวิธีการ LZW และเมื่อเปรียบเทียบระหว่างวิธีการที่ปรับเปลี่ยน วิธีการ TLZW ให้ประสิทธิภาพการอัดโดยเฉลี่ยต่ำกว่าวิธีการ THUFF เล็กน้อย แต่ใช้เวลาในการอัดและการขยายข้อมูลน้อยกว่าวิธีการ THUFF มาก โดยเฉพาะเวลาที่ใช้ในการอัดข้อมูลจะแตกต่างกันอย่างมาก

สำหรับการทดสอบระหว่างวิธีการ THUFF และวิธีการ TLZW ที่ใช้คำไทยเพิ่มขึ้นจาก 255 คำ เป็น 511 คำ ทั้ง 2 วิธีสามารถให้ประสิทธิภาพการอัดที่เพิ่มขึ้นโดยใช้เวลาในการอัดและการขยายข้อมูลของแต่ละวิธีไม่แตกต่างไปจากเดิมเท่าใดนัก

5.4 การทดสอบข้อความภาษาไทยทั่ว ๆ ไป

5.4.1 การทดสอบตามขั้นตอนที่ 5.2.1

ได้ผลการทดสอบประสิทธิภาพการอัด แสดงในตารางที่ 5.5 และแสดงเป็นกราฟดังรูปที่ 5.4 ส่วนผลการทดสอบเวลาที่ใช้ในการอัดข้อมูลและการขยายข้อมูลแสดงในตารางที่ 5.6 และแสดงเป็นกราฟดังรูปที่ 5.5 รูปที่ 5.6 ตามลำดับ

5.4.2 การทดสอบตามขั้นตอนที่ 5.2.2

ได้ผลการทดสอบประสิทธิภาพการอัด แสดงในตารางที่ 5.7 ส่วนผลการทดสอบเวลาที่ใช้ในการอัดข้อมูล และการขยายข้อมูลแสดงในตารางที่ 5.8

วิธีการ THUFF ให้ประสิทธิภาพการอัดข้อมูลสูงกว่าวิธีการ HUFF แต่ต้องใช้เวลาในการอัดข้อมูลมากขึ้นกว่าเดิม ส่วนการขยายข้อมูลใช้ทั้ง 2 วิธีใช้เวลาใกล้เคียงกัน สำหรับวิธีการ TLZW ให้ประสิทธิภาพการอัดข้อมูลสูงกว่าวิธีการ LZW ใช้เวลาในการอัดใกล้เคียงกันแต่ในการขยายข้อมูลใช้เวลาเพิ่มขึ้นเล็กน้อย

สำหรับการทดสอบระหว่างวิธีการ THUFF และวิธีการ TLZW ที่ใช้คำไทยเพิ่มขึ้นจาก 255 คำ เป็น 511 คำ ทั้ง 2 วิธี สามารถให้ประสิทธิภาพการอัดที่เพิ่มขึ้นโดยใช้เวลาในการอัดและการขยายข้อมูลไม่แตกต่างไปจากการใช้จำนวนคำที่น้อยกว่า



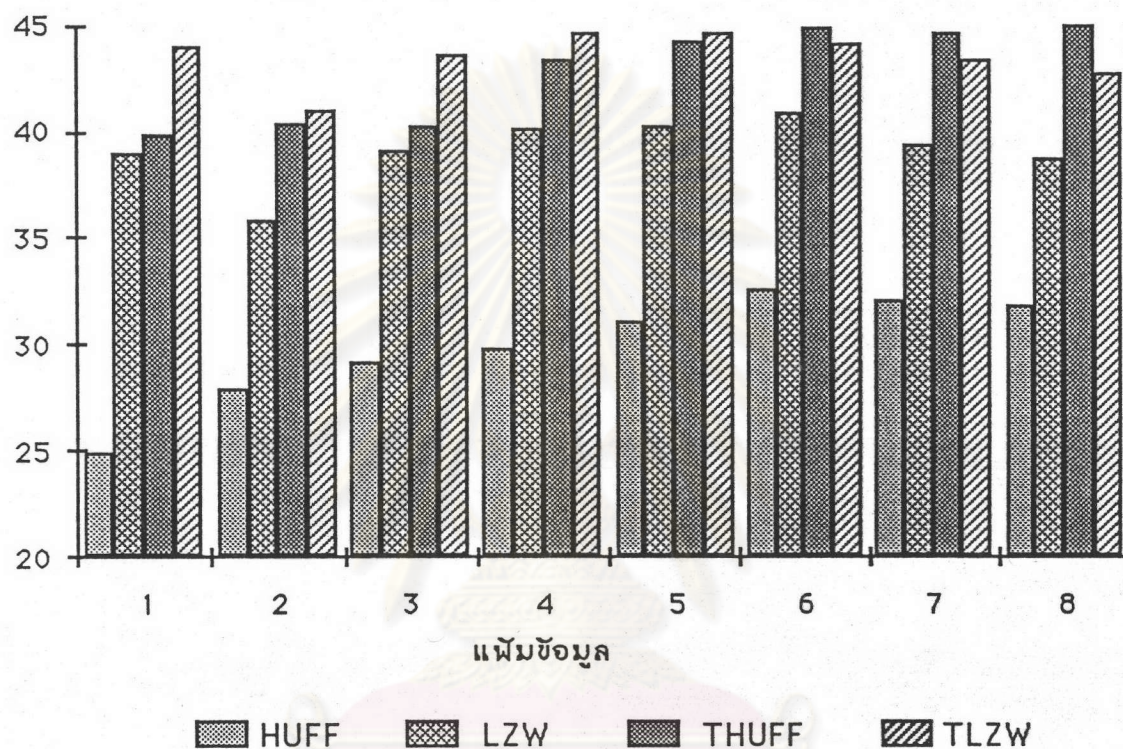
ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 5.1 การทดสอบประสิทธิภาพการอัดข้อมูลภาษาไทยล้วน
เปรียบเทียบระหว่างวิธีการเดิมและวิธีการที่ปรับเปลี่ยนที่ใช้ค่าไทย 255 คำ

แฟ้ม ลำดับ ที่	ขนาด ข้อมูล (ไบต์)	HUFF		LZW		THUFF		TLZW	
		ขนาด ข้อมูล ที่ถูกอัด	ประสิทธิภาพ (%)	ขนาด ข้อมูล ที่ถูกอัด	ประสิทธิภาพ (%)	ขนาด ข้อมูล ที่ถูกอัด	ประสิทธิภาพ (%)	ขนาด ข้อมูล ที่ถูกอัด	ประสิทธิภาพ (%)
1	5961	4477	24.90	3641	38.92	3586	39.84	3337	44.02
2	10569	7624	27.86	6785	35.80	6312	40.28	6242	40.94
3	14214	10074	29.13	8661	39.07	8494	40.24	8014	43.62
4	15697	11012	29.85	9400	40.12	8892	43.35	8706	44.54
5	20112	13880	30.99	12038	40.15	11214	44.24	11157	44.53
6	36452	24604	32.50	21565	40.84	20102	44.85	20396	44.05
7	50319	34191	32.05	30552	39.28	27875	44.60	28497	43.37
8	71582	48844	31.76	43884	38.69	39431	44.91	41018	42.70
ค่าเฉลี่ย	28113	19338	29.88	17066	39.11	15738	42.79	15921	43.47

จุฬาลงกรณ์มหาวิทยาลัย

ประสิทธิภาพ
(%)



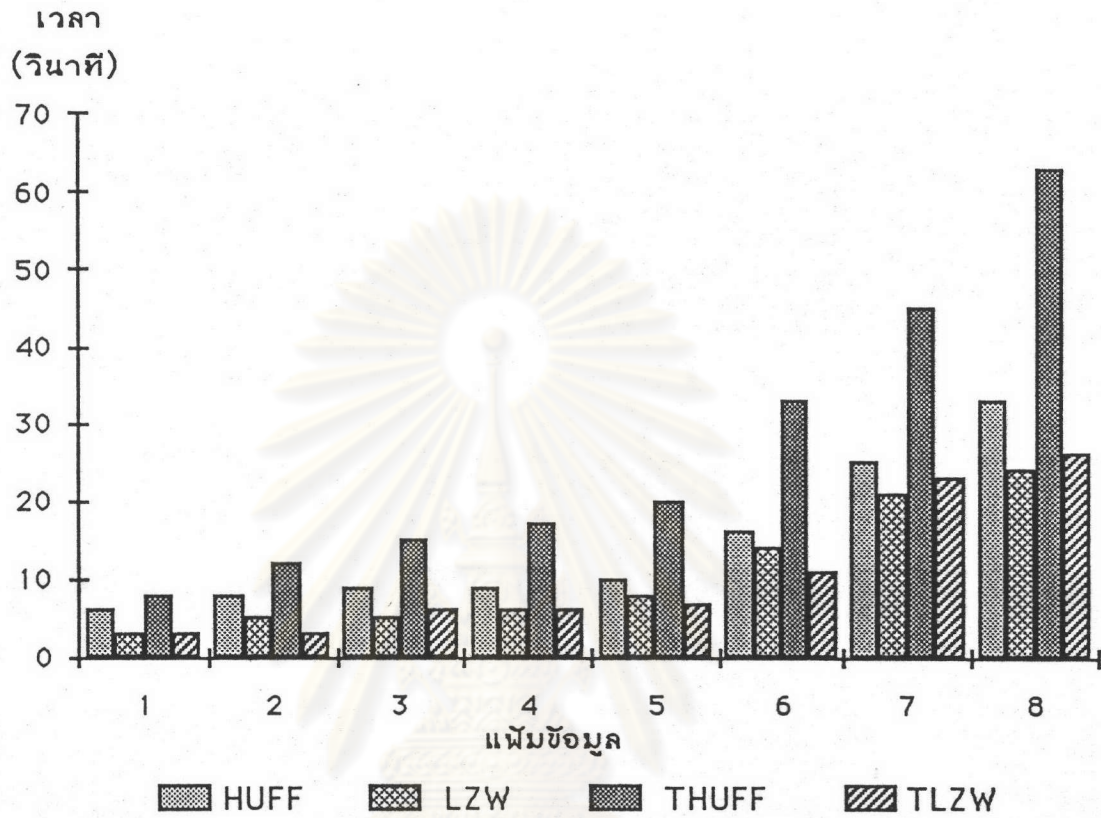
รูปที่ 5.1 กราฟแสดงประสิทธิภาพการอัดข้อมูลภาษาไทยล้วน
เปรียบเทียบระหว่างวิธีการเดิมและวิธีการที่ปรับเปลี่ยน
ที่ใช้คำไทย 255 คำ

จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 5.2 การทดสอบเวลาที่ใช้ในการอัดและการขยายข้อมูลภาษาไทยล้วน
เปรียบเทียบระหว่างวิธีการเดิมและวิธีการที่ปรับเปลี่ยนที่ใช้คำไทย 255 คำ

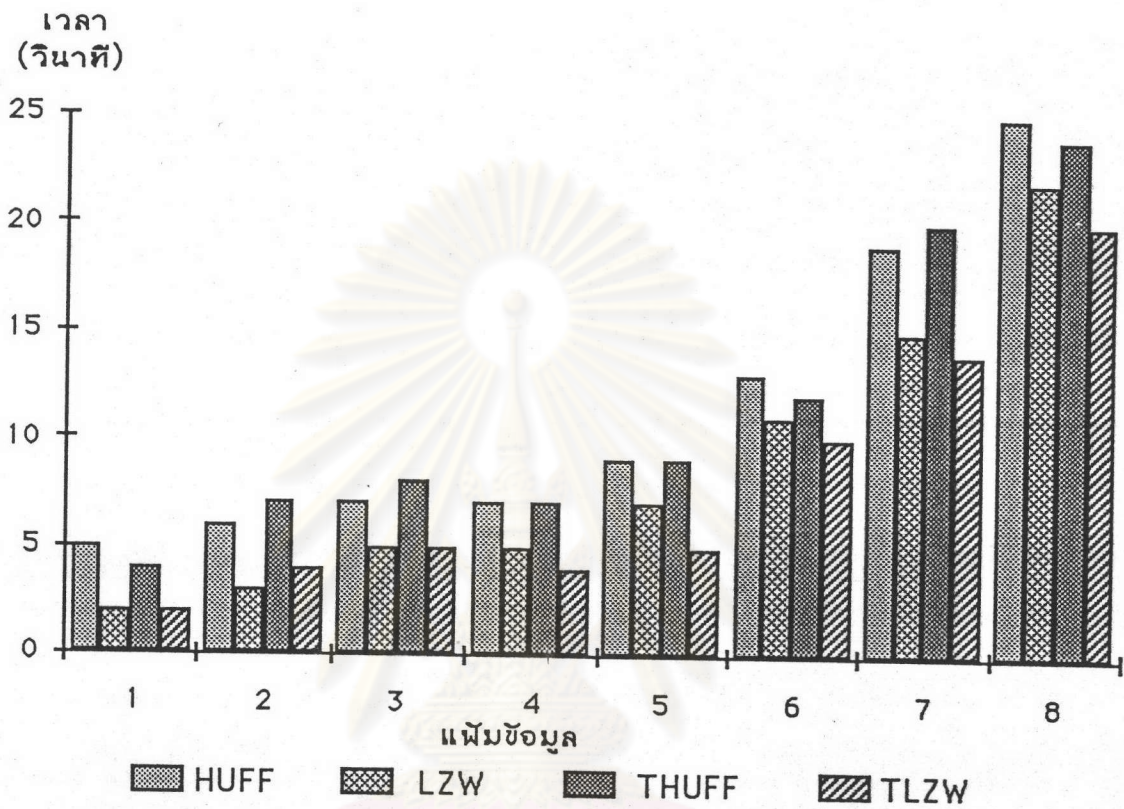
แฟ้ม ลำดับ ที่	ขนาด ข้อมูล (ไบต์)	HUFF		LZW		THUFF		TLZW	
		เวลาที่ ใช้ใน การอัด	เวลาที่ ใช้ใน การขยาย	เวลาที่ ใช้ใน การอัด	เวลาที่ ใช้ใน การขยาย	เวลาที่ ใช้ใน การอัด	เวลาที่ ใช้ใน การขยาย	เวลาที่ ใช้ใน การอัด	เวลาที่ ใช้ใน การขยาย
1	5961	6	5	3	2	8	4	3	2
2	10569	8	6	5	3	12	7	3	4
3	14214	9	7	5	5	15	8	6	5
4	15697	9	7	6	5	17	7	6	4
5	20112	10	9	8	7	20	9	7	5
6	36452	16	13	14	11	33	12	11	10
7	50319	25	19	21	15	45	20	23	14
8	71582	33	25	24	22	63	24	26	20
ค่าเฉลี่ย	28113	15	11	11	9	27	11	11	8

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 5.2 กราฟแสดงเวลาที่ใช้ในการอัดข้อมูลภาษาไทยล้วน
เปรียบเทียบระหว่างวิธีการเดิมและวิธีการที่ปรับเปลี่ยน
ที่ใช้คำไทย 255 คำ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 5.3 กราฟแสดงเวลาที่ใช้ในการขยายข้อมูลภาษาไทยล้วน
เปรียบเทียบระหว่างวิธีการเดิมและวิธีการที่ปรับเปลี่ยน
ที่ค่าไทย 255 ค่า

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 5.3 การทดสอบประสิทธิภาพการอัดข้อมูลภาษาไทยล้วน

เปรียบเทียบระหว่างวิธีการที่ปรับเปลี่ยนที่ใช้คำไทย 255 คำ และ 511 คำ

แฟ้ม ลำดับ ที่	ขนาด ข้อมูล (ไบต์)	คำไทย 255 คำ				คำไทย 511 คำ			
		THUFF		TLZW		THUFF		TLZW	
		ขนาด ข้อมูล ที่ถูกต้อง (%)	ประสิทธิภาพ (%)	ขนาด ข้อมูล ที่ถูกต้อง (%)	ประสิทธิภาพ (%)	ขนาด ข้อมูล ที่ถูกต้อง (%)	ประสิทธิภาพ (%)	ขนาด ข้อมูล ที่ถูกต้อง (%)	ประสิทธิภาพ (%)
1	5961	3586	39.84	3337	44.02	3390	43.13	3233	45.76
2	10569	6312	40.28	6242	40.94	6005	43.18	6106	42.23
3	14214	8494	40.24	8014	43.62	7916	44.31	7693	45.88
4	15697	8892	43.35	8706	44.54	8285	47.22	8369	46.68
5	20112	11214	44.24	11157	44.53	10473	47.93	10970	45.46
6	36452	20102	44.85	20396	44.05	19038	47.77	19942	45.29
7	50319	27875	44.60	28497	43.37	26190	47.95	27776	44.80
8	71582	39431	44.91	41018	42.70	36299	49.29	39632	44.63
ค่าเฉลี่ย	28113	15738	42.79	15921	43.47	14700	46.35	15465	45.09

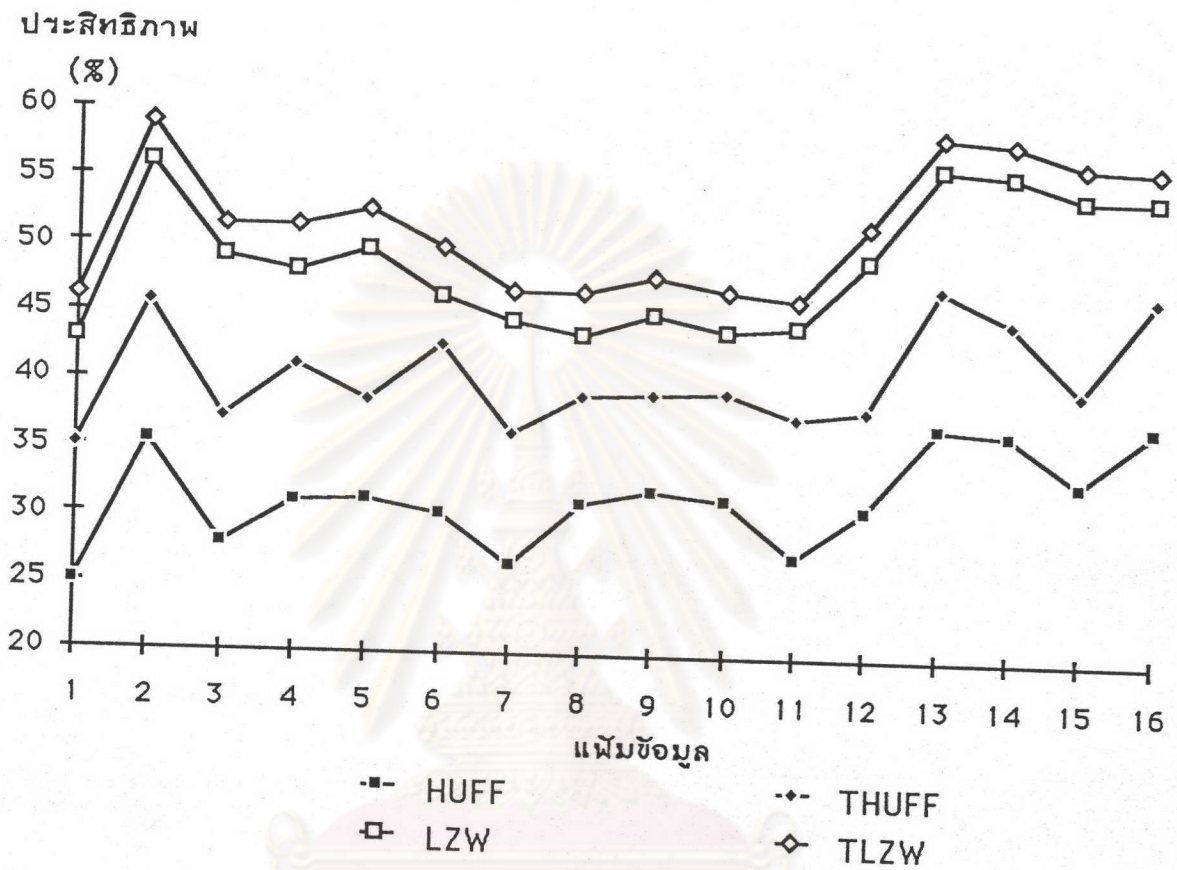
ตารางที่ 5.4 การทดสอบเวลาที่ใช้ในการอัดและการขยายข้อมูลภาษาไทยล้วน
เปรียบเทียบระหว่างวิธีการที่ปรับเปลี่ยนที่ใช้คำไทย 255 คำ และ 511 คำ

แฟ้ม ลำดับ ที่	ขนาด ข้อมูล (ไบต์)	คำไทย 255 คำ				คำไทย 511 คำ			
		THUFF		TLZW		THUFF		TLZW	
		เวลาที่ ใช้ใน การอัด	เวลาที่ ใช้ใน การขยาย	เวลาที่ ใช้ใน การอัด	เวลาที่ ใช้ใน การขยาย	เวลาที่ ใช้ใน การอัด	เวลาที่ ใช้ใน การขยาย	เวลาที่ ใช้ใน การอัด	เวลาที่ ใช้ใน การขยาย
1	5961	8	4	3	2	8	4	2	3
2	10569	12	7	3	4	12	6	5	4
3	14214	15	8	6	5	15	7	6	5
4	15697	17	7	6	4	16	7	7	5
5	20112	20	9	7	5	20	7	9	7
6	36452	33	12	11	10	34	12	15	12
7	50319	45	20	23	14	46	16	21	16
8	71582	63	24	26	20	58	21	25	23
ค่าเฉลี่ย	28113	27	11	11	8	26	10	11	9

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 5.5 การทดสอบประสิทธิภาพการอัดข้อมูลภาษาไทยทั่ว ๆ ไป
เปรียบเทียบระหว่างวิธีการเดิมและวิธีการที่ปรับเปลี่ยนที่ใช้คำไทย 255 คำ

แฟ้มลำดับที่	ขนาดข้อมูล (ไบต์)	HUFF		LZW		THUFF		TLZW	
		ขนาดข้อมูลที่ถูกอัด	ประสิทธิภาพ (%)	ขนาดข้อมูลที่ถูกอัด	ประสิทธิภาพ (%)	ขนาดข้อมูลที่ถูกอัด	ประสิทธิภาพ (%)	ขนาดข้อมูลที่ถูกอัด	ประสิทธิภาพ (%)
1	15201	11403	24.99	8668	42.98	9867	35.09	8197	46.08
2	20825	13441	35.46	9163	56.00	11279	45.84	8547	58.96
3	24690	17795	27.93	12580	49.05	15480	37.30	11972	51.51
4	25483	17541	31.17	13215	48.14	15002	41.13	12366	51.47
5	40447	27809	31.25	20323	49.75	24825	38.62	19173	52.60
6	45976	32078	30.23	24636	46.42	26288	42.82	23010	49.95
7	46368	34079	26.50	25737	44.49	29566	36.24	24716	46.70
8	49381	34024	31.10	27821	43.66	30100	39.05	26274	46.79
9	55182	37460	32.12	30320	45.05	33563	39.18	28788	47.83
10	64036	43905	31.44	35884	43.96	38854	39.32	33934	47.01
11	70522	51308	27.25	39239	44.36	43989	37.62	37791	46.41
12	85988	59411	30.91	43537	49.37	53173	38.16	41408	51.84
13	154731	97506	36.98	67811	56.17	81513	47.32	64093	58.58
14	156023	98731	36.72	68779	55.92	85803	45.01	65239	58.19
15	205850	137636	33.14	94308	54.19	123780	39.87	89350	56.59
16	264654	166357	37.14	121283	54.17	140280	46.99	115095	56.51
ค่าเฉลี่ย	82835	55030	31.52	40207	48.98	47710	40.60	38122	51.69

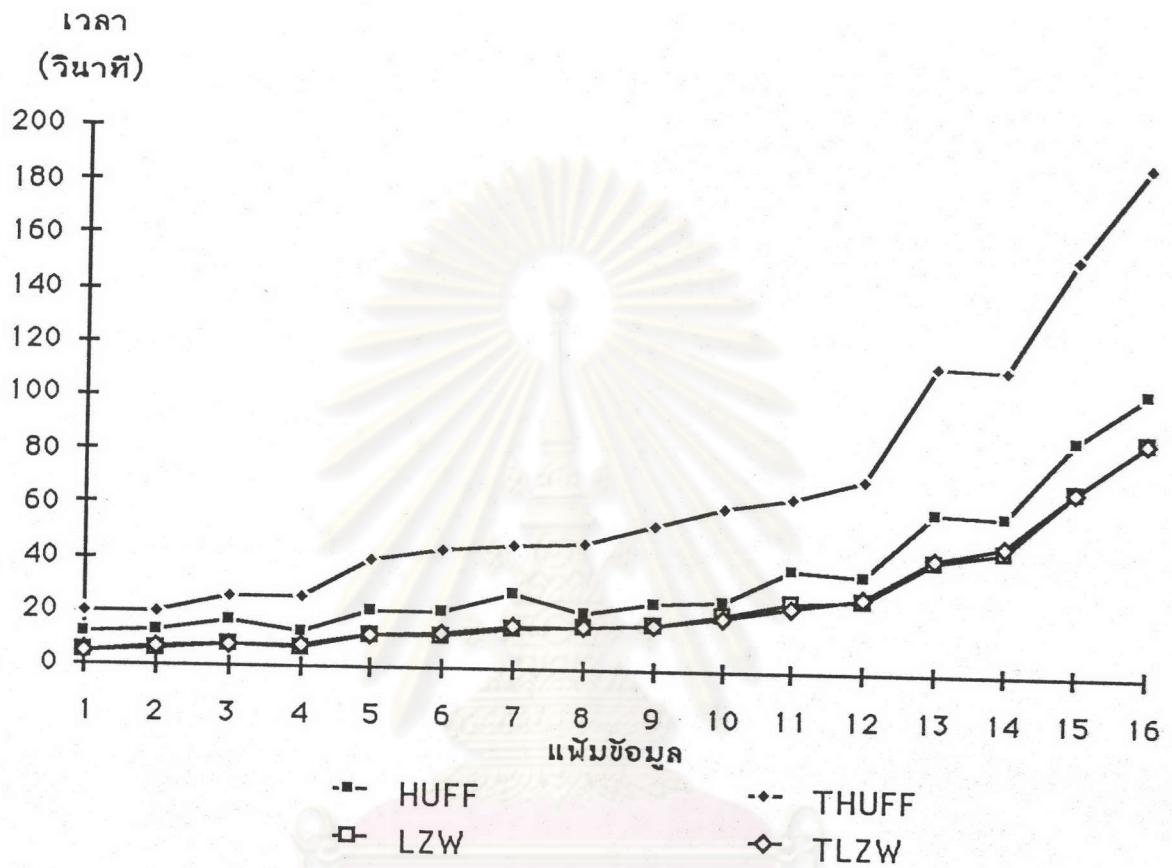


รูปที่ 5.4 กราฟแสดงประสิทธิภาพการอัดข้อมูลภาษาไทยทั่ว ๆ ไป
เปรียบเทียบระหว่างวิธีการเดิมและวิธีการที่ปรับเปลี่ยน
ที่ ๒๕ ธันวาคม ๒๕๕๕

ศูนย์บริการ
จุฬาลงกรณ์มหาวิทยาลัย

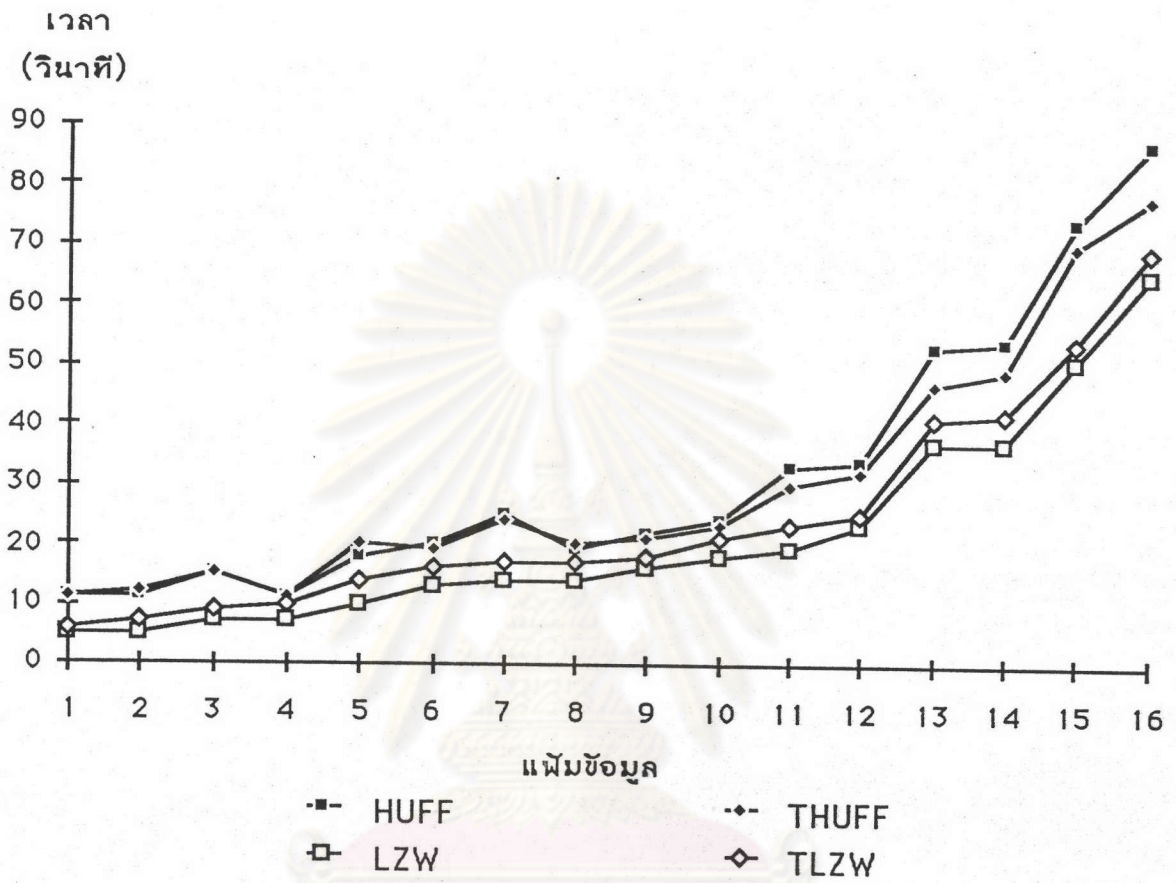
ตารางที่ 5.6 การทดสอบเวลาที่ใช้ในการจัดและการขยายข้อมูลภาษาไทยทั่วไป
เปรียบเทียบระหว่างวิธีการเดิมและวิธีการที่ปรับเปลี่ยนที่ใช้ค่าไทย 255 คำ

แฟ้ม ลำดับ ที่	ขนาด ข้อมูล (ไบต์)	HUFF		LZW		THUFF		TLZW	
		เวลาที่ ใช้ใน การจัด	เวลาที่ ใช้ใน การขยาย	เวลาที่ ใช้ใน การจัด	เวลาที่ ใช้ใน การขยาย	เวลาที่ ใช้ใน การจัด	เวลาที่ ใช้ใน การขยาย	เวลาที่ ใช้ใน การจัด	เวลาที่ ใช้ใน การขยาย
1	15201	12	11	5	5	20	11	5	6
2	20825	13	11	6	5	20	12	7	7
3	24690	17	15	8	7	26	15	8	9
4	25483	13	11	7	7	26	11	8	10
5	40447	21	18	12	10	40	20	12	14
6	45976	21	20	12	13	44	19	13	16
7	46368	28	25	15	14	46	24	16	17
8	49381	21	19	16	14	47	20	16	17
9	55182	25	22	17	16	53	21	17	18
10	64036	26	24	21	18	60	23	20	21
11	70522	38	33	26	19	64	30	24	23
12	85988	36	34	27	23	71	32	28	25
13	154731	59	53	42	37	114	47	43	41
14	156023	58	54	46	37	113	49	48	42
15	205850	87	74	68	51	154	70	68	54
16	264654	105	87	87	65	189	78	87	69
ค่าเฉลี่ย	82835	36	32	26	21	68	30	26	24



รูปที่ 5.5 กราฟแสดงเวลาที่ใช้ในการอัดข้อมูลภาษาไทยทั่ว ๆ ไป
 เปรียบเทียบระหว่างวิธีการเดิมและวิธีการที่ปรับเปลี่ยน
 ที่ใช้คำไทย 255 คำ

จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 5.6 กราฟแสดงเวลาที่ใช้ในการขยายข้อมูลภาษาไทยทั่ว ๆ ไป
เปรียบเทียบระหว่างวิธีการเดิมและวิธีการที่ปรับเปลี่ยน
ที่ใช้ค่าไทย 255 คำ

จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 5.7 การทดสอบประสิทธิภาพการอัดข้อมูลภาษาไทยทั่ว ๆ ไป
เปรียบเทียบระหว่างวิธีการที่ปรับเปลี่ยนที่ใช้คำไทย 255 คำ และ 511 คำ

แฟ้ม ลำดับ ที่	ขนาด ข้อมูล (ไบต์)	คำไทย 255 คำ				คำไทย 511 คำ			
		THUFF		TLZW		THUFF		TLZW	
		ขนาด ข้อมูล ที่ถูกต้อง	ประสิทธิภาพ (%)	ขนาด ข้อมูล ที่ถูกต้อง	ประสิทธิภาพ (%)	ขนาด ข้อมูล ที่ถูกต้อง	ประสิทธิภาพ (%)	ขนาด ข้อมูล ที่ถูกต้อง	ประสิทธิภาพ (%)
1	15201	9867	35.09	8197	46.08	9000	40.79	8007	47.33
2	20825	11279	45.84	8547	58.96	11008	47.14	8310	60.10
3	24690	15480	37.30	11972	51.51	14105	42.87	11738	52.46
4	25483	15002	41.13	12366	51.47	13220	48.12	12035	52.77
5	40447	24825	38.62	19173	52.60	22069	45.44	18823	53.46
6	45976	26288	42.82	23010	49.95	23205	49.53	22462	51.14
7	46368	29566	36.24	24716	46.70	27129	41.49	24047	48.14
8	49381	30100	39.05	26274	46.79	26774	45.78	25428	48.51
9	55182	33563	39.18	28788	47.83	30018	45.60	27986	49.28
10	64036	38854	39.32	33934	47.01	34476	46.16	32952	48.54
11	70522	43989	37.62	37791	46.41	39773	43.60	37001	47.53
12	85988	53173	38.16	41408	51.84	47240	45.06	40974	52.35
13	154731	81513	47.32	64093	58.58	77538	49.89	62257	59.76
14	156023	85803	45.01	65239	58.19	83341	46.58	63909	59.04
15	205850	123780	39.87	89350	56.59	109405	46.85	87359	57.56
16	264654	140280	46.99	115095	56.51	134656	49.12	111764	57.77
ค่าเฉลี่ย	82835	47710	40.60	38122	51.69	43935	45.88	37191	52.86

ตารางที่ 5.8 การทดสอบเวลาที่ใช้ในการอัดและการขยายข้อมูลภาษาไทยทั่วไป
เปรียบเทียบระหว่างวิธีการที่ปรับเปลี่ยนที่ใช้คำไทย 255 คำ และ 511 คำ

แฟ้ม ลำดับ ที่	ขนาด ข้อมูล (ไบต์)	คำไทย 255 คำ				คำไทย 511 คำ			
		THUFF		TLZW		THUFF		TLZW	
		เวลาที่ ใช้ใน การอัด	เวลาที่ ใช้ใน การขยาย	เวลาที่ ใช้ใน การอัด	เวลาที่ ใช้ใน การขยาย	เวลาที่ ใช้ใน การอัด	เวลาที่ ใช้ใน การขยาย	เวลาที่ ใช้ใน การอัด	เวลาที่ ใช้ใน การขยาย
1	15201	20	11	5	6	18	11	6	5
2	20825	20	12	7	7	18	12	7	6
3	24690	26	15	8	9	24	16	9	7
4	25483	26	11	8	10	23	12	10	8
5	40447	40	20	12	14	37	20	15	12
6	45976	44	19	13	16	40	18	17	14
7	46368	46	24	16	17	42	23	15	14
8	49381	47	20	16	17	42	18	18	15
9	55182	53	21	17	18	49	19	22	16
10	64036	60	23	20	21	56	20	24	19
11	70522	64	30	24	23	61	28	23	22
12	85988	71	32	28	25	65	28	27	24
13	154731	114	47	43	41	115	43	48	39
14	156023	113	49	48	42	105	46	44	41
15	205850	154	70	68	54	146	60	59	55
16	264654	189	78	87	69	180	72	73	72
ค่าเฉลี่ย	82835	68	30	26	24	64	28	26	23