



## บทที่ 4

### วิธีการทดลอง ผลการทดลอง และวิเคราะห์ผลการทดลอง

การเลือกข้อมูลเพื่อใช้ในการทดลอง มีผลอย่างมากต่อการทดลอง เมื่อนำข้อมูลมาป้อนให้ระบบทำงาน แล้วพบว่าในการทดลองนั้นไม่ได้ผลดีเท่าที่ควร อาจจะนำข้อมูลที่ไม่ได้ผลตามต้องการออกจากการทดลอง หรือทำการปรับปรุงแก้ไขโปรแกรมเพื่อให้ได้ผลลัพธ์ที่ต้องการ และมีประสิทธิภาพ แต่ก็อาจส่งผลกระทบต่อข้อมูลชุดอื่นๆ ที่นำเข้ามาทดลองได้ผลลดลงก็ได้ ดังนั้นจึงเป็นเรื่องยากที่จะหาข้อยุติว่าการพัฒนาระบบทดลองข้อมูลควรมีหลักเกณฑ์อย่างไร

สำหรับวิทยานิพนธ์นี้รวบรวมข้อมูลภาพอักขรที่ติดกันในแต่ละแบบ เพื่อทำการทดสอบประสิทธิภาพในการตัดตัวอักขรในรูปแบบการติดกันในลักษณะต่างๆ

#### วิธีการทดลอง

แบบตัวอักษรไทยที่นำมาใช้ทดสอบคือ AngsanaUPC, CordiaUPC, BrowalliaUPC, DilleniaUPC, EucrosiaUPC, FreesiaUPC, IrisUPC, JasmineUPC, KodchiangUPC, LilyUPC จะทดสอบที่ขนาด 12 14 16 18 20 22 24 26 28 36 point โดยพิมพ์ข้อความที่ใช้ในการทดลองด้วยโปรแกรม Microsoft Word แล้วพิมพ์โดยใช้เครื่องพิมพ์เลเซอร์ที่มีความละเอียด 300 จุดต่อนิ้ว (ภาพข้อมูลที่ใช้ทดลอง ได้แสดงดังรูป ก.11 ก.12 ก.13 ก.14 ก.15 ก.16 ก.17 ก.18 ก.19 ก.20 ในภาคผนวก)

แบบตัวอักษรไทยได้มาจากการการสแกนด้วยสแกนเนอร์ความละเอียด 300 จุดต่อนิ้ว และเป็นขาวดำ (Binary Image) จัดเก็บเป็น Bitmap (BMP)

ทำการอ่านภาพจากไฟล์ภาพที่สแกน แล้วทำการตัดแยกภาพตัวอักษรให้เป็นตัวอักษรเดี่ยวต่อไป

การตรวจสอบความถูกต้องใช้เกณฑ์ในการนับดังนี้

- ตัดอักษรเดี่ยวผิดเป็น 2 อักษรนับผิด 1 ที่
- ไม่ตัดอักษรที่ติด นับเท่าจำนวนอักษรที่ติด

ตารางที่ 4-1 แสดงผลการทดสอบการตัดตัวอักษรจำนวน 1,507 ตัวอักษร

แบบอักษร	ขนาดตัวอักษร	จำนวนตัวอักษร ที่ติดกัน	จำนวนที่ติดผิด	% ความถูกต้อง	เวลาที่ใช้
AngsanaUPC	12	46	18	60.87	21:56
	14	42	18	57.14	29:64
	16	46	16	65.22	37:23
	18	32	10	68.75	45:90
	20	20	4	80.00	56:80
	22	24	8	66.67	1:07:62
	24	22	10	54.55	1:08:04
	26	20	8	60.00	1:33:55
	28	22	10	54.55	1:45:92
	36	14	6	57.14	2:52:84
BrowalliaUPC	12	105	40	61.90	22:95
	14	90	20	77.78	29:34
	16	50	12	76.00	35:86
	18	56	12	78.57	45:23
	20	66	24	63.64	59:81
	22	58	18	68.97	1:10:31
	24	44	14	68.18	1:22:84
	26	48	14	70.83	1:35:98
	28	56	16	71.43	1:55:08
	36	36	14	61.11	3:03:62
CordiaUPC	12	108	44	59.26	23:31
	14	108	44	59.26	30:90
	16	82	20	75.61	39:90
	18	86	34	60.47	49:04
	20	74	12	83.78	59:18
	22	72	18	75.00	1:07:04
	24	64	14	78.13	1:25:21
	26	60	12	80.00	1:37:23
	28	58	14	75.86	1:52:88

แบบอักษร	ขนาดตัวอักษร	จำนวนตัวอักษร ที่ติดกัน	จำนวนที่ติดผิด	% ความถูกต้อง	เวลาที่ใช้
CordiaUPC	36	48	18	62.50	3:01:51
DilleniaUPC	12	116	60	48.28	21:01
	14	104	44	57.69	27:93
	16	99	50	49.49	28:17
	18	103	44	57.28	35:10
	20	103	54	47.57	42:54
	22	101	44	56.44	49:45
	24	48	20	58.33	53:02
	26	72	32	55.56	1:30:05
	28	48	14	70.83	1:41:80
	36	42	20	52.38	2:42:41
EucrosiaUPC	12	90	32	64.44	22:56
	14	58	20	65.52	28:62
	16	58	16	72.41	35:67
	18	54	18	66.67	45:17
	20	38	4	89.47	53:81
	22	40	6	85.00	1:04:76
	24	38	4	89.47	1:22:56
	26	38	4	89.47	1:34:22
	28	36	4	88.89	1:46:80
	36	36	4	88.89	2:52:19
FreesiaUPC	12	109	42	61.47	23:36
	14	96	40	58.33	30:06
	16	61	26	57.38	38:15
	18	72	32	55.56	45:70
	20	63	24	61.90	55:82
	22	59	22	62.71	1:09:50
	24	40	6	85.00	1:22:10
	26	44	8	81.82	1:34:84
	28	46	6	86.96	1:49:96
	36	40	6	85.00	2:58:19

แบบอักษร	ขนาดตัวอักษร	จำนวนตัวอักษร ที่ติดกัน	จำนวนที่ติดผิด	% ความถูกต้อง	เวลาที่ใช้
IrisUPC	12	112	42	62.50	22:02
	14	104	30	71.15	27:02
	16	88	34	61.36	36:56
	18	96	30	68.75	42:95
	20	86	28	67.44	52:51
	22	76	26	65.79	1:05:40
	24	74	32	56.76	1:17:82
	26	80	24	70.00	1:27:82
	28	74	22	70.27	1:40:65
	36	72	22	69.44	2:42:01
JasmineUPC	12	124	76	38.71	21:89
	14	108	62	42.59	27:80
	16	114	62	45.61	36:10
	18	60	30	50.00	43:56
	20	56	30	46.43	52:76
	22	48	24	50.00	1:04:90
	24	52	28	46.15	1:13:04
	26	40	18	55.00	1:24:14
	28	40	20	50.00	1:41:49
	36	40	20	50.00	2:49:66
KodchiangUPC	12	136	88	35.29	19:21
	14	64	34	46.88	24:65
	16	56	32	42.86	29:62
	18	50	28	44.00	37:02
	20	50	30	40.00	43:58
	22	46	24	47.83	53:64
	24	48	22	54.17	1:02:09
	26	50	24	52.00	1:11:07
	28	46	24	47.83	1:22:84
	36	40	20	50.00	2:18:46
LilyUPC	ไม่สามารถทดสอบได้เนื่องจากตัวอักษรหนา และติดกันมาก				

## การวิเคราะห์ผลการทดลอง

จำนวนตัวอักษรที่ทำการทดสอบทั้งหมด 13,563 ตัวอักษร ประกอบด้วยตัวอักษรที่ติดกันจำนวน 5,739 ตัวอักษร ความถูกต้องในการตัดแยกตัดอักษรที่ติดกัน 63.14% ใช้เวลาในการตัดแยกตัวอักษรทั้งหมดโดยเฉลี่ย 22.07 ตัวอักษรต่อวินาที จากตารางที่ 4-1 จะเห็นได้ว่า อัตราความผิดพลาดที่ตัวอักษรขนาด 12 point สูงกว่าขนาดอื่นๆ เนื่องจากอักษรมีขนาดเล็กทำให้จุดใกล้กันมาก เมื่อตัวอักษรมีขนาดตั้งแต่ 14 point ขึ้นไปความผิดพลาดเริ่มลดลง ได้ผลความถูกต้องสูงขึ้น และพบว่าตัวอักษรแบบ AngsanaUPC นั้นมีการติดกันน้อยกว่าตัวอักษรในแบบอื่นๆ ส่วนแบบตัวอักษรที่มีการติดกันของตัวอักษรมากได้แก่ ตัวอักษรในแบบ KodchiangUPC ตัวอักษรแบบ JasmineUPC นั้นไม่มีหัวและมีลักษณะตัวอักษรอ่อนจะทำให้พิจารณาจุดตัดได้ยาก กรณีที่ติดกันในแนวดิ่ง ตัวอักษรแบบ DilleniaUPC จะมีลักษณะตัวอักษรผอม และมีการติดกันในแนวนอนมากกว่าแบบอื่นๆ ตัวอักษรแบบ LilyUPC จะเป็นตัวอักษรที่มีความกว้างมาก และตัวอักษรหนาเป็นผลให้ติดกันต่อเนื่องมาก ทำให้ไม่สามารถตัดแยกได้ ดังตัวอย่าง รูปที่ 4-1

สำหรับเวลาในการวิเคราะห์และตัดแยกตัวอักษรจะเพิ่มขึ้นตามขนาดของตัวอักษร เนื่องจากถ้าหน้ากระดาษขนาดเท่ากันตัวอักษรขนาดเล็กจะมีจำนวนตัวอักษรมากกว่า แต่ใช้เวลาในการหาเส้นตัดคอลลัมน์บรรทัดใกล้เคียงกันเพราะต้องนับจุดตามแนวดิ่ง และแนวนอนเท่ากัน ดังนั้นตัวอักษรขนาดใหญ่จะมีจำนวนแผ่นมากกว่า จึงใช้เวลานานกว่า

AngsanaUPC สติปัญญา ได้ใช้วิตามินซี วิริยะ  
 BrowalliaUPC สติปัญญา ได้ใช้วิตามินซี วิริยะ  
 CordiaUPC สติปัญญา ได้ใช้วิตามินซี วิริยะ  
 DilleniaUPC สติปัญญา ได้ใช้วิตามินซี  
 EucrosiaUPC สติปัญญา ได้ใช้วิตามินซี วิริยะ  
 FreesiaUPC สติปัญญา ได้ใช้วิตามินซี วิริยะ  
 IrisUPC สติปัญญา ได้ใช้วิตามินซี วิริยะ  
 JasmineUPC สติปัญญา ได้ใช้วิตามินซี วิริยะ  
 KodchiangUPC สติปัญญา ได้ใช้วิตามินซี วิริยะ  
 LilyUPC สติปัญญา ได้ใช้วิตามินซี วิริยะ

รูปที่ 4-1 แสดงแบบตัวอักษรต่างๆ

จากการทดลองพบว่าความผิดพลาดส่วนใหญ่มาจากส่วนตัดบรรทัดผิด อักขรติด และวิเคราะห์รูปแบบการติดกันของตัวอักษรผิดพลาด

เราสามารถวิเคราะห์ความผิดพลาดโดยสามารถแยกตามการเกิดได้ดังนี้

1. ความผิดพลาดจากการตัดแบ่งบรรทัดผิด
2. วิเคราะห์รูปแบบการติดกันของตัวอักษรผิดพลาด
3. การตัดผิดตำแหน่ง
4. การแยกตัวอักษรที่ไขว้กัน หรือตัดแยกตามเส้นตัวอักษร

#### ความผิดพลาดจากการตัดบรรทัด

สาเหตุสำคัญของความผิดพลาดที่เกิดกับตัวอักษรระดับบน และ ล่างซึ่งถ้าตัวอักษรที่มีขนาดเล็กจะทำให้สระล่างลงมามีเส้นแนวระดับเท่ากับสระบน ของตัวอักษรในบรรทัดล่าง เป็นผลทำให้ส่วนตัดบรรทัดทำงานผิดพลาด เมื่อตัวอักษรที่มีขนาดใหญ่ขึ้น ส่วนตัดบรรทัดจะเข้าใจผิดว่าช่องระหว่างตัวอักษรบนกับตัวอักษรกลางหรือ ตัวอักษรกลางกับ ตัวอักษรล่างเป็นเส้นแบ่งระหว่างบรรทัด จึงทำให้ตัดบรรทัดผิดพลาดได้

#### ความผิดพลาดจากการวิเคราะห์แบบตัวอักษรที่ติดกันผิด และการตัดผิดตำแหน่ง

สาเหตุในการวิเคราะห์แบบตัวอักษรที่ติดกันผิดพลาด เนื่องมาจาก ตัวอักษรที่ติดกันบางแบบมีลักษณะที่ใกล้เคียงกัน อีกทั้งแบบตัวอักษรมีหลายแบบ ลักษณะที่ติดกันของแบบตัวอักษรแบบหนึ่งอาจใกล้เคียงกับการติดกันของตัวอักษรในอีกลักษณะหนึ่งของแบบตัวอักษรอีกแบบหนึ่ง ทำให้ส่วนตัดแยก ทำการตัดผิดตำแหน่งไปด้วยเนื่องจากรูปแบบการติดกันแต่ละแบบจะใช้วิธีการตัดที่ไม่เหมือนกัน หรือกรณีที่ตัวอักษรที่มีความกว้างมากๆ ทำให้ส่วนตัดแยกเข้าใจว่าเป็นตัวอักษรที่ติดกันก็จะทำการตัดแยกผิดพลาด เช่น ณ ญ เป็นต้น หรือส่วนตัดแยกพบว่ามีตัวอักษรที่ติดกัน แล้วมีความกว้างน้อย ทำให้ไม่ทำการตัดแยก เช่น เร หรือในกรณีตัวอักษรแบบ JasmineUPC ที่ไม่มีหัว ตัวอักษรบางตัวสั้นมาก และการติดกันทำให้ส่วนวิเคราะห์เข้าใจว่าเป็นตัวเดียว เช่น ที่ จึงไม่ตัดไม้เอกออกเนื่องจากเข้าใจว่าเป็นสระอีเท่านั้น

### ความผิดพลาดจากการแยกตัวอักษรที่ไขว้กัน

การตัดแยกตัวอักษรที่ติดกันในลักษณะที่ไขว้กัน ส่วนตัดแยกในลักษณะนี้ไม่สามารถทำงานได้ดีกับตัวอักษรบางประเภทเช่น DilleniaUPC เนื่องจากตัวอักษรมีลักษณะที่ผอมและไขว้กันมากกว่าแบบอื่นๆ ทำการการตัดแยกผิดพลาด เช่น ร้ไ