บทที่ 3

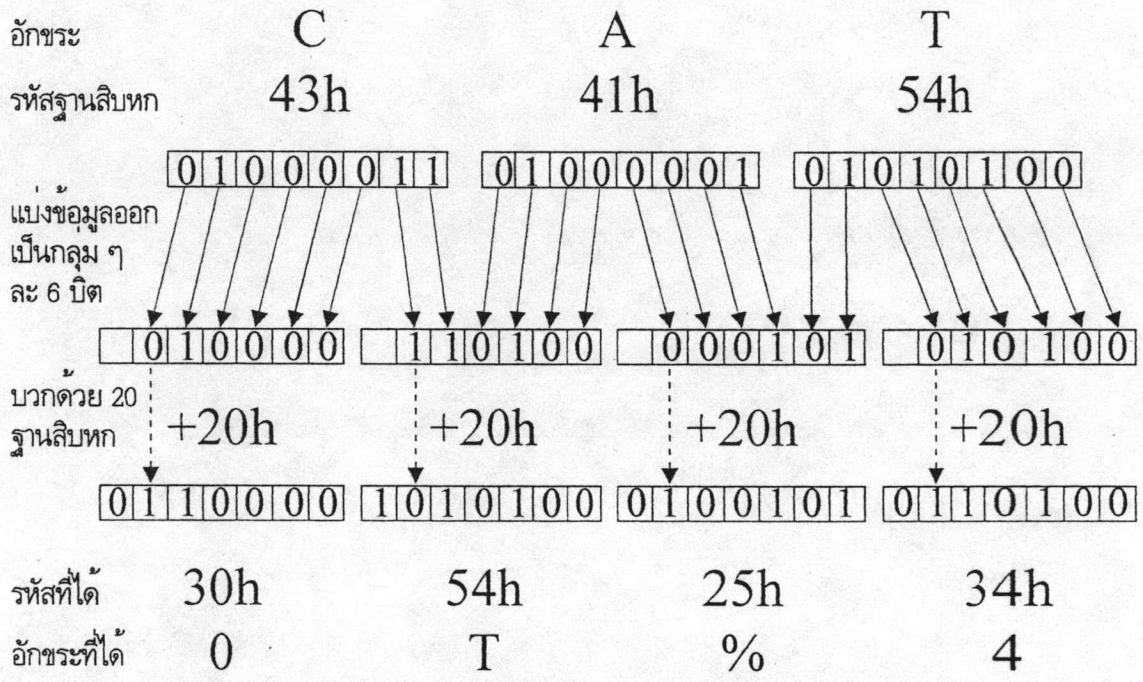
การแปลงรหัส

เราสามารถแปลงรหัสมาตรฐาน 8 บิต ดังกล่าวเป็นรหัสกลาง 7 บิตได้หลายวิธีในที่นี้จะอธิบายถึงวิธีการแปลงเป็นรหัสกลาง และแปลงกลับเป็นรหัสดั้งต้น 3 วิธี คือ วิธีของโปรแกรมยูยูเอ็นโค้ด วิธีการแทรกอักขระพิเศษ และวิธีใช้อักขระคู่ (double character)

การเข้ารหัสโดยวิธีของโปรแกรมยูยูเอ็นโค้ด

โปรแกรมยูยูเอ็นโค้ดเป็นโปรแกรมอรรถประโยชน์ที่ใช้แปลงข้อมูล 8 บิตเป็นรหัสกลางขนาด 7 บิต จากการศึกษาพบว่าโปรแกรมดังกล่าวจะรับข้อมูลขนาด 8 บิตมาทีละกลุ่ม กลุ่มละ 3 ไบต์รวมเป็น 24 บิต จากนั้นนำมาแปลงเป็นรหัสกลางจำนวน 4 ไบต์ โดยแบ่งข้อมูลที่รับมา 24 บิตนั้นออกเป็น 4 กลุ่ม ๆ ละ 6 บิต ข้อมูลทั้ง 6 บิตจะวางอยู่ในตำแหน่งบิตที่ 0 ถึงบิตที่ 5 ของทั้ง 4 ไบต์ ส่วนบิตที่ 6 และ 7 จะมีค่าเป็น 0 เสมอ จากนั้นจะบวกด้วยรหัสแอสกี 20h เพื่อเลี่ยงการเป็นอักขระควบคุม (control character) ส่วนการถอดรหัสทำได้โดยวิธีการกลับกันโดยที่นำแต่ละไบต์ทั้งสี่ไบต์มาลบด้วย 20h ก่อน แล้วนำบิตของข้อมูลตั้งแต่บิตที่ 0 ถึงบิตที่ 5 ทั้ง 4 ไบต์มารวมกันแล้วแบ่งออกเป็นไบต์ขนาด 8 บิตได้ 3 ไบต์

ตัวอย่างการแปลงรหัสโดยวิธีของโปรแกรมยูยูเอ็นโค้ด เช่นคำว่า "CAT" เมื่อแปลงเป็นรหัสกลางโดยวิธีนี้จะได้เป็นสายอักขระ (string) "OT%4" ดังแสดงในรูป 3.1 โดยที่อักขระแต่ละตัวในคำนี้ เมื่อแปลงเป็นเลขฐานสองจะได้เป็นกลุ่มตัวเลข 3 กลุ่มดังนี้ 01000011 01000001 01010100 เมื่อนำมาเรียงกันแล้วแบ่งกลุ่มใหม่เป็น 4 กลุ่ม กลุ่มละ 6 บิตจะได้ 010000, 110100, 000101, 010100 จากนั้นบวกแต่ละกลุ่มด้วย 20h จะเป็น 0110000, 1010100, 0100101, 1010100 ขณะนี้จะได้รหัส 7 บิตทั้งสิ้น 4 อักขระเมื่อเทียบในตารางรหัสแอสกีจะเป็นสายอักขระ "OT%4"



โดยวิธีของโปรแกรมยูยูเอ็นโค้ดข้อความ"CAT" จะได้รับการเข้ารหัสเป็น"0T%4"

รูป 3.1 แสดงตัวอย่างการเข้ารหัสโดยวิธีของโปรแกรมยูยูเอ็นโค้ด

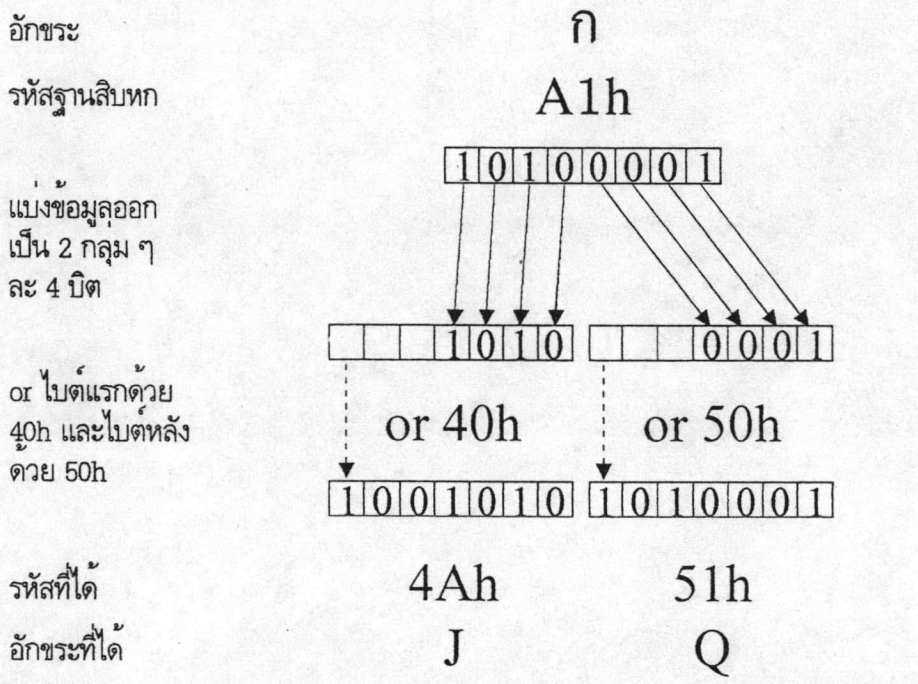
ข้อดีของวิธีนี้คือได้รหัสกลางที่มีขนาดเพิ่มขึ้นจากเดิมเพียง 33% เหมาะสำหรับการแปลงแฟ้มแบบไบนารี และเป็นวิธีเข้ารหัสที่รู้จักกันดีแล้ว ส่วนข้อเสียคือเมื่อรหัสกลางบางส่วนสูญหายจะทำให้การแปลงรหัสกลับผิดพลาดเริ่มตั้งแต่ตำแหน่งที่ข้อมูลสูญหายจนจบเอกสาร และจำเป็นต้องรับข้อมูลมาทีละกลุ่ม กลุ่มละสามไบต์จึงจะสามารถแปลงข้อมูลเป็นรหัสกลางได้สี่ไบต์ ในทำนองเดียวกันในการถอดรหัสต้องรับข้อมูลมาทีละ4ไบต์ให้ครบก่อนจึง จะถอดรหัสได้ข้อมูลเดิมสามไบต์

การเข้ารหัสโดยวิธีอักขระคู่ (double character)

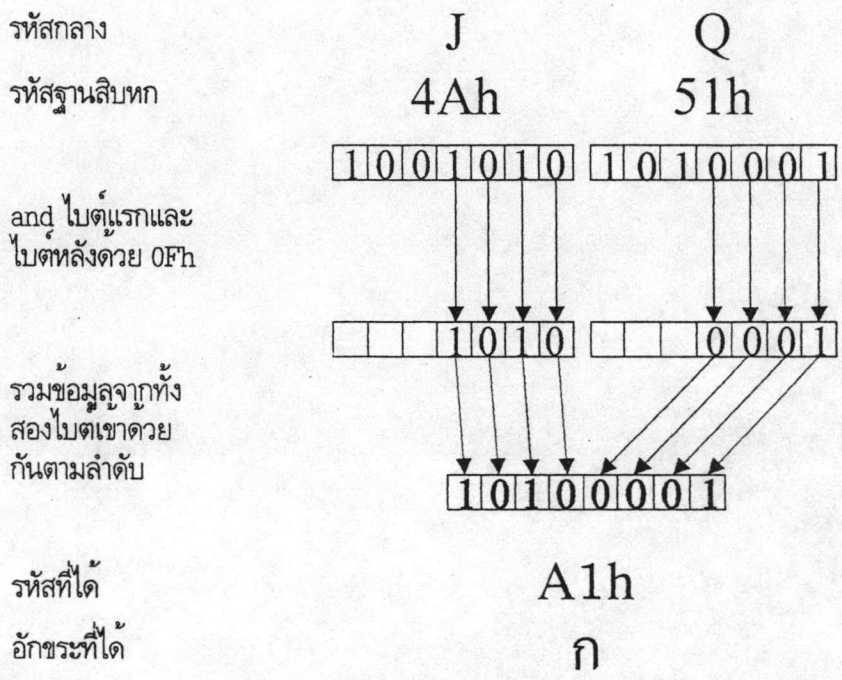
วิธีนี้จะการแปลงข้อมูลที่อยู่ในรูปรหัส 8 บิต 1 ไบต์ เป็นรหัส 7 บิต จำนวน 2 ไบต์ข้อดีของวิธีนี้คือทำให้เราสามารถเพิ่มการตรวจสอบความถูกต้องของข้อมูลลงไปในรหัสกลางได้ รหัสกลางในวิธีนี้จะเป็นคู่ ๆ มีขนาด 14 บิต

การเข้ารหัสโดยวิธีอักขระคู่ทำได้โดยจากข้อมูลขนาด 8 บิต แบ่งออกเป็น สองส่วน ส่วนละ 4 บิต นำข้อมูลแต่ละส่วนแยกกันเก็บไว้ที่ บิตที่ 0 ถึง 3 ของไบต์แรกและไบต์หลังตามลำดับ จากนั้นกำหนดให้บิตที่ 4 ทำหน้าที่บอกว่าไบต์นั้นเป็นไบต์แรกหรือไบต์หลังโดยที่เมื่อเป็นไบต์แรกจะมีค่าเป็นศูนย์ และเมื่อเป็นไบต์หลังจะมีค่าเป็นหนึ่ง เพื่อบังคับให้ผลจากการแปลงรหัสไม่เป็นอักขระควบคุมและเป็นอักขระภาษาอังกฤษตัวใหญ่ (upper case letter) จะกำหนดให้บิตที่ 6 มีค่าเป็น 1 เสมอ ดังนั้นรหัสที่ได้จะอยู่ในช่วง 40h ถึง 4Fh เมื่อเป็นไบต์หน้าและอยู่ในช่วง 50h ถึง 5Fh เมื่อเป็นไบต์หลัง รูป 3.2 แสดงตัวอย่างการเข้ารหัสโดยวิธีอักขระคู่สำหรับอักขระ "ก" มีรหัสแอสกี A1h จะแยกออกเป็นสองไบต์ ได้ 0Ah และ 01h จากนั้น ดำเนินการทางตรรกศาสตร์ "หรือ" กับไบต์แรกด้วย 40h และไบต์หลังด้วย 50h จะได้ผลลัพธ์เป็น 4Ah และ 51h ซึ่งตรงกับอักขระ J และ Q

ในการถอดรหัส จะรับข้อมูลมาทีละคู่ถ้าพบว่าข้อมูลไม่ครบคู่จะทิ้งข้อมูลชุดนั้น เมื่อได้ข้อมูลครบคู่แล้ว จะนำเอาเฉพาะส่วนที่เก็บข้อมูลในรหัสกลางคือ บิต 0 ถึงบิต 3 ของไบต์แรกและไบต์หลังมารวมกันเป็นหนึ่งไบต์ที่มีขนาด 8 บิตดังแสดงในรูป 3.3



รูป 3.2 แสดงตัวอย่างการเข้ารหัสโดยวิธีอักขระคู่



รูป 3.3 แสดงการถอดรหัสเมื่อรหัสกลางเป็นชนิดอักขระคู่

ข้อเสียของวิธีนี้คือรหัสที่ได้จะมีขนาดเพิ่มขึ้นจากเดิมเป็น 2 เท่า ส่วนข้อดีคือในกรณีที่มีข้อผิดพลาดในการสื่อสารเกิดขึ้นทำให้ข้อมูลบางส่วนขาดหายไป วิธีนี้จะจำกัดความเสียหายต่อการถอดรหัสเฉพาะคู่ที่ผิดพลาดเท่านั้น เพราะการถอดรหัสเป็นอิสระจากข้อมูลที่มาก่อนหน้า

การเข้ารหัสโดยการแทรกอักขระพิเศษ

การเข้ารหัสโดยการแทรกอักขระพิเศษนี้เป็นการแทรกรหัสบางอย่างเพื่อบอกจุดเริ่มต้นการแปลงเป็นรหัสกลาง และเพื่อบอกจุดสิ้นสุดการแปลงรหัส อักขระนี้อาจเป็นอักขระใดก็ได้ โดยมากนิยมใช้ลำดับอักขระหลีก (escape sequence) วิธีการนี้เป็นวิธีหนึ่งที่ กำหนดขึ้นใช้สำหรับภาษาตะวันออกที่ใช้ชุดอักขระคู่ หรือดีบีซีเอส (double byte character set,DBCS) เช่นภาษาจีน เกาหลี และญี่ปุ่น โดยจะกำหนดอักขระพิเศษเพื่อบอกจุดเริ่มต้นและเพื่อบอกจุดจบของดีบีซีเอส

ตัวอย่างการแปลงรหัสระหว่างรหัสJIS-Romanเป็นรหัส JIS X 0208-1983ซึ่งเป็นรหัสมาตรฐานสำหรับระบบข่าวสารในภาษาญี่ปุ่น สำหรับข้อความว่า "คะนะคันจิ" หรือ "かな漢字" ในภาษาญี่ปุ่น โดยใช้ชุดอักขระพิเศษ 3 ตัวคือ <ESC> $ B ทำหน้าที่เลือกชุดอักขระ JIS X 0208-1983 และใช้ <ESC> (J เพื่อเลือก JIS-Roman หรือชุดอักขระปกติ โดยจะได้รับการเข้ารหัสดังแสดงตามรูป 3.4

ข้อความ				か	な	漢	字			
ชุดอักขระพิเศษ	<ESC>	$	B					<ESC>	(	J
ข้อมูลที่เข้ารหัสแล้ว	1Bh	24h	42h	242Bh	244Ah	3441h	3B7Ah	1Bh	28h	4Ah
รหัสแอสกี				$+	$J	4A	;z			

รูป 3.4 แสดงตัวอย่างการแปลงรหัสจาก JIS-Roman เป็น JIS X 0208-1983

โดยวิธีการเดียวกันนี้เราสามารถนำมาใช้ในการแปลงรหัสสำหรับอักขระไทยเป็นรหัสกลางได้โดยการตัดบิตที่ 7 ของสายอักขระไทยทิ้งแล้วแทรกอักขระพิเศษที่ปลายสายอักขระทั้งสองด้านเพื่อบอกให้ทราบว่ารหัสที่อยู่ระหว่างกลางนั้นเป็นรหัสกลาง ซึ่งจะกล่าวถึงโดยละเอียดในหัวข้อถัดไป

การเข้ารหัสที่ใช้ในวิทยานิพนธ์ฉบับนี้

ในวิทยานิพนธ์ฉบับนี้เลือกใช้วิธีการแทรกรหัสพิเศษ เพราะวิธีนี้ไม่จำเป็นต้องรอให้มีข้อมูลครบจำนวนก่อนที่จะแปลงรหัสเหมือนวิธีของโปรแกรมยูยูเอ็นโค้ด ซึ่งเป็นวิธีที่เหมาะสำหรับใช้ในการแปลงรหัสที่ต้องทำทันทีที่ได้รับอักขระมา และโดยวิธีนี้อักขระที่จะได้รับการแปลงจะจำกัดเฉพาะที่เป็นภาษาไทยเท่านั้นข้อความที่เป็นภาษาอังกฤษจะคงเดิม เหมาะจะใช้ในตัวเลียนแบบเทอร์มินัลเพราะทำให้สามารถนำเข้าคำสั่งของยูนิกซ์ได้เหมือนปรกติ ซึ่งต่างจากการแปลงข้อมูลเป็นรหัสกลางโดยวิธีอักขระคู่ซึ่งจะแปลงอักขระทุกตัวเป็นรหัสใหม่จะทำให้ยูนิกซ์ไม่สามารถเข้าใจคำสั่งที่เป็นรหัสกลางนั้นได้

การเข้ารหัสโดยวิธีนี้ใช้ลำดับอักขระหลีก 4 ลำดับเพื่อใช้ในการแปลงรหัส ลำดับอักขระหลีกนี้ทำหน้าที่บอกจุดเริ่มของชุดอักขระไทย บอกจุดจบของชุดอักขระไทย บอกจุดเริ่มของชุดอักขระควบคุม และบอกจุดจบของชุดอักขระควบคุมรหัสทั้งสี่นี้แสดงในตาราง 3.1 โดยที่อักขระไทยในที่นี้หมายถึงอักขระที่มีรหัสแอสกีมากกว่าหรือเท่ากับ 80h และอักขระควบคุมหมายถึงอักขระที่มีรหัสแอสกีน้อยกว่า 20h

ลำดับอักขระหลีก	ฐานสิบหก	ชื่อย่อ	หน้าที่
<ESC> { T	1B 7B 54	SO	บอกจุดเริ่มของชุดอักขระไทย
<ESC> { L	1B 7B 4C	SI	บอกจุดจบของชุดอักขระไทย
<ESC> { O	1B 7B 4F	SCO	บอกจุดเริ่มของชุดอักขระควบคุม
<ESC> { I	1B 7B 49	SCI	บอกจุดจบของชุดอักขระควบคุม

ตาราง 3.1 แสดงลำดับอักขระหลีกที่ใช้ในการแปลงรหัส

ลำดับอักขระหลีก	ค่าของตัวแปร SOmode	ค่าของตัวแปร SCOmode
<ESC> { T	1	ไม่มีการเปลี่ยนแปลงค่า
<ESC> { L	0	ไม่มีการเปลี่ยนแปลงค่า
<ESC> { O	ไม่มีการเปลี่ยนแปลงค่า	0
<ESC> { I	ไม่มีการเปลี่ยนแปลงค่า	1

ตาราง 3.2 แสดงค่าของตัวแปร SOmode และ SCOmode ที่ใช้ในการถอดรหัส

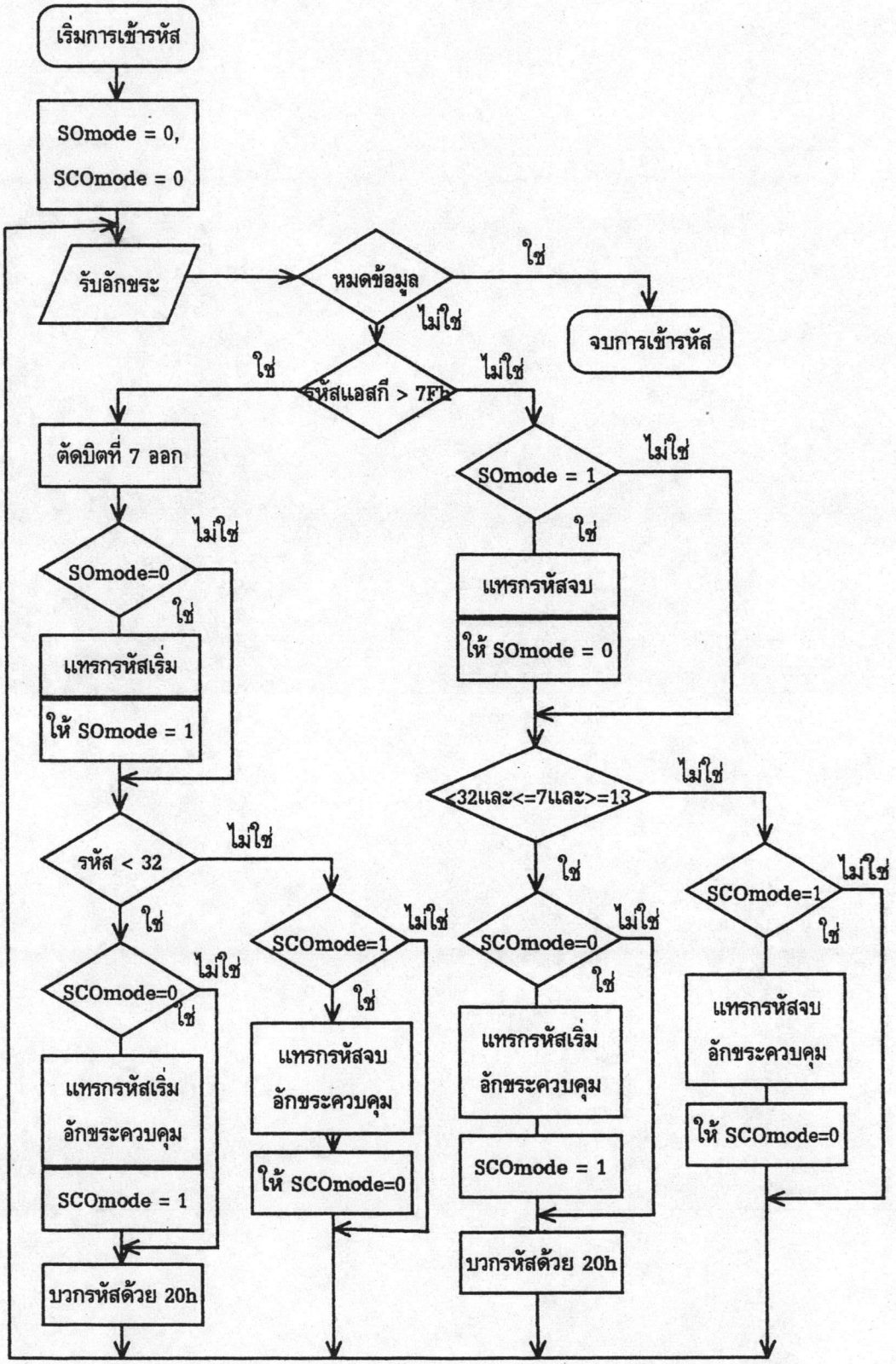
ขั้นตอนการแปลงข้อมูลเป็นรหัสกลางแสดงตามผังงานในรูป 3.5 เริ่มต้นจะกำหนดให้ตัวแปรแสดงชนิดของชุดอักขระ SOmode และ SCOmode ให้มีค่าเป็นศูนย์ อ่านอักขระเข้ามาตรวจสอบว่าเป็นอักขระที่มีค่ามากกว่า 7Fh ถ้าเป็นจะตัดบิตที่ 7 ทิ้งแล้วตรวจตัวแปร SOmode ว่าเป็นศูนย์หรือไม่ถ้าเป็นศูนย์จะแทรกรหัส SO ซึ่งเป็นลำดับอักขระหลีกบอกจุดเริ่มต้นของชุดอักขระไทยแล้วตั้งค่า SOmode ใหม่ให้เป็นหนึ่ง จากนั้นจะตรวจว่ารหัสที่ตัดบิตที่ 7 แล้วเป็นรหัสที่มีค่าน้อยกว่า 32 หรือ 20h หรือไม่ถ้าเป็นจะตรวจดูว่าขณะนั้นตัวแปร SOmode เป็นศูนย์หรือไม่ถ้าเป็นจะแทรกรหัส SCO ซึ่งเป็นลำดับอักขระหลีกบอกจุดเริ่มต้นของชุดอักขระควบคุมแล้วกำหนดให้ตัวแปรSCOmode มีค่าเป็นหนึ่ง จากนั้นบวกรหัสที่ได้ด้วย 20h เพื่อให้รหัสที่ได้ไม่เป็นรหัสควบคุมแล้วกลับไปรับอักขระใหม่ ในกรณีที่รหัสที่ตัดบิตที่ 7 แล้วมีค่ามากกว่าหรือเท่ากับ 32 หรือ 20h จะพิจารณาตัวแปร SCOmode ถ้ามีค่าเป็นหนึ่งจะแทรกรหัส SCIเพื่อบอกจุดจบของชุดอักขระควบคุมแล้วตั้งค่า SCOmode เป็นศูนย์แล้วกลับไปรับอักขระใหม่ ในกรณีที่เป็นอักขระที่น้อยกว่าหรือเท่ากับ 7Fh จะพิจารณาตัวแปร SOmode ถ้าเป็นหนึ่งจะแทรกรหัสจบชุดอักขระไทยแล้วตั้งตัวแปร SOmode เป็นศูนย์ จากนั้นพิจารณาว่าอักขระนั้นเป็นอักขระควบคุมที่ไม่ใช่ 07h หรือ 0Dh จะพิจารณาตัวแปร SCOmode ถ้าเป็นศูนย์จะแทรกรหัสเริ่มชุดอักขระควบคุมและตั้งตัวแปร SCOmode เป็นหนึ่งจากนั้นบวกรหัสด้วย 20h เพื่อไม่ให้เป็นอักขระควบคุม จากนั้นจะกลับไปรับอักขระใหม่ ในกรณีที่เป็นอักขระที่มีค่ามากกว่า 23 หรือเป็นรหัส 07h หรือ 0Dh จะพิจารณาตัวแปร SCOmode ถ้าเป็นหนึ่งแทรกลำดับอักขระหลีกบอกจบชุดอักขระควบคุม และลบล้างตัวแปร SCOmode เป็นศูนย์แล้วกลับไปรับอักขระใหม่ทำซ้ำจนหมดข้อมูลที่จะแปลงเป็นรหัสกลาง

การถอดรหัสแสดงตามผังงานในรูป 3.6 เริ่มต้นโดยลบล้างตัวแปร SOmode และ SCOmode จากนั้นรับอักขระนำมาพิจารณาว่าเป็นชุดอักขระหลีกทั้งสี่ข้างต้นหรือไม่ เมื่อเป็นชุดอักขระหลีกเริ่มต้นอักขระไทยจะตั้งตัวแปร SOmode เป็นหนึ่งเมื่อเป็นชุดอักขระหลีกจบชุดอักขระไทยจะตั้งตัวแปร SOmode เป็นศูนย์ เมื่อเป็นชุดอักขระหลีกเริ่มชุดอักขระควบคุมจะตั้งตัวแปร SCOmode เป็นหนึ่ง และเมื่อเป็นชุดอักขระหลีกจบชุดอักขระควบคุมจะตั้งตัวแปร SCOmode เป็นศูนย์ จากนั้นจะรับอักขระถัดไป ค่าของตัวแปร SOmode และ SCOmode แสดงในตาราง 3.2 ในรอบถัดมาถ้าอักขระดังกล่าวไม่ใช่ลำดับอักขระหลีกจะพิจาณาตัวแปร SCOmode ถ้ามีค่าเป็นหนึ่งแสดงว่ารหัสอยู่ในช่วงของอักขระควบคุม จะลบรหัสนั้นด้วย 20h จากนั้นจะพิจารณาตัว

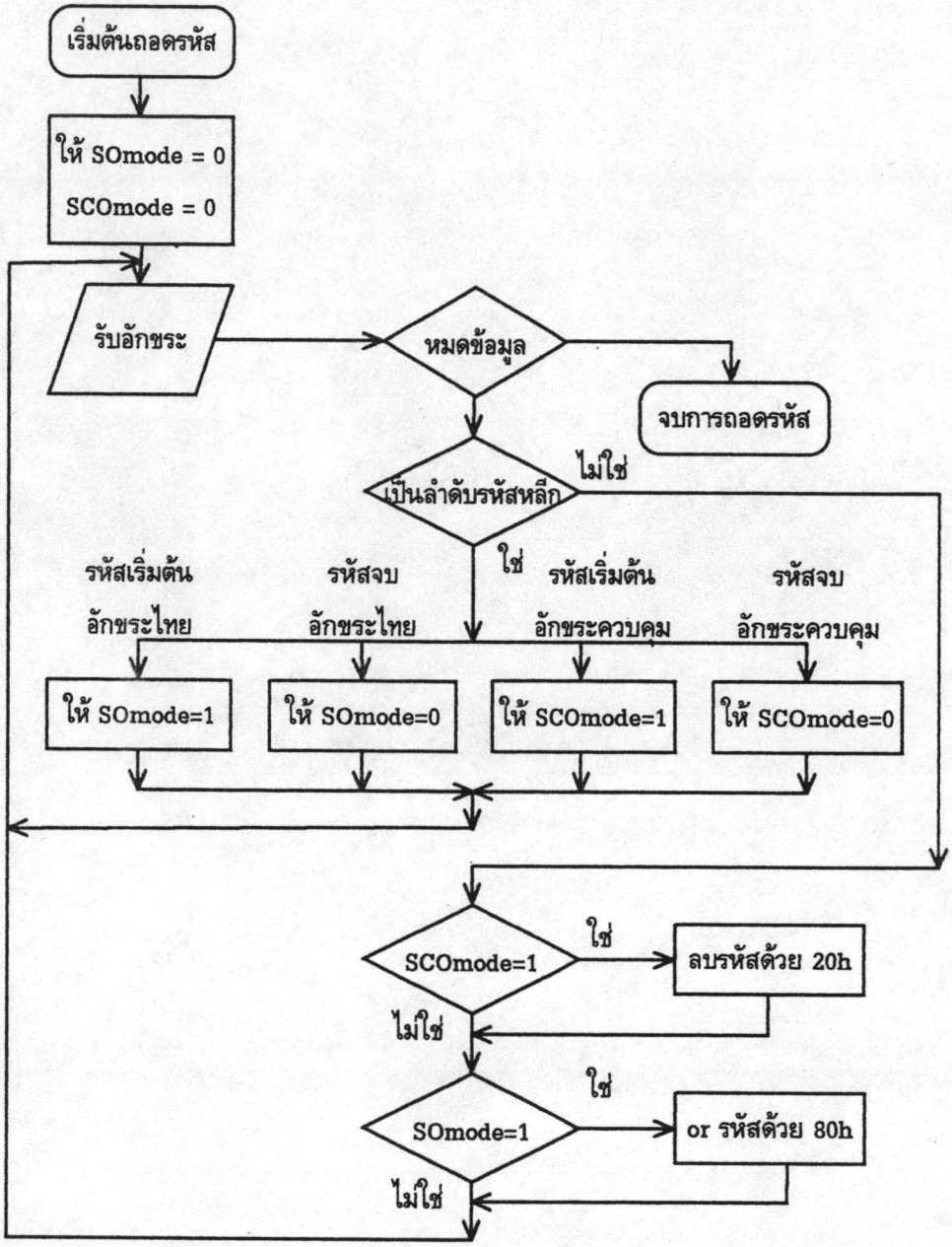
แปร SOmode ถ้ามีค่าเป็นหนึ่งแสดงว่าเป็นอักขระไทยดังนั้นจะนำรหัสมาดำเนินการทางตรรก "หรือ" กับเลข 80h เพื่อทำให้บิตที่ 7 มีค่าเป็นหนึ่ง สุดท้ายจะได้ผลลัพท์เป็นอักขระที่ได้รับการถอดรหัสแล้ว

รูป 3.7 แสดงตัวอย่างการแปลงเป็นรหัสกลางของข้อความ "Example ภาษาไทย encoding" โดยใช้วิธีแทรกรหัสพิเศษแบบที่ใช้ในวิจัยนี้ เริ่มแรกเมื่ออ่านข้อความพบคำแรกคือ "Example" ซึ่งเป็นภาษาอังกฤษ จะไม่มีการแปลงอักขระแต่อย่างใด เมื่ออ่านอักขระถัดมาพบว่าเป็นชุดอักขระที่มีค่ามากกว่า 7Fh คือคำว่า "ภาษาไทย" จะแทรกลำดับอักขระหลีกบอกจุดเริ่มชุดอักขระไทย ตามตาราง 3.1 จะเป็น 1Bh 7Bh 54h หรือ <ESC> { T ในเลขฐานสิบหก จากนั้นแปลงคำว่า "ภาษาไทย" ซึ่งมีรหัสตามมาตรฐานอุตสาหกรรมเป็น C0h D2h C9h D2h E4h B7h C2h จากนั้นตัดบิตที่8 ออกจะได้รหัสเป็น 40h 52h 49h 52h 64h 37h 42h ซึ่งจะตรงกับสายอักขระ @RIRd7B ตามด้วยลำดับอักขระหลีกบอกจบชุดอักขระไทยในงานวิจัยนี้ใช้รหัส 1Bh 7Bh 4Ch หรือ <ESC> { L ตามด้วยคำภาษาอังกฤษ "encoding" ในการถอดรหัส เมื่อพบรหัสกลางดังกล่าวด้วยวิธีตามผังงานในรูป 3.6 จะแปลงข้อความ "Example <ESC>{T@RIRd7B<ESC>{L encoding" เป็นข้อความ "Example ภาษาไทย encoding" โดยตัดลำดับอักขระหลีกออกและทำให้บิตที่ 7 ของสายอักขระ "@RIRd7B" เป็นหนึ่ง ดังนั้นสายอักขระดังกล่าวจะกลายเป็น "ภาษาไทย"

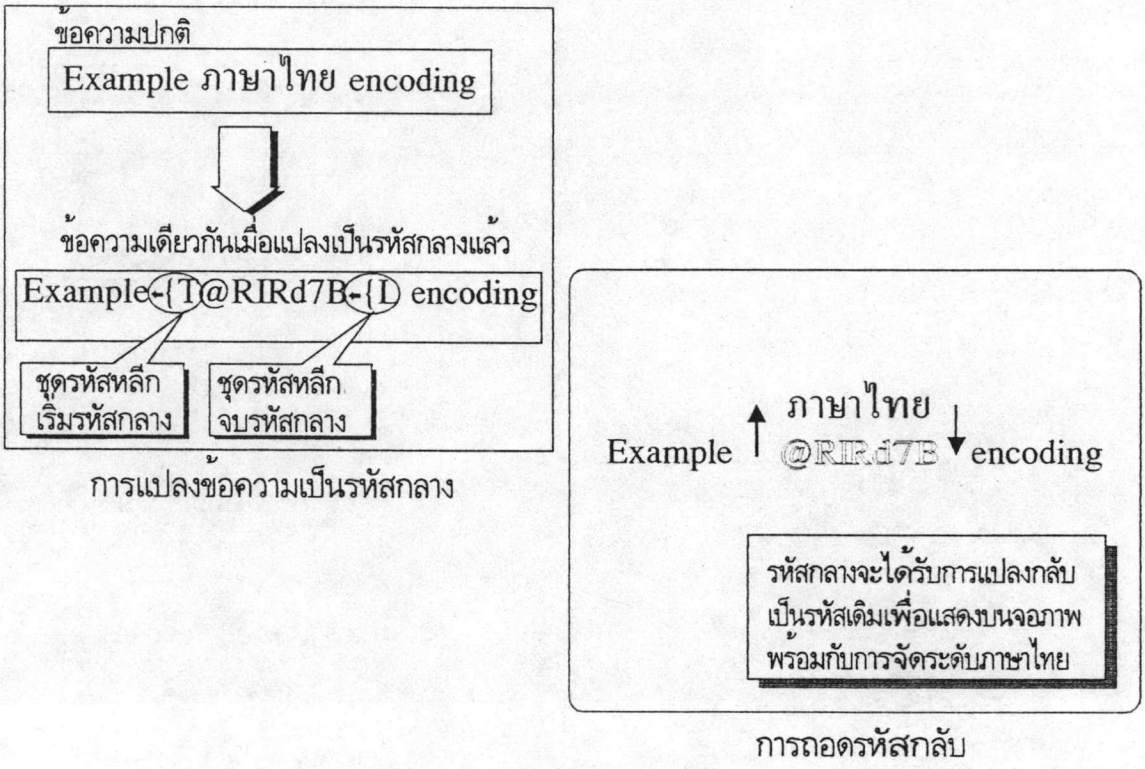
ข้อดีของวิธีนี้คือไม่จำเป็นต้องรอให้ข้อมูลเป็นกลุ่มก่อนที่จะแปลงรหัสเหมือนวิธี ของโปรแกรมยูยูเอ็นโคด และขนาดของข้อมูลรหัสกลางจะมีขนาดไม่ใหญ่กว่าข้อมูลเดิมมากนักส่วนข้อเสียที่สำคัญคือเมื่อมีชุดอักขระที่บอกจุดเริ่มต้นและสิ้นสุดตัวใดสูญหายไปจะทำให้การแปลงข้อมูลกลับผิดพลาด



รูป 3.5 ผังงานแสดงการแปลงเป็นรหัสกลางโดยแทรกลำดับอักขระหลีกที่ใช้ในวิทยานิพนธ์ฉบับนี้



รูป 3.6 ผังงานแสดงการถอดรหัสสำหรับวิธีแทรกลำดับอักขระหลีกที่ใช้ในวิทยานิพนธ์ฉบับนี้



รูป 3.7 แสดงตัวอย่างการแปลงข้อความเป็นรหัสกลางโดยวิธีแทรกลำดับอักขระหลีก