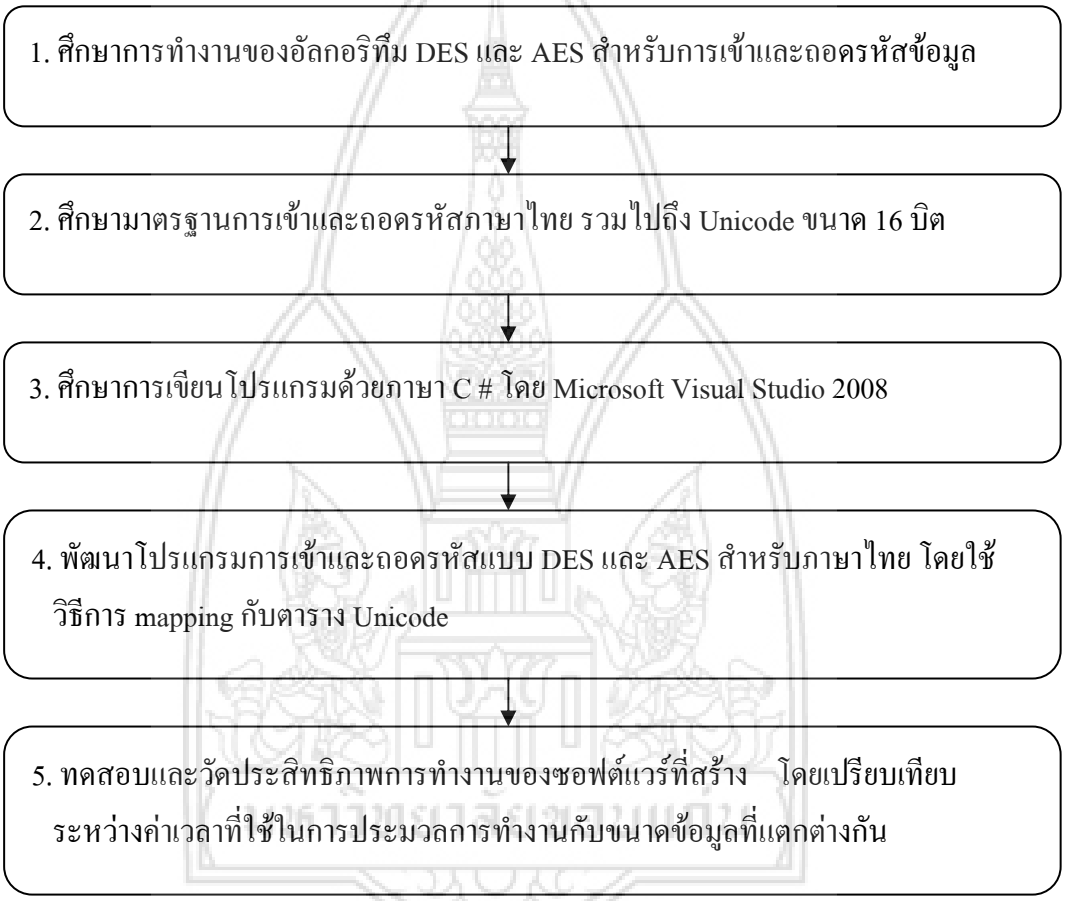
บทที่ 3
วิธีดำเนินการวิจัย

บทนี้เป็นการนำเสนอการออกแบบและวิธีการสร้างแบบการจำลองการเข้ารหัสและถอดรหัส โดยอธิบายวิธีการวิจัยและพัฒนาโปรแกรมประยุกต์ในส่วนต่างๆ



ภาพที่ 27 ขั้นตอนการวิธีการวิจัยและพัฒนาโปรแกรมประยุกต์

3.1 มาตรฐานการใช้อักษรไทย

การพัฒนาระบบซอฟต์แวร์นั้นมีความจำเป็นอย่างยิ่ง ในการศึกษาถึงมาตรฐานที่เกี่ยวข้องกับระบบงานสำหรับการใช้ภาษาไทย ในการพัฒนาด้านสารสนเทศในประเทศไทยและประเทศเพื่อนบ้านนั้นประเทศไทยถูกจัดให้อยู่ในกลุ่มอินโดจีน ได้แก่ภาษาพม่า ไทย ลาว

และกัมพูชา ซึ่งมีอักขรวิธีคล้ายของไทย และอักขรวิธีของกลุ่มนี้ตรงกับวิธีที่ชาวศรีลังกาใช้เขียนภาษาสิงหลและทมิฬด้วย

อักขรวิธี คือ วิธีการเขียนข้อความ ซึ่งการนำข้อความเหล่านั้นนำเข้ามาประมวลผลและส่งออกนั้นจะมีรูปแบบที่แตกต่างกันแต่ละกลุ่มประเทศจะอาศัยโปรแกรมคอมพิวเตอร์ร่วมกันได้ในกลุ่มเดียวกัน การศึกษาเกี่ยวกับภาษาธรรมชาติจะเป็นรากฐานสำคัญในการจัดทำระบบซอฟต์แวร์สากล (Software internationalization)

การเข้ารหัสอักษรเพื่อสนับสนุนการแสดงภาษาไทยมีอยู่หลายมาตรฐาน เช่น TIS -620 หรือ มอก.620 ซึ่งเป็นมาตรฐานของรหัสตัวอักษร ซึ่งกำหนดโดยสำนักงานมาตรฐานอุตสาหกรรม หรือ สมอ. (TISI: Thai Industrial Standard Institute) Windows-874 เป็นมาตรฐานที่ถูกพัฒนาขึ้นโดยบริษัทไมโครซอฟต์ ใช้สำหรับการแสดงผลภาษาไทยบนระบบปฏิบัติการวินโดวส์ และ Unicode เป็นมาตรฐานการเข้ารหัสที่รองรับการเก็บอักษรทุกภาษาทั่วโลกได้โดยอาศัยรหัสเพียงชุดเดียวเพราะเป็นรหัสอักษรแบบ 16 บิต [10]

ในการจัดเก็บข้อมูลบนระบบคอมพิวเตอร์นั้นมีกรรมวิธีในการเก็บและใช้ข้อมูลในรูปของศูนย์และหนึ่ง โดยลักษณะดังกล่าวเรียกว่า “Binary System” ค่าศูนย์และหนึ่ง เรียกว่า “Bit” (Binary Digit) ซึ่งจัดว่าเป็นหน่วยของข้อมูลขนาดเล็กที่สุด ที่จะถูกจัดเก็บภายในระบบคอมพิวเตอร์และใช้เลขฐานสองในการแทนค่าข้อมูลเช่นตัวอักษรต่างๆ เมื่อข้อมูลมีความหลากหลายมากขึ้นจึงได้มีการจัดเก็บในรูปแบบเลข ฐานแปด ฐานสิบ และฐานสิบหก เพื่อสะดวกในการคำนวณและการติดต่อสื่อสารระหว่างบุคคลและเครื่องคอมพิวเตอร์ การใช้รหัสแทนตัวอักษรมีหลายมาตรฐาน เช่น ECB , EBCDI, ASCII และ Unicode เป็นต้น ในการจัดเก็บข้อมูลแบบ ASCII นั้นจะต้องทำการจัดเก็บแบบ 8 bits เป็น 1 byte ซึ่งมีความสามารถในการจัดเก็บได้ เพียง 256 อักขระ หากทำการจัดเก็บข้อมูลแบบ Unicode นั้นจะทำการจัดเก็บ 16 bits เป็น 2 bytes ต่อ หนึ่งตัวอักขระ จึงมีความสามารถในการจัดเก็บได้ ถึง 65536 อักขระ ซึ่งค่ามากพอสำหรับการจัดเก็บภาษาได้ทั่วโลก

รหัสมาตรฐาน Unicode [9] ในรูปแบบของเลขฐานสิบหก ตั้งแต่ 0E01-0E5A ซึ่งมีขนาดจัดเก็บ 2 ไบท์ ถูกจัดเป็นค่าสำหรับการแสดงผลภาษาไทย โดยมีการจัดแบ่งเทียบ ทั้งรหัส ACSII ฐานสิบตั้งแต่ 161-250 ซึ่งมีขนาดจัดเก็บ 1 ไบท์ และ Unicode ฐานสิบ ตั้งแต่ 3585 - 3674 ดังตารางที่ 15 ซึ่งเป็นตารางแสดงค่า Unicode สำหรับแสดงผลภาษาไทย โดยมีตารางที่ 14 เป็นตารางอธิบายการอ่านค่าในตารางที่ 15

ตารางที่ 14 อธิบายแถวในการอ่านค่าในตารางที่ 15 สำหรับการแสดงผลภาษาไทย

ตัวเลขแถวที่ 1	แทน อักขระภาษาไทย	ก
ตัวเลขแถวที่ 2	แทน รหัส ASCII : ฐานสิบ (Dec)	161
ตัวเลขแถวที่ 3	แทน รหัส Unicode : ฐานสิบ (Dec)	3585
ตัวเลขแถวที่ 4	แทน รหัส Unicode : ฐานสิบหก (Hex)	E01

ตารางที่ 15 แสดงค่า Unicode สำหรับการแสดงผลภาษาไทย

ก 161 3585 E01	ข 162 3586 E02	ฃ 163 3587 E03	ค 164 3588 E04	ฅ 165 3589 E05	ฆ 166 3590 E06	ง 167 3591 E07	จ 168 3592 E08	ฉ 169 3593 E09	ช 170 3594 E0A
ซ 171 3595 E0B	ฌ 172 3596 E0C	ญ 173 3597 E0D	ฎ 174 3598 E0E	ฏ 175 3599 E0F	ฐ 176 3600 E10	ฑ 177 3601 E11	ฒ 178 3602 E12	ณ 179 3603 E13	ด 180 3604 E14
ต 181 3605 E15	ถ 182 3606 E16	ท 183 3607 E17	ธ 184 3608 E18	น 185 3609 E19	บ 186 3610 E1A	ป 187 3611 E1B	ผ 188 3612 E1C	ฝ 189 3613 E1D	พ 190 3614 E1E
ฟ 191 3615 E1F	ภ 192 3616 E20	ม 193 3617 E21	ย 194 3618 E22	ร 195 3619 E23	ฤ 196 3620 E24	ล 197 3621 E25	ฦ 198 3622 E26	ว 199 3623 E27	ศ 200 3624 E28
ษ 201 3625 E29	ส 202 3626 E2A	ห 203 3627 E2B	ฬ 204 3628 E2C	อ 205 3629 E2D	ฮ 206 3630 E2E	ฯ 207 3631 E2F	ะ 208 3632 E30	◌ั 209 3633 E31	า 210 3634 E32
◌ำ 211 3635 E33	◌ิ 212 3636 E34	◌ี 213 3637 E35	◌ึ 214 3638 E36	◌ื 215 3639 E37	◌ุ 216 3640 E38	◌ู 217 3641 E39	◌ฺ 218 3642 E3A	◌ 219 3643 E3B	◌ 220 3644 E3C
◌ 221 3645 E3D	◌ 222 3646 E3E	฿ 223 3647 E3F	เ 224 3648 E40	แ 225 3649 E41	โ 226 3650 E42	ใ 227 3651 E43	ไ 228 3652 E44	ๅ 229 3653 E45	ๆ 230 3654 E46
◌็ 231 3655 E47	◌่ 232 3656 E48	◌้ 233 3657 E49	◌๊ 234 3658 E4A	◌๋ 235 3659 E4B	◌์ 236 3660 E4C	◌ํ 237 3661 E4D	◌๎ 238 3662 E4E	◉ 239 3663 E4F	๐ 240 3664 E50
๑ 241 3665 E51	๒ 242 3666 E52	๓ 243 3667 E53	๔ 244 3668 E54	๕ 245 3669 E55	๖ 246 3670 E56	๗ 247 3671 E57	๘ 248 3672 E58	๙ 249 3673 E59	๚ 250 3674 E5A

และในสำหรับการแสดงผลภาษาอังกฤษนั้น ก็ได้มีการจัดแบ่งรหัสสำหรับการแสดงผล คือ รหัสมาตรฐาน Unicode ในรูปแบบของเลขฐานสิบหก ตั้งแต่ 1F-BE ซึ่งมีขนาดจัดเก็บ 2 ไบท์ รหัส ACSII ฐานสิบตั้งแต่ 31-190 ซึ่งมีขนาดจัดเก็บ 1 ไบท์ และ Unicode ฐานสิบ ตั้งแต่ 31-190 ดังตารางที่ 17 ซึ่งเป็นตารางแสดงค่า Unicode สำหรับแสดงผลภาษาอังกฤษ โดยมีตารางที่ 16 เป็นตารางอธิบายการอ่านค่าในตารางที่ 17

ตารางที่ 16 อธิบายแถวในการอ่านค่าในตารางที่ 17 สำหรับการแสดงผลภาษาอังกฤษ

ตัวเลขแถวที่ 1	แทน อักขระภาษาไทย	A
ตัวเลขแถวที่ 2	แทน รหัส ASCII : ฐานสิบ (Dec)	65
ตัวเลขแถวที่ 3	แทน รหัส Unicode : ฐานสิบ (Dec)	65
ตัวเลขแถวที่ 4	แทน รหัส Unicode : ฐานสิบหก (Hex)	41

ตารางที่ 17 แสดงค่า Unicode สำหรับการแสดงผลภาษาอังกฤษ

		!	"	#	$	%	&	'	(
31	32	33	34	35	36	37	38	39	40
31	32	33	34	35	36	37	38	39	40
1F	20	21	22	23	24	25	26	27	28
)	*	+	,	-	.	/	0	1	2
41	42	43	44	45	46	47	48	49	50
41	42	43	44	45	46	47	48	49	50
29	2A	2B	2C	2D	2E	2F	30	31	32
3	4	5	6	7	8	9	:	;	<
51	52	53	54	55	56	57	58	59	60
51	52	53	54	55	56	57	58	59	60
33	34	35	36	37	38	39	3A	3B	3C
=	>	?	@	A	B	C	D	E	F
61	62	63	64	65	66	67	68	69	70
61	62	63	64	65	66	67	68	69	70
3D	3E	3F	40	41	42	43	44	45	46

ตารางที่ 17 แสดงค่า Unicode สำหรับการแสดงผลภาษาอังกฤษ (ต่อ)

G	H	I	J	K	L	M	N	O	P
71	72	73	74	75	76	77	78	79	80
71	72	73	74	75	76	77	78	79	80
47	48	49	4A	4B	4C	4D	4E	4F	50
Q	R	S	T	U	V	W	X	Y	Z
81	82	83	84	85	86	87	88	89	90
81	82	83	84	85	86	87	88	89	90
51	52	53	54	55	56	57	58	59	5A
[	\	]	^	_	`	a	b	c	d
91	92	93	94	95	96	97	98	99	100
91	92	93	94	95	96	97	98	99	100
5B	5C	5D	5E	5F	60	61	62	63	64
e	f	g	h	i	j	k	l	m	n
101	102	103	104	105	106	107	108	109	110
101	102	103	104	105	106	107	108	109	110
65	66	67	68	69	6A	6B	6C	6D	6E
o	p	q	r	s	t	u	v	w	x
111	112	113	114	115	116	117	118	119	120
111	112	113	114	115	116	117	118	119	120
6F	70	71	72	73	74	75	76	77	78
y	z	{	|	}	~		€		,
121	122	123	124	125	126	127	128	129	130
121	122	123	124	125	126	127	128	129	130
79	7A	7B	7C	7D	7E	7F	80	81	82
f	„	…	†	‡	ˆ	‰	S	‹	Œ
131	132	133	134	135	136	137	138	139	140
131	132	133	134	135	136	137	138	139	140
83	84	85	86	87	88	89	8A	8B	8C
	Ž			‘	’	“	”	•	–
141	142	143	144	145	146	147	148	149	150
141	142	143	144	145	146	147	148	149	150
8D	8E	8F	90	91	92	93	94	95	96
—	˜	™	š	›	œ		ž	Y	
151	152	153	154	155	156	157	158	159	160
151	152	153	154	155	156	157	158	159	160
97	98	99	9A	9B	9C	9D	9E	9F	A0
¡	¢	£	¤	¥	¦	§	¨	©	ª
161	162	163	164	165	166	167	168	169	170
161	162	163	164	165	166	167	168	169	170
A1	A2	A3	A4	A5	A6	A7	A8	A9	AA

ตารางที่ 17 แสดงค่า Unicode สำหรับการแสดงผลภาษาอังกฤษ (ต่อ)

«	¬		®		°	±	²	³	´
171	172	173	174	175	176	177	178	179	180
171	172	173	174	175	176	177	178	179	180
AB	AC	AD	AE	AF	B0	B1	B2	B3	B4
µ	¶	·	¸	¹	º	»	¼	½	¾
181	182	183	184	185	186	187	188	189	190
181	182	183	184	185	186	187	188	189	190
B5	B6	B7	B8	B9	BA	BB	BC	BD	BE

จากค่าในตารางที่ 15 และ ตารางที่ 17 จะสังเกตเห็นว่าค่า รหัส ASCII ทั้งภาษาไทยและภาษาอังกฤษมีค่าที่ซ้ำกันเกิดขึ้น นั้นหมายความว่าหากมีการใช้รหัส ASCII ในการแสดงผลจะเกิดการแสดงผลที่อาจคลาดเคลื่อนได้ยกตัวอย่างการแทนค่าเลขฐานสำหรับตัวอักษรภาษาไทยและภาษาอังกฤษที่มีค่าซ้ำกัน

ตัวอย่างที่ 1 ตัวอักษร ฉ

แทน รหัส ASCII : ฐานสิบ (Dec) คือ 169

แทน รหัส Unicode : ฐานสิบ (Dec) คือ 3593

แทน รหัส Unicode : ฐานสิบหก (Hex) คือ E09

ตัวอย่างที่ 2 สัญลักษณ์ ©

แทน รหัส ASCII : ฐานสิบ (Dec) คือ 169

แทน รหัส Unicode : ฐานสิบ (Dec) คือ 169

แทน รหัส Unicode : ฐานสิบหก (Hex) คือ A9

จากทั้งสองตัวอย่างจะเห็นว่า จะเกิดปัญหาขึ้นหากนำมาตรฐาน รหัส ASCII มาใช้ในงานจะเกิดค่าซ้ำซ้อนกันเกิดขึ้น คือ 169 ดังนั้นการใช้ Unicode ในการทำงานจะช่วยลดปัญหาดังกล่าวที่เกิดขึ้นเพราะค่า รหัส Unicode ฐานสิบสำหรับภาษาไทย คือ 3585 – 3674 จะไม่ซ้ำกับ ค่ารหัส Unicode ฐานสิบสำหรับภาษาอังกฤษ คือ 31 – 190 เป็นต้น

3.2 การเข้าและถอดรหัสของ อัลกอริทึมแบบ Data Encryption Standard และ Advance Encryption Standard สำหรับภาษาไทย โดยใช้ วิธีการ mapping กับตาราง Unicode

ในการสร้างโปรแกรมสำหรับการเข้าและถอดรหัส เพื่อรองรับการใช้งานได้หลายภาษานั้นมีความจำเป็นอย่างยิ่ง ที่จะต้องเข้าใจรูปแบบการทำงานของแต่ละอัลกอริทึมในการใช้งาน โดยในงานวิจัยชิ้นนี้ได้เลือกรูปแบบการเข้าและถอดรหัสแบบสมมาตรในการศึกษา โดยเลือกอัลกอริทึมที่เป็นที่นิยมและยอมรับ มาทำการ ศึกษาและเปรียบเทียบรูปแบบการทำงาน คือ อัลกอริทึมแบบ DES และ AES เมื่อทำการเลือกอัลกอริทึมเป็นที่เรียบร้อยขั้นตอนต่อไป จึงทำการเลือกมาตรฐานในการใช้งานภาษาไทย โดยรูปแบบที่เลือกใช้คือ Unicode ซึ่งมีการจัดเก็บข้อมูลต่อตัวอักขระ คือ 2 ไบท์ และในการทำงานครั้งนี้ได้สรุปขั้นตอนหลักในการศึกษาเป็นส่วนๆ ดังนี้

3.2.1 Data Encryption Standard (DES)

ในเนื้อหาส่วนนี้กล่าวถึงองค์ประกอบหลักของการดำเนินงาน ในส่วนของ การเข้าและถอดรหัสแบบ DES ซึ่งเป็นส่วนหนึ่งในการดำเนินงานครั้งนี้ ในการออกแบบการสร้างซอฟต์แวร์ในส่วนนี้ จะประกอบไปด้วยตารางค่าอ้างอิงต่างๆและขั้นตอนหลักในการทำงาน ดังนี้

(1) Initial Permutation (IP) Table

(2) Expansion (E)

(3) S-box

(4) Primitive function P

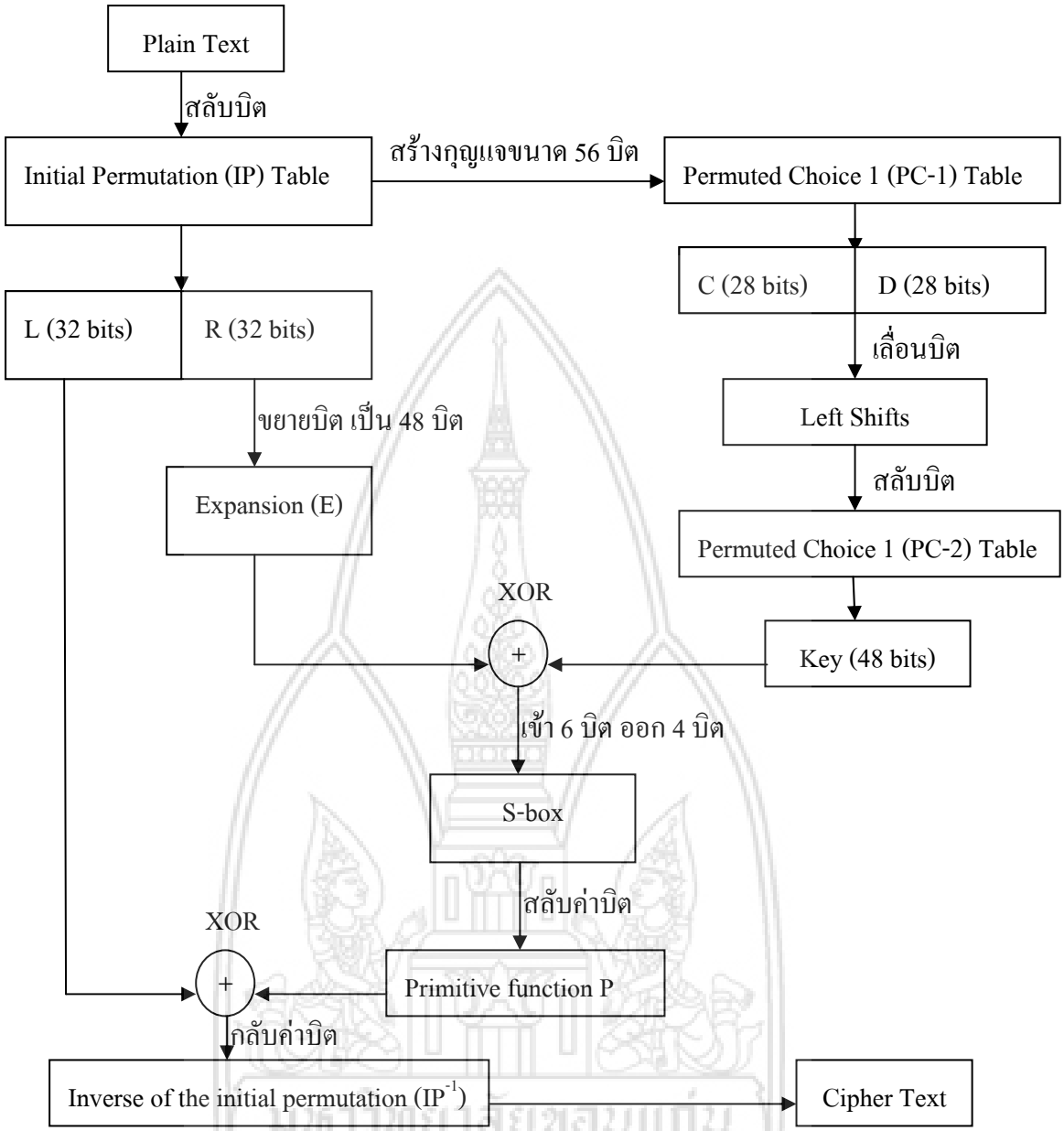
(5) Inverse of the initial permutation (IP^{-1})

(6) Permuted Choice 1 (PC-1) Table

(7) Permuted Choice 1 (PC-2) Table

(8) Left Shifts

จากตารางค่าดังกล่าว นำมาสร้าง โดยการนำเข้าชุดข้อมูลแบบบล็อกขนาด 64 บิต และใช้กุญแจขนาด 56 บิต มีจำนวนรอบในการทำงานเท่ากับ 16 รอบ เพื่อสร้างกุญแจย่อยให้มีจำนวน 16 ดอก โดยระหว่างการเข้ารหัสนั้นแต่ละรอบการทำงานจะมีกุญแจขนาด 48 บิต โดยสามารถอธิบายการทำงานเบื้องต้นได้ดังไดอะแกรมด้านล่างนี้



ภาพที่ 28 แสดงการทำงานเบื้องต้นของอัลกอริทึม DES

3.2.2 AES

ในเนื้อหาส่วนนี้กล่าวถึงองค์ประกอบหลักของการดำเนินงานโดยสรุป ในส่วนของ การเข้าและถอดรหัสแบบ AES ซึ่งจำเป็นอย่างยิ่งในการออกแบบการสร้างซฮฟต์แวร์ โดยอัลกอริทึม AES มีหลักการทำงานคร่าวๆ ดังนี้

(1) ขนาดของ blockที่ใช้ในการคำนวณ มีขนาด 128-bit

(2) ขนาดของกุญแจ ที่ใช้ในการคำนวณ มีขนาด 128, 192 หรือ 256

(3) ในการทำงานมีทางเลือกมากกว่า Feistel cipher (คล้ายกับ IDEA)

(4) มีการจัดการขนาดข้อมูล 4 กลุ่ม ของ 4 ไบท์

(5) ในการทำงานแต่ละรอบนั้นจะบรรจุไปด้วย

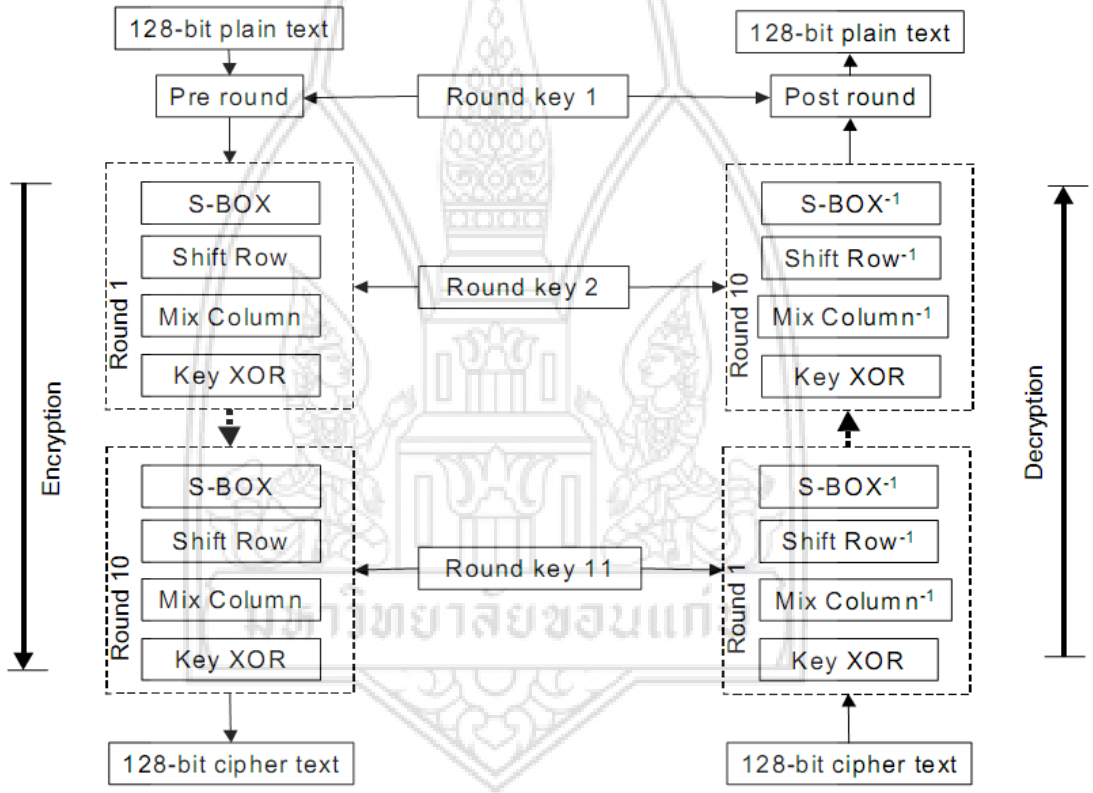
(5.1) ขั้นตอนในการแทนค่า ไบท์ (SubBytes)

(5.2) ขั้นตอนในการเลื่อนแถว (ShiftRows)

(5.3) ขั้นตอนในการผสมคอลัมภ์ (MixColumns)

(5.4) ขั้นตอนในการเพิ่มรอบของกุญแจ (AddRoundKey)

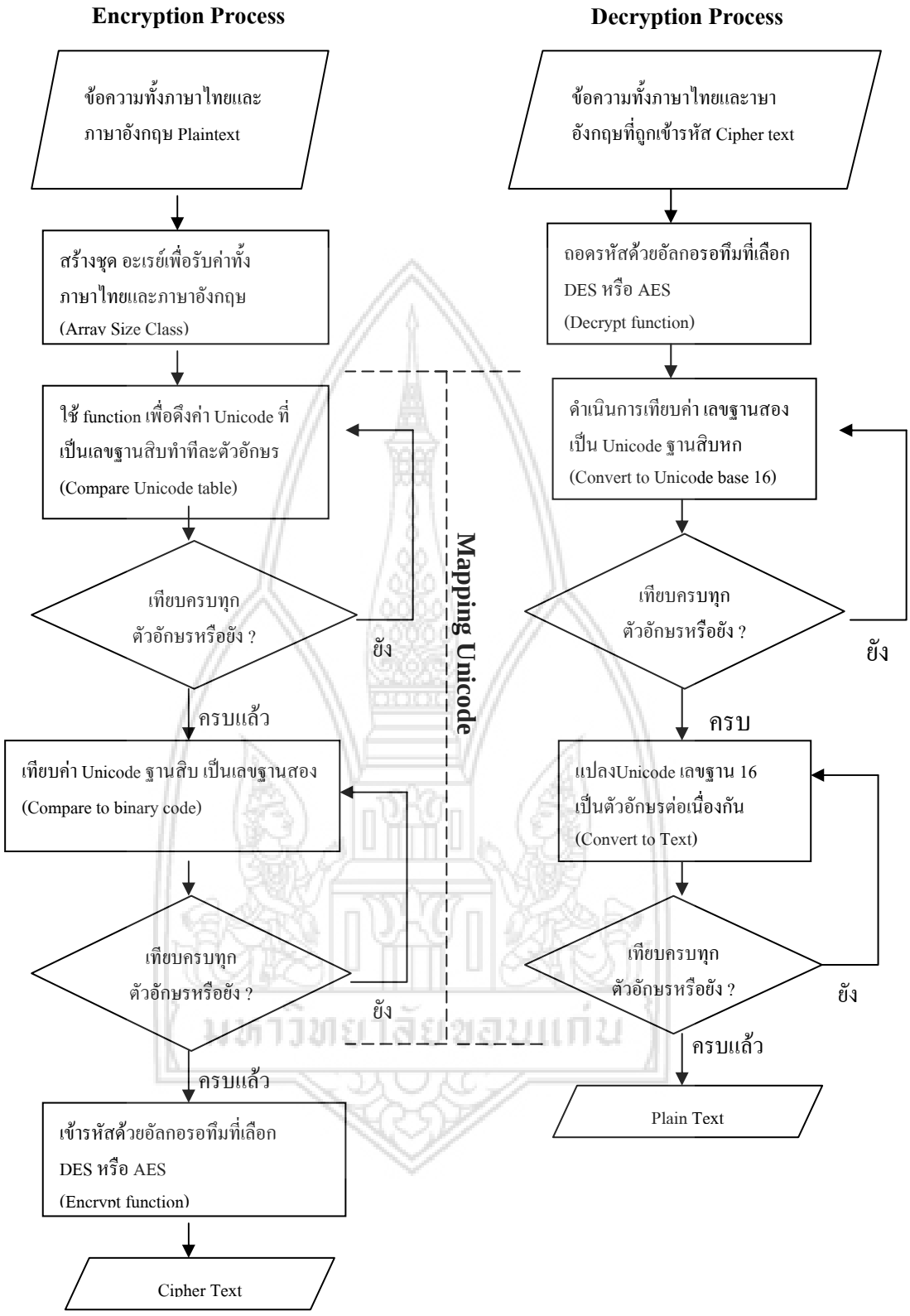
(6) การดำเนินการทั้งหมดถูกนำมาเปรียบเทียบ xor และ ตาราง lookups ดังนั้นจึงทำให้การทำงานมีทั้งประสิทธิภาพความเร็วที่เพิ่มมากขึ้น โดยคำนวณทั้งหมด 10 รอบ



ภาพที่ 29 แสดงการทำงานเบื้องต้นของอัลกอริทึม AES

3.2.3 วิธีการ mapping กับตาราง Unicode

ในการพัฒนาโปรแกรมการเข้าและถอดรหัสแบบ DES และ AES สำหรับภาษาไทย โดยใช้ วิธีการ mapping กับตาราง Unicode นั้น เป็นการเพิ่มขั้นตอนการ mapping เลขฐานสองของตัวอักขระแต่ละตัวที่นำเข้ามา โดยกำหนดให้ 1 ตัวอักขระนั้นมีขนาด 2 ไบท์ จากนั้นนำอักขระหนึ่งตัวนั้นมาทำการเทียบค่ากับตาราง Unicode ที่สร้างเป็นตารางอ้างอิงในโปรแกรม เมื่อทำการเทียบค่าจนครบทุกตัวอักขระตามความยาวของข้อมูลที่นำเข้ามาเป็นที่เรียบร้อยแล้ว จึงทำการเข้ารหัสตามรูปแบบของอัลกอริทึมที่เลือก คือ DES หรือ AES โดยสามารถแบ่งกระบวนการทำงานออกเป็นส่วนๆ ได้ดังภาพ



ภาพที่ 30 แสดงการเข้าและถอดรหัสด้วยเทคนิค

3.3 การสร้างระบบและจำลองการทำงาน

ในขั้นตอนการวิจัยครั้งนี้ ได้ทำการจัดเตรียมอุปกรณ์ทั้งด้านซอฟต์แวร์ และ ฮาร์ดแวร์ เพื่อจำลองการทำงานของกระบวนการเข้าและถอดรหัส โดยแบ่งรูปแบบการศึกษาและวิเคราะห์การทำงานออกเป็นส่วนๆ ดังนี้

3.3.1 System Parameters

Hardware

1. Intel® Core™ 2 Duo Duo CPU , T6500 @ 2.10 GHz
2. 2GB of RAM
3. Hard disk 320 GB

Software

1. Microsoft Visual Studio 2008
2. Window XP SP2

3.3.2 Experiment Factors

ในการเปรียบเทียบประสิทธิภาพการทำงานของโปรแกรมที่สร้างนั้น จะทำการนำเข้าข้อความทั้งภาษาไทยและภาษาอังกฤษที่มีความหลากหลายของคำที่ใช้ โดยกำหนดขนาดความยาวตั้งแต่ 1000 ไบท์ จนถึง 5000 ไบท์ ซึ่งโปรแกรมที่สร้างนั้นจะทำการบันทึกเวลาในการทำงานแต่ละครั้ง แล้วนำผลที่ได้มาทำการหาค่าเฉลี่ย เพื่อศึกษาถึงความถูกต้องในการเข้าและถอดรหัสคำภาษาไทย