

คำในภาษาไทยคำหนึ่ง ๆ อาจประกอบไปด้วย พยัญชนะ สระ วรรณยุกต์ต่าง ๆ ซึ่งอยู่ในบรรทัดเดียวกันหรืออยู่คนละบรรทัด โดยที่สระอาจมีทั้งอยู่เหนือบรรทัดหรือใต้บรรทัดของพยัญชนะ และยังมีวรรณยุกต์อยู่อีกหนึ่งบรรทัด ดังนั้นจึงพอสรุปได้ว่าลักษณะโครงสร้างของตัวอักษรในภาษาไทยแบ่งออกเป็น 4 ระดับดังนี้

ระดับที่ 1 เป็นระดับของตัวอักษรประเภทวรรณยุกต์และตัวการันต์มีทั้งหมด 5 ตัว ได้แก่ ไม้ทัณฑฆาต และ ไม้ยมก

ระดับที่ 2 เป็นระดับของตัวอักษรประเภทสระบางประเภทรวมทั้งไม้ไต่คู้ และ หยาคน้ำค้างมีทั้งหมด 7 ตัว ได้แก่ ไม้ไต่คู้ ไม้ยมก และ (หยาคน้ำค้าง)

ระดับที่ 3 เป็นระดับที่สำคัญของภาษาไทยมีตัวอักษรมากที่สุดมีทั้งหมด 56 ตัว เป็น พยัญชนะ 46 ตัว (ก-ฮ และ ฦ ฼) และสระอีก 10 ตัว (รวมทั้ง ๆ และ ๆ) ได้แก่ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙ และ ๐

ระดับที่ 4 เป็นระดับของตัวอักษรประเภทสระมีทั้งหมด 2 ตัว ได้แก่ ๑ และ ๒

รวมตัวอักษรในภาษาไทยทั้งหมดมี 70 ตัว

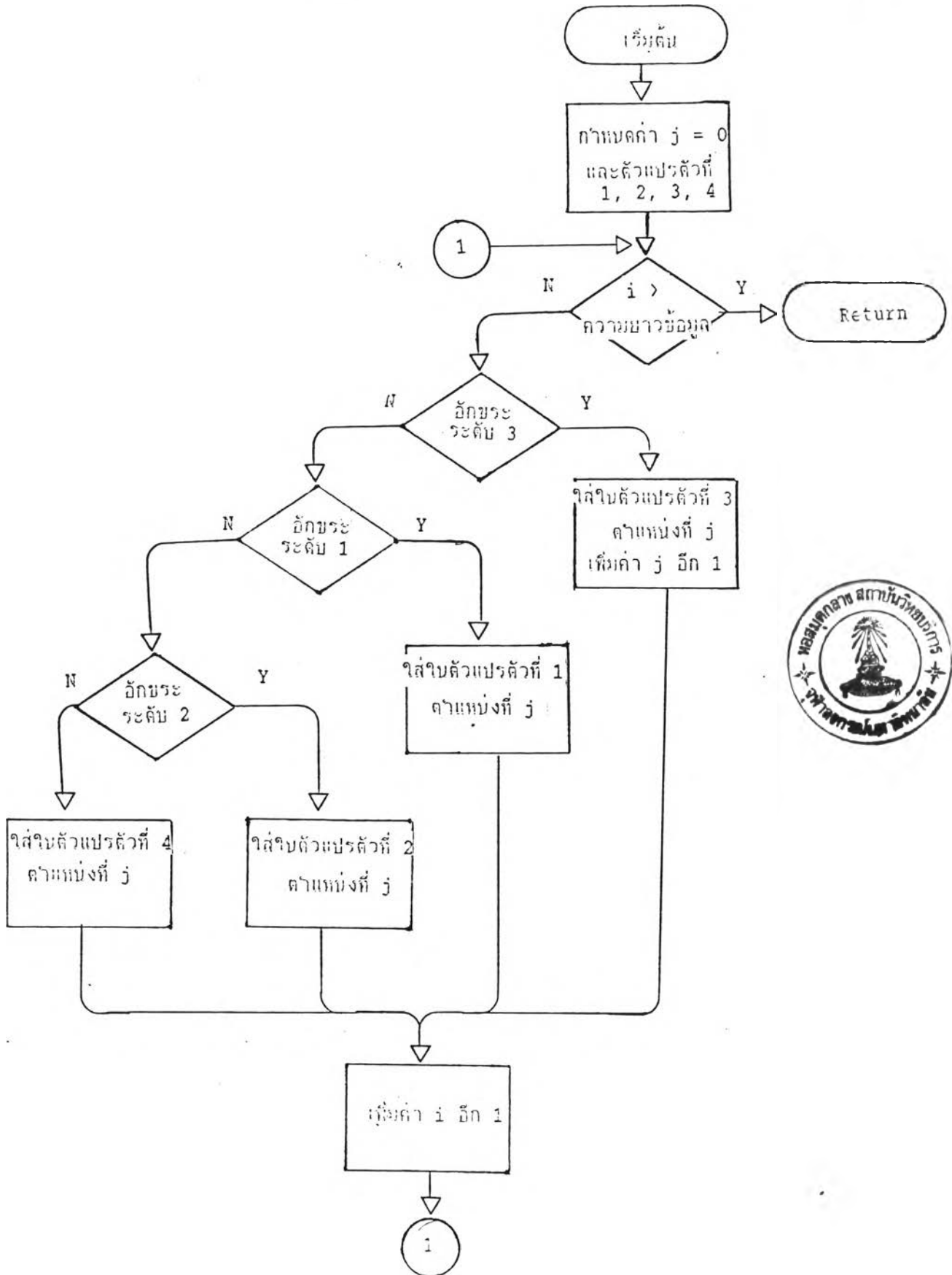
3.2 การจัดเรียงลำดับตัวอักษรในภาษาไทย เพื่อบันทึกเข้าสู่ระบบคอมพิวเตอร์

ดังที่กล่าวมาแล้วว่า ระบบคอมพิวเตอร์ได้ออกแบบมาเพื่อใช้กับภาษาที่มีโครงสร้างของตัวอักษรอยู่ในบรรทัดเดียวกันหรือระดับเดียว ดังนั้นข้อมูลที่เป็นอักษรในภาษาไทยจึงจำเป็นต้องจัดลำดับหรือตำแหน่งของอักขระก่อนเพื่อให้อยู่ในรูปแบบที่ระบบคอมพิวเตอร์สามารถอ่านได้ เช่น คำว่า "สื่อสาร" จะต้องจัดตัวอักขระให้อยู่ในบรรทัดเดียวกันและเรียงลำดับหรือตำแหน่งของตัวอักขระดังนี้ "ส อ สาร" หรือ "ส อ สาร" เป็นต้น

จากกฎเกณฑ์ที่พอสรุปได้ว่าถ้าตำแหน่งใดมีตัวอักขระมากกว่าหนึ่งตัว ตัวอักษรประเภทพยัญชนะจะนำหน้าตัวอักขระอื่น ๆ แล้วจึงตามด้วยอักขระประเภทอื่น ตัวอย่างเช่น คำว่า "คู" อาจเป็น "คู" หรือ "คู" ก็ได้ การสลับตำแหน่งของตัวอักษรในระดับ 1 2 และ 4 ไม่มีผลต่อการจัดเรียงพิมพ์

3.3 การเรียงพิมพ์ภาษาไทย

เนื่องด้วยข้อมูลอักขระไทยที่จะเข้าสู่ระบบคอมพิวเตอร์ ต้องจัดตัวอักขระให้อยู่ในบรรทัดเดียวกันและจัดลำดับหรือตำแหน่งตัวอักขระก่อน เช่นคำว่า "สื่อสาร" อาจถูกจัดเป็น "ส อสื่อสาร" หรือ "ส อสื่อสาร" ดังที่กล่าวมาแล้ว ดังนั้นในการพิมพ์ภาษาไทยจึงต้องมีการจัดเรียงพิมพ์เพื่อจัดระดับและตำแหน่งของตัวอักขระให้กลับไปอยู่ในลักษณะ 4 ระดับตามเดิม ดังนั้นการเรียงพิมพ์ภาษาไทยจึงต้องประกอบด้วยการจัดประเภทของอักขระว่าอยู่ในระดับใด ตำแหน่งการพิมพ์จะเคลื่อนไปทางขวาถึงตำแหน่ง เมื่ออักขระที่พบเป็นประเภทหรืออยู่ในระดับที่ 3 เท่านั้น เนื่องจากการสลับตำแหน่งของตัวอักขระในระดับ 1 2 และ 4 ไม่มีผลต่อการจัดเรียงพิมพ์ ฉะนั้นการจัดเรียงพิมพ์อักขระไทยจะจัดเรียงตามประเภท 3 ประเภท 2 ประเภท 1 และประเภท 4 ตามลำดับ ซึ่งอาจสรุปเป็นผังดำเนินการได้ตามผังงานที่ 3.1



สำหรับรหัสที่ใช้แทนตัวอักษรภาษาไทย - อังกฤษ ตัวเลขและตัวอักษรพิเศษใน
ระบบเอ็บซีดีค (EBCDIC) มีรหัสดังตารางที่ 3.1

bit 8	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
7	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
6	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
bit 5	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
4 3 2 1																
0 0 0 0	nul	dle	ds	sp	&	-			ก	ท	ม	า				0
0 0 0 1	soh	dcl	sos		/				ข	ฃ	ช	า	A	J		1
0 0 1 0	stx	dc2	fs	syn					ข	ฃ	ว		B	K	S	2
0 0 1 1	etx	tm							ค	ฅ	ณ		C	L	T	3
0 1 0 0	pf	res	byp	pn					ค	ฅ	น		D	M	U	4
0 1 0 1	ht	nl	lf	rs					ท	ก	ก		E	N	V	5
0 1 1 0	lc	bs	etb	uc					ง	ท	ว		F	O	W	6
0 1 1 1	del	ll	esc	eot					จ	ฆ	ศ		G	P	X	7
1 0 0 0	ge	can							ฉ	บ	ช		H	Q	Y	8
1 0 0 1	rlf	em							ช	บ	ส		I	R	Z	9
1 0 1 0	smm	cc	sm	C	!	!			ช	ป	ท					
1 0 1 1	vt	cul	cu2	cu3	.	S	.	C	ง	จ	ฑ					
1 1 0 0	ff	ifs	dc4	<	*	%	@		ณ	ณ	ภ					
1 1 0 1	cr	igs	enq	nak	()	-	'	ภ	พ	ช					
1 1 1 0	so	irs	ack	+	:	>	#		ภ	พ	ช					
1 1 1 1	si	ius	bel	sub	-	?			ฐ	ณ						

ตารางที่ 3.1 อัญมรพทตรวจสอบ EBCDIC สมบูรณ์ของไอบีเอ็ม ระบบ 370