การออกแบบพจนานุกรมสำหรับการเรียนรู้ อักษรคัดลายมือไทย-อังกฤษอัตโนมัติ บนเครื่องไมโครคอมพิวเตอร์

A Design Of Dictionary For Automatic Handprinted Thai-English Characters Recognition On Microcomputer

นายชาย เกษมอมรกุล

CHAI KASEM-AMORNKUL

อาจารย์ที่ปรึกษา

รองศาสตราจารย์ ดร.ชม กิ้มปาน

ADVISOR

ASSOCIATE PROFESSOR DR.CHOM KIMPAN

วิทยานิพนธ์สำหรับปริญญาวิศวกรรมศาสตร์มหาบัณฑิต

สาขาวิศวกรรมไฟฟ้า

บัณฑิตวิทยาลัย

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2532

ISBN. 974-8154-42-4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บัณฑิตวิทยาลัย

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

แบบฟอร์มการให้คะแนนการสอบวิทยานิพนธ์

หัวข้อวิทยานิพนธ์ การออกแบบพจนานุกรม สำหรับการเรียนรู้อักษรคัดลายมือไทย-อังกฤษ อัตโนมัติบนเครื่องไมโครคอมพิวเตอร์(A Design of Dictionary for Automatic Handprinted Thai-English Characters Recognition on Microcomputer)

ชื่อนักศึกษา นายชาย เกษมอมรกุล **รหัสประจำตัว** 30126-038

หลักสูตร วิศวกรรมศาสตรมหาบัณฑิต **สาขาวิชา** วิศวกรรมไฟฟ้า

ภาควิชา วิศวกรรมคอมพิวเตอร์

อาจารย์ผู้ควบคุมวิทยานิพนธ์ รศ.ดร.ชม กิ้มปาน

ชื่ออาจารย์ผู้ควบคุมการสอบ	ลายมือชื่อ
รศ.ดร.ชม กิ้มปาน	
รศ.ประทีป บัญญัตินพรัตน์	
ดร.วรวัฒน์ ลิ้มโภคา	
ผศ.ดร.รัตติกร วราทูลศิริพันธุ์	
ผศ.ดร.บุญวัฒน์ อัตชู	

ค่าระดับคะแนนรวมที่เป็นเอกฉันท์จากคณะกรรมการสอบ G : GOOD (ดี)

วัน/เดือน/ปี ที่สอบ 8 สิงหาคม 2533 **เวลา** 13.30 น.

สถานที่สอบ ห้องประชุมคณะวิศวกรรมศาสตร์ (A-305)

บัณฑิตวิทยาลัยรับรองแล้ว

(นายศักชัย รัตนภาส)

รองอธิการบดีฝ่ายวิชาการ

รักษาราชการแทนคณบดีบัณฑิตวิทยาลัย

วันที่ 22 **เดือน** สิงหาคม **พ.ศ.** 2533

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทคัดย่อ

โดยทั่วไปลักษณะของอักษรทั้งภาษาไทยและภาษาอังกฤษนั้น จะประกอบไปด้วยลักษณะการวางตัวของเส้นในลักษณะต่างๆกัน เป็นโครงสร้างตามลักษณะเฉพาะของตัวอักษรตัวนั้นๆ ดังนั้นอาจกล่าวได้ว่าโครงสร้างของอักษรเป็นลักษณะเฉพาะอย่างหนึ่งของอักษร

วิทยานิพนธ์ฉบับนี้ได้นำเสนอถึง การออกแบบพจนานุกรมสำหรับการรู้จำอักษรคัดลายมืออัตโนมัติ โดยอาศัยลักษณะเด่นทางโครงสร้าง (Feature Concentration) มาใช้จำแนกตัวอักษร เพื่อให้เครื่องไมโครคอมพิวเตอร์สามารถรู้จำอักษรภาษาไทยและภาษาอังกฤษที่เขียนอยู่บนกระดาษ โดยป้อนผ่านเครื่องตรวจกวาดภาพ (Image Scanner) เข้ามา ส่งข้อมูลนั้นให้เครื่องไมโครคอมพิวเตอร์ แล้วส่งไปยังส่วนของการรู้จำเพื่อวิเคราะห์โครงร่างของข้อมูล ว่าเป็นตัวอักษรอะไร เมื่อรู้ว่าเป็นอักษรอะไรก็ให้รหัส ASCII ของอักษรตัวนั้น เป็นการทำงานแทนระบบป้อนข้อมูลของแป้นพิมพ์อักษรที่ใช้อยู่เดิม ไมโครคอมพิวเตอร์ก็จะนำรหัส ASCII ไปใช้งานต่อไป จากการศึกษาและทดลองได้ผลการรู้จำอักษรถูกต้องประมาณ 98% ด้วยความเร็วการรู้จำตัวอักษรประมาณ 1.2 วินาทีต่ออักษร จากฐานความรู้ของการรู้จำอักษร (พจนานุกรมการเรียนรู้) อักษรคัดลายมือภาษาไทยและภาษาอังกฤษจำนวน 147 อักษรๆละ 100 ตัว บนเครื่องไมโครคอมพิวเตอร์ AT ขนาด 16 bit ความเร็วของหน่วยประมวลผล 12 MHz

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Abstract

General characteristic of Thai and English characters compose of line segments lying with it own specific structure. With such feature , we can conclude that the structure of one character is the particular characteristic of that character.

This thesis proposes "A Design of Dictionary for Automatic Handprinted Thai-English Characters Recognition on Microcomputer". The algorithm consists of the several routines which is finally resulting successfully characters recognition. Thai and English characters, written on the paper, are extracted by the Hewlett Packard Scanner and become input data for the recognition section. Structure of each character is analyzed. ASCII code of that character is given once such character is recognized, otherwise the identification is requested for an automatic learning. Traditional data entry system thought the keyboard is now made easier by this algorithm. The experiment had been conducted, 147 handprinted Thai and English characters are tested 100 times each. Resulting in 98% of successfully character recognized with an average time-taken of 1.2 seconds per character which proceeded by 16 bits'microcomputer, PC/AT with clock speed at 12 MHz.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อ	2
Abstract	3
สารบัญ	4
สารบัญรูปประกอบ	6
สารบัญตารางประกอบ	9
บทที่ 1 บทนำ	10
เนื้อหาโดยรวมของวิทยานิพนธ์	12
บทที่ 2 วัตถุประสงค์ ประวัติความเป็นมาและขั้นตอนการรู้จำอักษร	13
2.1 วัตถุประสงค์	13
2.2 ประวัติความเป็นมาของการรู้จำรูปแบบอักษร	14
2.3 ขั้นตอนการรู้จำตัวอักษร	15
2.3.1 การแยกตัวอักษร	16
2.3.2 การรู้จำตัวอักษร	18
บทที่ 3 การแยกตัวอักษร	20
3.1 การแปลงข้อมูลภาพจากเครื่องตรวจกวาดภาพ	22
3.1.1 Details of Header Section	22
3.1.2 Details of the Directory	23
3.1.3 Organization of each Tag Entry	24
3.1.4 Bit-Map Image	25
3.2 การแยกแยะภาพอักษรหนึ่งหน้าออกเป็นบรรทัด	30
3.3 การแยกแยะภาพอักษรหนึ่งบรรทัดออกเป็นตัวอักษร	31
บทที่ 4 ทฤษฎีการรู้จำตัวอักษร	35
4.1 การสร้างรหัสเบื้องต้น (Initial Feature Extraction)	35
4.1.1 เนื้อตัวอักษร (P-code)	36
4.1.2 พื้นตัวอักษร (Q-code)	40

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

4.2 การปรับรหัสพื้นตัวอักษร (Unification) 44

4.3 การสร้างรหัสรวม (Concentration Code) 46

4.3.1 รหัสรวมที่ใช้เนื้อตัวอักษรเป็นหลัก 46

4.3.2 รหัสรวมที่ใช้พื้นตัวอักษรเป็นหลัก 47

4.3.3 การจัดโครงสร้างของรหัสรวม 48

บทที่ 5 การปรับปรุงการปรับรหัสพื้นตัวอักษรเพิ่มเติมและการสร้างพจนานุกรม 53

5.1 การปรับปรุงรหัสพื้นตัวอักษรเพิ่มเติม 53

5.2 การสร้างพจนานุกรมการรู้จำตัวอักษร 57

บทที่ 6 โปรแกรมการรู้จำอักษรคัดลายมือ 69

บทที่ 7 ผลการทดลองโปรแกรมการรู้จำอักษร 75

บทที่ 8 บทสรุปและข้อเสนอแนะ 80

กิตติกรรมประกาศ 82

หนังสืออ้างอิง 83

ภาคผนวก ก รหัสแอสกี้ของอักขระในรหัส สมอ. 87

ภาคผนวก ข การใช้งานโปรแกรมเครื่องตรวจกวาดภาพ 92

ภาคผนวก ค รูปแบบของ TIFF FILE 101

ภาคผนวก ง ตัวโปรแกรมการรู้จำอักษรคัดลายมือ 118

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูปประกอบ

หน้า

รูปที่ 1-1 บล็อกไดอะแกรมของระบบการรู้จำตัวอักษร 11

รูปที่ 2-1 แสดงลักษณะข้อมูลภาพหนึ่งหน้ากระดาษ 16

รูปที่ 2-2 แสดงลักษณะข้อมูลภาพหนึ่งบรรทัด 17

รูปที่ 2-3 แสดงข้อมูลภาพหนึ่งตัวอักษร 17

รูปที่ 3-1 รูปแบบกระดาษเพื่อการสร้างพจนานุกรม (ฐานความรู้) 21

รูปที่ 3-2 รูปแบบกระดาษเพื่อการทดสอบพจนานุกรม (ฐานความรู้) 21

รูปที่ 3-3 แสดงรายละเอียดโครงสร้างของ TIFF file 23

รูปที่ 3-4 แสดงรายละเอียด TIFF Header Section 23

รูปที่ 3-5 แสดงรายละเอียดของ Directory Section 24

รูปที่ 3-6 แสดงรายละเอียดข้อมูลแต่ละ Tag 25

รูปที่ 3-7 แสดงตัวอย่างในการพิจารณาโครงสร้างของข้อมูลภาพ 25

รูปที่ 3-8 แสดงข้อมูลที่ได้จากการแปลงข้อความจากเครื่องตรวจกวาดภาพ 29

รูปที่ 3-9 แสดงการแยกตัวอักษรออกจากกันทีละตัวอักษร 31

รูปที่ 3-10 แสดงผลของการแยกตัวอักษร 31

รูปที่ 3-11 (ก) แสดงโฟล์ชาร์ตของการแยกตัวอักษร 33

รูปที่ 3-11 (ข) แสดงโฟล์ชาร์ตของการแยกตัวอักษร (ต่อ) 34

รูปที่ 4-1 แสดงทิศทางที่ใช้ในการเข้าหาข้อมูล 4 ทิศทาง 35

รูปที่ 4-2 แสดงตัวอย่างข้อมูล 1 ตัวอักษร 36

รูปที่ 4-3 Initial Feature Extraction โดยพิจารณาเนื้อตัวอักษร 36

รูปที่ 4-4 (ก) แสดงการวางตัวในแนวนอน 37

รูปที่ 4-4 (ข) แสดงการวางตัวในแนวตั้ง 37

รูปที่ 4-4 (ค) แสดงการวางตัวในแนวทแยง 37

รูปที่ 4-4 (ง) แสดงการวางตัวในลักษณะที่ถูกล้อมรอบด้วย 1 37

รูปที่ 4-5 (ก) แสดงการพิจารณารหัสเบื้องต้นจากเนื้อตัวอักษร 39

รูปที่ 4-5 (ข) แสดงการแทนรหัสเบื้องต้นจากเนื้อตัวอักษร 39

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูปประกอบ(ต่อ)

หน้า

รูปที่ 4-6 Initial Feature Extraction โดยพิจารณาพื้นตัวอักษร 40

รูปที่ 4-7 (ก) แสดงการพิจารณารหัสเบื้องต้นจากพื้นตัวอักษร 42

รูปที่ 4-7 (ข) แสดงการแทนรหัสเบื้องต้นจากพื้นตัวอักษร 43

รูปที่ 4-8 แสดงข้อมูลที่ผ่านขั้นตอนของการกำหนดรหัสเบื้องต้น 43

รูปที่ 4-9 แสดงข้อมูลที่มีลักษณะไม่สมบูรณ์ 44

รูปที่ 4-10 แสดงข้อมูลที่ผ่านขั้นตอน การปรับรหัสพื้นตัวอักษร 46

รูปที่ 4-11 Concentration โดยการพิจารณาเนื้อตัวอักษร 47

รูปที่ 4-12 Concentration โดยการพิจารณาพื้นตัวอักษร 47

รูปที่ 4-13 (ก) แสดงการพิจารณาโครงสร้างของรหัสรวม 20 บิต 48

รูปที่ 4-13 (ข) แสดงการแจกแจงโครงสร้างของรหัสรวม 20 บิต 49

รูปที่ 4-14 (ก) แสดงการพิจารณาโครงสร้างของรหัสรวม 16 บิต 49

รูปที่ 4-14 (ข) แสดงการแจกแจงโครงสร้างของรหัสรวม 16 บิต 50

รูปที่ 4-15 แสดงการหารหัสรวม (Concentrated Word) 51

รูปที่ 4-16 (ก) แสดงการลดขนาดของรหัสรวม EEEES 51

รูปที่ 4-16 (ข) แสดงการลดขนาดของรหัสรวม OOEEE 52

รูปที่ 5-1 (ก) แสดงข้อมูลที่มีโครงสร้างโดยรวมไม่ชัดเจนของอักษร "ก" 54

รูปที่ 5-1 (ข) แสดงข้อมูลที่มีโครงสร้างโดยรวมไม่ชัดเจนของอักษร "ป" 54

รูปที่ 5-2 (ก) แสดงข้อมูลที่ผ่านการปรับรหัสพื้นตัวอักษรของอักษร "ก" 56

รูปที่ 5-2 (ข) แสดงข้อมูลที่ผ่านการปรับรหัสพื้นตัวอักษรของอักษร "ป" 56

รูปที่ 5-3 แสดงโครงสร้างของพจนานุกรม 64

รูปที่ 5-4 (ก) โฟล์ชาร์ตแสดงการทำงานของการรู้จำอักษรคัดลายมือ 67

รูปที่ 5-4 (ข) โฟล์ชาร์ตแสดงโปรแกรมการรู้จำอักษรคัดลายมือ 68

รูปที่ 6-1 การจัดการระบบการแสดงผลภาษาไทยด้วยโปรแกรม THAITIP 70

รูปที่ 6-2 แสดงรายละเอียดหน้าจอภาพของโปรแกรมการรู้จำ 71

รูปที่ 6-3 การสร้างพจนานุกรม 71

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูปประกอบ(ต่อ)

หน้า

รูปที่ 6-4 การเพิ่มเติมพจนานุกรม 72

รูปที่ 6-5 การทดสอบตัวอักษร 72

รูปที่ 6-6 แสดงการทดสอบอักษร 73

รูปที่ 6-7 แสดงการเรียนรู้ด้วยเครื่องไมโครคอมพิวเตอร์ 74

รูปที่ 7-1 แสดงเครื่องไมโครคอมพิวเตอร์และเครื่องตรวจกวาดภาพ 75

รูปที่ ข-1 แสดงชื่อโปรแกรม ปี ค.ศ. บริษัทที่ผลิต 93

รูปที่ ข-2 แสดง Window ของการ Edit และ Scan 94

รูปที่ ข-3 ตัวอย่างอักษรลายมือเขียนที่ใช้ 94

รูปที่ ข-4 แสดงการทำ Preview Scan 95

รูปที่ ข-5 การกำหนดเงื่อนไขของ Preview Scan 96

รูปที่ ข-6 แสดงการกำหนดขอบเขตของข้อมูลที่ต้องการเก็บ 96

รูปที่ ข-7 แสดงการทำ Final Scan 97

รูปที่ ข-8 แสดงการกำหนดเงื่อนไขของ Final Scan 98

รูปที่ ข-9 แสดงการเก็บข้อมูลลง File 98

รูปที่ ข-10 แสดงการออกจาก Scanner Window 99

รูปที่ ข-11 แสดง Image Editor Window 100

รูปที่ ค-1 แสดงการประยุกต์ใช้งาน TIFF File 101

รูปที่ ค-2 แสดงโครงสร้างของ TIFF File 102

รูปที่ ค-3 แสดงส่วนหัวเรื่องของข้อมูลภาพ 102

รูปที่ ค-4 แสดงส่วนเงื่อนไขข้อมูลภาพ 103

รูปที่ ค-5 แสดงส่วนแสดงขนาดของเงื่อนไขข้อมูลภาพ 104

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตารางประกอบ

หน้า

ตารางที่ 3-1 แสดงการพิจารณาโครงสร้างของแฟ้มข้อมูล 26-27

ตารางที่ 3-2 แสดงโครงสร้างทั่วไปของแฟ้มข้อมูล 27-29

ตารางที่ 4-1 การกำหนดค่ารหัส P 38

ตารางที่ 4-2 การกำหนดค่ารหัส Q 41

ตารางที่ 4-3 กฎการปรับค่ารหัส Q 45

ตารางที่ 5-1 แสดงการปรับรหัสพื้นตัวอักษรเพิ่มเติม 55

ตารางที่ 5-2 (ก) แสดงการแจกแจงความถี่ของรหัสร่วมของอักษร "ก" 57

ตารางที่ 5-2 (ข) แสดงการแจกแจงความถี่ของรหัสร่วมของอักษร "บ" 58

ตารางที่ 5-2 (ค) แสดงการแจกแจงความถี่ของรหัสร่วมของอักษร "B" 59

ตารางที่ 5-2 (ง) แสดงการแจกแจงความถี่ของรหัสร่วมของอักษร "D" 60

ตารางที่ 5-3 แสดงการแยกแยะอักษรด้วยรหัสร่วมที่มีความถี่สูงสุด 61

ตารางที่ 5-4 แสดงการแจกแจงความถี่ร่วมของอักษร 62

ตารางที่ 5-5 แสดงตัวอย่างรหัสแทนอักษรบางตัวในพจนานุกรม 63

ตารางที่ 5-6 ตารางแจกแจงความถี่ที่คุณสฤษดิ์ ปัตตะโชติ ใช้ในโครงการวิจัย 65

ตารางที่ 7-1 เปรียบเทียบในลักษณะของอักษรตัวพิมพ์ 76

ตารางที่ 7-2 เปรียบเทียบในลักษณะของอักษรลัดลายมือ 78

ตารางที่ 7-3 แสดงการเปรียบเทียบด้วยเงื่อนไขการสร้างพจนานุกรม 78

ตารางที่ 7-4 แสดงประสิทธิภาพของการรู้จำกับวิธีการรู้จำที่ใช้ 79

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

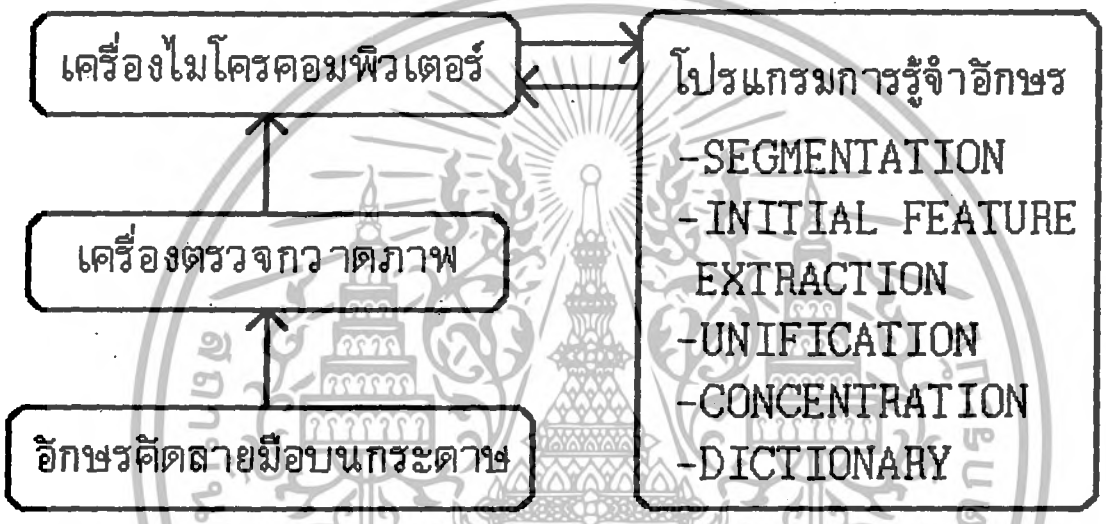
บทที่ 1
บทนำ

ในขณะที่โลกของเรากำลังหมุนไปทุกๆวินาทีนั้น ความก้าวหน้าของเทคโนโลยีต่างๆ ก็กำลังก้าวไปด้วยเช่นกัน ในเรื่องราวของคอมพิวเตอร์ก็ได้มีการเคลื่อนไหวอย่างไม่หยุดยั้ง ได้มีการพัฒนาในเรื่องราวต่างๆไปในทุกวิถีทางอย่างรีบเร่ง เพื่อให้ทันกับความก้าวหน้าของโลกเรา ในทุกวันนี้คอมพิวเตอร์ได้เข้ามามีบทบาทสำคัญในหลายๆเรื่อง มีการพัฒนาให้คอมพิวเตอร์สามารถเรียนรู้เรื่องราวต่างๆ เพื่อให้มีความสามารถสูงขึ้นและได้มีส่วนมาแบ่งเบาภาระอันหนักอึ้งของมนุษย์เรา ได้มีการนำเอาคอมพิวเตอร์ไปใช้งานในหลายๆลักษณะ ได้สร้างเครื่องไม้เครื่องมือตลอดจนอุปกรณ์ต่างๆ ซึ่งควบคุมด้วยไมโครโปรเซสเซอร์หรือคอมพิวเตอร์และให้สามารถทำงานเลียนแบบลักษณะการทำงานของคนเราให้มากที่สุด งานวิจัยในส่วนของวิทยานิพนธ์ฉบับนี้ก็เป็นอีกลักษณะหนึ่งที่พยายามที่จะทำให้เครื่องไมโครคอมพิวเตอร์สามารถเลียนแบบลักษณะการทำงานของมนุษย์ โดยให้เครื่องไมโครคอมพิวเตอร์สามารถรู้จำตัวอักษร หรือให้เครื่องไมโครคอมพิวเตอร์สามารถอ่านอักษรต่างๆได้ "การรู้จำเป็นคุณสมบัติประจำตัวอย่างหนึ่งของมนุษย์เรา ในระบบสิ่งมีชีวิตอย่างอื่นก็เหมือนกัน รูปแบบเป็นลักษณะการบอกวัตถุประสงค์ เราต้องจดจำสิ่งต่างๆ ทุกๆย่างก้าวในชีวิต การจดจำของคนเราเป็นระบบความรู้ที่เปลี่ยนแปลงไปมาซ่อนเงื่อนงำทันสมัยมาก ทั้งนี้เพราะเราสามารถจำรูปแบบในลักษณะต่างๆ อย่างยอดเยี่ยมได้" [1] จากการศึกษาลักษณะการรู้จำของมนุษย์เรานั้นเป็นกระบวนการที่ซับซ้อนมาก อย่างไรก็ตามเราก็ได้พยายามหาวิธีการต่างๆเพื่อจะสอนให้คอมพิวเตอร์เรียนรู้กระบวนการดังกล่าวได้อย่างคนเรา ซึ่งสามารถสรุปเป็นกระบวนการในการอ่านตัวอักษรหรืออ่านข้อความต่างๆออกได้โดยเริ่มตั้งแต่มีตัวอักษรหรือข้อความที่เขียนอยู่บนกระดาษเข้ามา และเริ่มรับรู้ตัวอักษรต่างๆด้วยการมองเห็นโดยใช้สายตามอง เพื่อเก็บลักษณะอักษรนั้นๆแล้วจึงเรียนรู้อักษรต่างๆเหล่านั้น หรือบอกความหมายของอักษรที่ได้เห็นด้วยการใช้ส่วนของสมอง ซึ่งจากลักษณะการเรียนรู้ของมนุษย์ในลักษณะดังกล่าว สามารถที่จะนำมาประยุกต์แทนโดยใช้อุปกรณ์ทางด้านอิเลกทรอนิก เพื่อเป็นการแบ่งเบาภาระการทำงานของมนุษย์เรา ในรูปของการทำงานแทนโดย ในส่วนของการมองเห็นด้วยตาเพื่อเก็บลักษณะของอักษรหรือภาพต่างๆนั้น สามารถแทนได้ด้วยเครื่องตรวจกวาดภาพ (Image Scanner) ซึ่งมีความสามารถในการเก็บสัญญาณแสงที่ได้จากการสะท้อนกลับของการเปล่งแสงไปยังอักษรหรือภาพที่มากั้น และแปลงสัญญาณแสงที่ได้เป็นสัญญาณไฟฟ้า ซึ่งแสดงลักษณะของอักษรหรือภาพที่ผ่านเข้ามา และในส่วนของสมองก็สามารถแทนได้ในส่วนของสมองกลเครื่องไมโครคอมพิวเตอร์รวมกับโปรแกรมการรู้จำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อักษร ซึ่งเป็นส่วนในการนำเอาสัญญาณไฟฟ้าที่ได้จากเครื่องตรวจกวาดภาพมาพิจารณาจัดลำดับประมวลผล โดยมีวิธีการและขั้นตอนในการพิจารณาจากโปรแกรมในการรู้จำตัวอักษร สามารถเขียนแสดงการทำแทนการมองเห็นและการเรียนรู้ของมนุษย์ ในรูปของส่วนประกอบของระบบการรู้จำอักษรโดยเครื่องไมโครคอมพิวเตอร์ ดังในรูปที่ 1-1

<u>ส่วนประกอบของระบบ</u>



รูปที่ 1-1 บล็อกไดอะแกรมของระบบการรู้จำตัวอักษร

ในส่วนของงานวิจัยให้ คอมพิวเตอร์สามารถรู้จำตัวอักษร (Characters Recognition) เป็นแนวทางในการพัฒนาให้คอมพิวเตอร์มีความสามารถเลียนแบบมนุษย์ และสามารถนำไปประยุกต์ใช้งาน โดยเป็นส่วนของการป้อนข้อมูลในลักษณะที่เป็นอัตโนมัติ (Auto Matic Data Entry) วิทยานิพนธ์ฉบับนี้จะเน้นถึง การออกแบบพจนานุกรมการรู้จำตัวอักษรคัดลายมือทั้งภาษาไทยและภาษาอังกฤษ โดยการพิจารณาโครงสร้างของอักษรแต่ละตัว ทั้งส่วนที่เป็นเนื้อตัวอักษรและส่วนที่เป็นพื้นหลังตัวอักษร แล้วนำมาหารหัสรวม (Concentrate Code) แล้วพิจารณารหัสรวมเพื่อให้เป็นลักษณะที่ใช้บอกลักษณะเฉพาะของอักษรแต่ละตัว ซึ่งมีรายละเอียดเนื้อหาแสดงในบทต่างๆ ดังต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1 เป็นบทนำ กล่าวถึงเนื้อหาโดยรวมของวิทยานิพนธ์และโครงสร้างของระบบการรู้จำอักษร

บทที่ 2 กล่าวถึงวัตถุประสงค์ ประวัติความเป็นมา ขั้นตอนของการรู้จำอักษร

บทที่ 3 กล่าวถึงวิธีการในการแยกตัวอักษร (Segmentation)

บทที่ 4 กล่าวถึงทฤษฎีการรู้จำตัวอักษร

1. การสร้างรหัสเบื้องต้น (Initial Feature Extraction)
2. การปรับรหัสพื้นตัวอักษร (Unification)
3. การสร้างรหัสรวม (Concentration)

บทที่ 5 กล่าวถึงการปรับปรุงการปรับรหัสพื้นตัวอักษรเพิ่มเติมและการสร้างพจนานุกรม

บทที่ 6 กล่าวถึงโปรแกรมการรู้จำอักษรคัดลายมือ

บทที่ 7 กล่าวถึงผลการทดลองโปรแกรมการรู้จำตัวอักษร และวิเคราะห์ผลการทดลอง

บทที่ 8 เป็นบทสุดท้าย กล่าวถึงบทสรุป และข้อเสนอแนะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
วัตถุประสงค์ ประวัติความเป็นมาและขั้นตอนการรู้จำอักษร

ในส่วนของการสอนให้คอมพิวเตอร์สามารถรู้จำอักษรต่างๆนั้น ได้มีการศึกษาและทำการวิจัยในเรื่องราวต่างๆที่เกี่ยวข้องกันมาอย่างแพร่หลาย ได้มีการเรียนรู้และสร้างทฤษฎีของการรู้จำกันอย่างมากมาย ซึ่งในแต่ละวิธีการก็จะมีความเหมาะสมและใช้แก้ปัญหาที่เกิดขึ้น ดังจะเห็นได้จากหนังสือ นิตยสาร และวารสารทั้งในประเทศและต่างประเทศ [1]-[19] ได้ลงบทความและกล่าวถึงเรื่องราวของการรู้จำตัวอักษร (Pattern Recognition) ในส่วนของการศึกษาก็เน้นกันหลายๆวิธีการด้วยกัน โดยอาจแบ่งโดยพิจารณาจากลักษณะอักษรในรูปแบบที่แตกต่างกัน เช่นการรู้จำอักษรตัวพิมพ์ [3]-[8],[10]-[12],[14],[15],[18],[20],[22],[23],[26],[27] การรู้จำอักษรลายมือเขียน [2],[9],[13],[16],[17],[19],[21],[24],[25],[28],[29] หรือโดยอาจแบ่งโดยการพิจารณาจากวิธีการในการรู้จำ เช่นการรู้จำตัวอักษรโดยวิธีการแบบซ้อนทับ (Matching) [3]-[8],[11],[13],[16],[20]-[23],[25],[26] การรู้จำตัวอักษรโดยวิธีการพิจารณาโครงสร้างของตัวอักษร [2],[9],[10],[14],[17]-[19],[24],[26]-[29] ซึ่งแต่ละวิธีการก็มีลักษณะที่เด่นซึ่งสามารถอธิบายลักษณะอักษร หรือรู้จำอักษรแต่ละประเภท และสามารถใช้แก้ปัญหาได้แตกต่างกันออกไป ส่วนของงานวิจัยในวิทยานิพนธ์ฉบับนี้ ส่วนของการรู้จำอักษรได้ทำการประมวลผลโดยเลือกใช้วิธีการรู้จำโดยการพิจารณาโครงสร้างของตัวอักษร โดยการพิจารณาทั้งในส่วนของเนื้อตัวอักษรและส่วนที่เป็นพื้นตัวอักษร แล้วพิจารณาหาลักษณะเด่นโดยการแจกแจงความถี่หารหัสรวมที่เกิดขึ้นมากโดยวิธีการทางสถิติ จากรหัสรวม (Concentrate Code) ซึ่งเกิดจากการรวมรหัสในส่วนของเนื้อตัวอักษรและพื้นตัวอักษรเข้าด้วยกัน เพื่อการรู้จำอักษรจากอักษรคัดลายมือทั้งอักษรภาษาไทยและอักษรภาษาอังกฤษ

2.1 วัตถุประสงค์

วิทยานิพนธ์ฉบับนี้มีวัตถุประสงค์ เพื่อเป็นการพัฒนาระบบของการป้อนข้อมูลอัตโนมัติโดยใช้เครื่องไมโครคอมพิวเตอร์ ร่วมกับอุปกรณ์สนับสนุนอย่างอื่นๆ เช่นเครื่องตรวจกวาดภาพ (Image Scanner) เป็นต้น แทนการป้อนข้อมูลด้วยแป้นพิมพ์ (Key Board) อีกทั้งยังเป็นการศึกษาเปรียบเทียบกับวิธีการของการรู้จำตัวอักษรระหว่างวิธีการที่นำเสนอกับวิธีการอื่นที่ใกล้เคียงกัน โดยวิทยานิพนธ์ฉบับนี้จะได้นำเสนอวิธีการของการรู้จำอักษร โดยใช้วิธีการพิจารณาโครง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สร้างของอักษรจากลักษณะเด่นของรหัสรวม (Feature Concentration) โดยพิจารณาหาลักษณะรหัสรวม (Concentrate code) เป็นลักษณะเด่นซึ่งใช้อธิบายลักษณะอักษรตัวนั้นๆเก็บไว้เป็นพจนานุกรม(ฐานความรู้) ของการรู้จำอักษร และนำเสนอการจัดทำพจนานุกรมอัตโนมัติของการรู้จำอักษร โดยเป็นพจนานุกรมที่มีลักษณะที่ยืดหยุ่น สามารถเปลี่ยนแปลงและสามารถเพิ่มเติมพจนานุกรมได้เองจากการเรียนรู้อักษรที่มีลักษณะใหม่ ซึ่งไม่สอดคล้องกับลักษณะเดิมที่มีอยู่ และสามารถนำวิธีการนี้ไปประยุกต์ใช้งานในเรื่องอื่นๆ เช่น การจดจำลายนิ้วมือ และอื่นๆได้อีกต่อไป

2.2 ประวัติความเป็นมาของการรู้จำรูปแบบอักษร

เรื่องราวของการรู้จำรูปแบบอักษร (Characters Pattern Recognition) นั้นได้มีการศึกษากันอย่างกว้างขวางในหลายปีที่ผ่านมา ทั้งในรูปแบบอักษรที่เป็นตัวพิมพ์(Printed Characters) และรูปแบบของอักษรคัดลายมือ (Handprinted Characters) ได้มีการศึกษาและทำการวิจัยทั้งในประเทศและต่างประเทศ ดังปรากฏผลงานที่ได้ตีพิมพ์ในวารสารหรือเอกสารประชุม สัมมนา วิชาการต่างๆ [2]-[29] ซึ่งผลที่ได้จากการใช้เทคนิค ตลอดถึงวิธีการต่างๆนั้นก็ขึ้นอยู่กับลักษณะของอักษรที่ใช้ เทคนิควิธีการรู้จำอักษรและเงื่อนไข ตลอดจนข้อจำกัดของการใช้งานนั้นๆ

เทคนิคและวิธีการในการรู้จำอักษรที่ใช้กันอยู่สามารถแบ่งออกได้เป็นประเภทใหญ่ๆ ได้ 2 วิธี ด้วยกันคือ

1. วิธีการซ้อนทับ (Matching Method)

ในส่วนวิธีการรู้จำอักษรด้วยวิธีการแบบซ้อนทับนี้ ถือได้ว่าเป็นวิธีการรู้จำอย่างธรรมชาติ โดยหลักการแล้วการรู้จำตัวอักษรด้วยวิธีการนี้ จะไม่สนใจพิจารณาถึงรายละเอียดของตัวอักษร เช่น เนื้อตัวอักษร พื้นเบื้องหลังตัวอักษร เป็นต้น แต่จะสนใจพิจารณาเพียงลักษณะความคล้ายคลึง หรือลักษณะความเหมือนกันของตัวอักษรเท่านั้น ซึ่งโดยทั่วไปวิธีการรู้จำตัวอักษรด้วยการซ้อนทับนี้จะมีขั้นตอนในการรู้จำที่ค่อนข้างซับซ้อน และใช้คณิตศาสตร์ในการประมวลผลค่อนข้างมากเมื่อเทียบกับวิธีการอื่น แต่จะใช้ได้ผลดีกับตัวอักษรที่มีลักษณะคงตัวมากๆ เช่นอักษรตัวพิมพ์ หรือกับอักษรที่มีลักษณะไม่สมบูรณ์ เช่น ลักษณะของเนื้อตัวอักษรเว้า แหว่ง มีสัญญาณรบกวน ซึ่งจะทำให้ตัวอักษรมีลักษณะที่เปลี่ยนไป แต่ด้วยวิธีการรู้จำแบบนี้จะไม่สนใจกับลักษณะที่เกิดขึ้น แต่จะพิจารณาเฉพาะความคล้ายคลึงกับอักษรที่มีอยู่เท่านั้น ซึ่งเป็นลักษณะเด่นของวิธีการนี้

2. วิธีการวิเคราะห์โครงสร้างของอักษร (Structural Analysis)

ในส่วนของวิธีการรู้จำอักษรแบบนี้จะมีลักษณะที่ตรงข้ามกับวิธีการแบบซ้อนทับ โดยจะให้ความสนใจและความสำคัญกับลักษณะที่เป็น ส่วนเนื้อตัวอักษรและส่วนพื้นเบื้องหลังตัวอักษร เช่น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การพิจารณาในส่วนที่เป็นวงปิด หรือหัวตัวอักษรในภาษาไทย การพิจารณาในส่วนของจุดแยก จุดต่อ จุดตัด และจุดปลาย การพิจารณาลักษณะของลายเส้น(เนื้อตัวอักษร) เป็นต้น ซึ่งการพิจารณาโครงสร้างตัวอักษรในลักษณะดังกล่าวจะสามารถใช้ได้ผลดีกับอักษรในหลายๆลักษณะ แต่จะไม่ได้ผลมากนักกับอักษรที่มีลักษณะที่ไม่สมบูรณ์ เช่นอักษรที่มีลักษณะเว้า แหว่ง ขาด หรือมีสัญญาณรบกวน ซึ่งทำให้รูปร่างของอักษรเปลี่ยนไป ก็จะทำให้การพิจารณาผิดพลาดไปด้วย

ในส่วนประเทศของเรานั้นก็ได้มีการศึกษากันอย่างกว้างขวางในหลายๆรูปแบบเช่นเดียวกัน จากหลักการใหญ่ของวิธีการในการรู้จำตัวอักษรจะเห็นได้ว่า ได้มีการประยุกต์และพัฒนาวิธีการต่างๆออกมาอย่างมากมาย ดังจะเห็นได้จากผลงานที่ได้ตีพิมพ์เป็นบทความ สัมมนาประชุมวิชาการ [2]-[19] ในวาระต่างๆ วิทยานิพนธ์ [20]-[27] วิทยานิพนธ์มหาบัณฑิต [28], [29] ทั้งในรูปแบบของอักษรตัวพิมพ์ภาษาไทย [3]-[8],[10]-[12],[14],[15],[18], [20],[22],[23],[26],[27] และอักษรคัดลายมือภาษาไทย [9],[13],[16],[17],[19] ,[21],[24],[25],[28],[29] ซึ่งจากผลการศึกษาและวิจัยเกี่ยวกับการรู้จำอักษรในหลายๆวิธีก็ได้มีการปรับปรุงพัฒนาขึ้นมาตามลำดับ คาดว่าจะสามารถนำไปประยุกต์ใช้งานได้ในระยะเวลาอันใกล้นี้ ในส่วนวิทยานิพนธ์ฉบับนี้ ได้ศึกษาและเลือกใช้วิธีการรู้จำอักษรโดยการวิเคราะห์โครงสร้างของอักษร (Structural Analysis) ด้วยวิธีการพิจารณารหัสรวมที่เกิดจากเนื้อตัวอักษรและพื้นเบื้องหลังตัวอักษร ซึ่งเรียกว่า Feature Concentration โดยการหารหัสเบื้องต้นแทนแต่ละจุดภาพทั้งที่เป็นเนื้อตัวอักษรและพื้นตัวอักษร ด้วยวิธีการ Initial Feature Extraction แล้วพิจารณาปรับรหัสที่แทนพื้นตัวอักษร ด้วยวิธีการ Unification แล้วพิจารณาหารหัสรวมของเนื้อตัวอักษรและพื้นตัวอักษรที่ใช้แทนจุดภาพนั้นๆ แล้วใช้วิธีการทางสถิติ โดยการทดสอบแจกแจงความถี่ของรหัสรวมที่เกิดขึ้นในแต่ละตัวอักษร ซึ่งปรากฏว่ารหัสรวมที่มีความถี่สูงสุด 5 อันดับแรก สามารถใช้เป็นลักษณะเด่น สามารถแยกแยะอักษรออกจากกันได้

2.3 ขั้นตอนการรู้จำตัวอักษร

แบ่งออกเป็น 2 ส่วนใหญ่ๆด้วยกันคือ

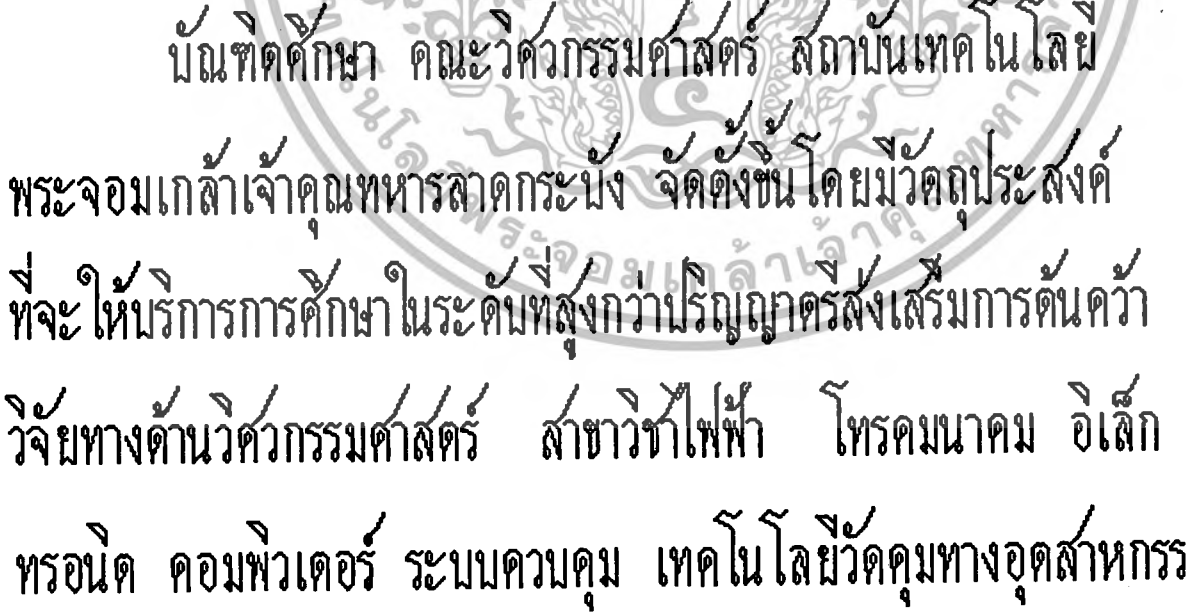
1. การแยกตัวอักษรแต่ละตัวออกมาจากข้อมูลทั้งหมด (Segmentation)
2. การรู้จำอักษร (Recognition)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.1 การแยกตัวอักษร (Character Segmentation)

เป็นกระบวนการในการเตรียมข้อมูลเบื้องต้นก่อนเข้าสู่กระบวนการรู้จำอักษร โดยเมื่อป้อนแฟ้มข้อมูลภาพที่ได้จากเครื่องตรวจกวาดภาพ ต้องทำการแปลงข้อมูลที่ได้ให้อยู่ในรูปข้อมูลเพื่อการประมวลผลเสียก่อน โดยอยู่ในรูปของรหัสแทน 0 ซึ่งใช้แทนพื้นตัวอักษรและ 1 ซึ่งใช้แทนเนื้อตัวอักษร แล้วจึงทำการจัดแบ่งแยกภาพอักษรที่ไม่ต่อเนื่องออกจากกัน และจัดลำดับการเข้าประมวลผล โดยมีขั้นตอนดังต่อไปนี้

1. หลังจากแปลงข้อมูลให้อยู่ในรูปรหัสแทนทั้ง 0 และ 1 ซึ่งแทนข้อมูลภาพอักษรหนึ่งหน้ากระดาษ ต้องทำการแบ่งข้อมูลภาพอักษรหนึ่งหน้ากระดาษออกจากกันเป็นบรรทัด(ตามแนวนอน) โดยอาศัยช่องว่างของข้อมูลตามแนวนอน ซึ่งจากข้อมูลภาพหนึ่งหน้ากระดาษซึ่งแทนรหัสในส่วนของเนื้อตัวอักษรด้วย 1 และส่วนที่เป็นพื้นอักษรด้วย 0 ในการแยกแต่ละบรรทัดอักษรออกจากกันก็โดยการพิจารณาหารหัสที่เป็น 0 ที่มีการต่อเนื่องกันตามแนวนอนเท่ากับความยาวของภาพอักษร โดยแยกการพิจารณาในส่วนของระยะด้านบนของบรรทัดข้อความ และส่วนของระยะด้านล่างของแต่ละบรรทัด และทำการพิจารณาต่อเนื่องไปจนกว่าจะสิ้นสุดข้อมูลภาพหนึ่งหน้ากระดาษ ซึ่งสามารถแสดงลักษณะของข้อความหนึ่งหน้ากระดาษจากอักษรคัดลายมือดังในรูปที่ 2-1



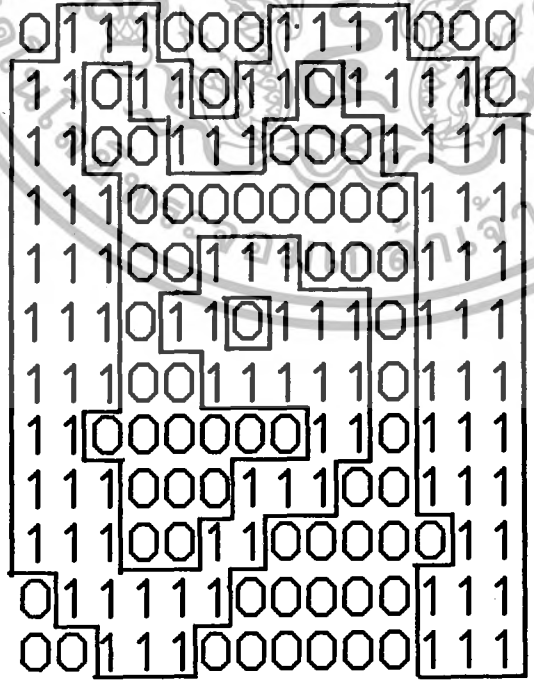
รูปที่ 2-1 แสดงลักษณะข้อมูลภาพหนึ่งหน้ากระดาษ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. หลังจากการพิจารณาหาระยะด้านบนและระยะด้านล่าง เป็นขอบเขตของข้อมูลแต่ละบรรทัดแล้ว ดังแสดงในรูปที่ 2-2 ต้องทำการแยกข้อมูลของอักษรแต่ละตัวออกจากกัน โดยอาศัยวิธีการคล้ายกับการแยกข้อมูลภาพแต่ละบรรทัด โดยอาศัยช่องว่างระหว่างอักษรแต่ละตัว พิจารณาหารหัส 0 ซึ่งแทนส่วนที่เป็นพื้นตัวอักษรในแนวตั้ง โดยพิจารณาในช่วงของระยะบรรทัด (ระยะด้านบนและระยะด้านล่างของบรรทัด) และพิจารณาหาระยะทางด้านหน้าและระยะทางด้านหลังของแต่ละตัวอักษร ซึ่งจะได้ระยะทั้ง 4 ทิศทาง (ระยะด้านบน ระยะด้านล่าง ระยะด้านหน้าและระยะด้านหลัง) เป็นขอบเขตของอักษรแต่ละตัว ดังแสดงในรูปที่ 2-3 และพิจารณาต่อเนื่องไปจนกว่าจะหมดขอบเขตของบรรทัด จึงไปพิจารณาในบรรทัดถัดไป

บัณฑิตศึกษา คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยี
พระจอมเกล้าเจ้าคุณทหารลาดกระบัง จัดตั้งขึ้นโดยมีวัตถุประสงค์

รูปที่ 2-2 แสดงลักษณะข้อมูลภาพหนึ่งบรรทัด



รูปที่ 2-3 แสดงข้อมูลภาพหนึ่งอักษร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.2 การรู้จำอักษร (Recognition)

เป็นกระบวนการที่ต่อเนื่องจากกระบวนการแยกตัวอักษร (Segmentation) โดยเมื่อได้ข้อมูลอักษรที่แทนอักษรหนึ่งตัวจากกระบวนการข้างต้นแล้ว ดังในรูปที่ 2-3 จะพิจารณาหารหัสแทนทั้งในส่วนที่เป็นเนื้อตัวอักษรและพิจารณาหารหัสแทนในส่วนที่เป็นพื้นตัวอักษร แล้วนำรหัสเหล่านี้ไปหารหัสเฉพาะเพื่อใช้แทนอักษรตัวนั้นๆ และแยกแยะอักษรแต่ละตัวออกจากกัน โดยมีขั้นตอนดังต่อไปนี้

1. เมื่อได้ข้อมูลของอักษรแต่ละตัว ก็พิจารณาหารหัสเบื้องต้น (Initial Feature Extraction) แทนทุกๆจุด โดยการหารหัสแทนทั้งในส่วนเนื้อตัวอักษร โดยการพิจารณาการวางตัวของจุดที่อยู่ติดทั้ง 4 ทิศทางจะได้รหัสแทนออกมา 4 แบบด้วยกันคือ h,v,s,i และหารหัสแทนส่วนพื้นตัวอักษร โดยการพิจารณาการวางของจุดที่แทนเนื้อตัวอักษรตัวแรกในแต่ละทิศทางทั้ง 4 ทิศ ซึ่งจะได้รหัสแทนออกมา 16 แบบด้วยกันคือ 0,1,...,F

2. ทำการปรับรหัสแทนพื้นตัวอักษร (Unification) ให้เหมาะสม และทำการรวมรหัสแทนเนื้อตัวอักษรและพื้นตัวอักษรเข้าด้วยกัน (Concentration) เพื่อให้ได้รหัสรวมตามการจัดโครงสร้างรวมของแต่ละจุด ซึ่งเรียกว่า Concentrate Code ซึ่งต้องทำทุกๆจุดภาพ

3. พิจารณาเฉพาะรหัสรวม (Concentrate Code) โดยทำการทดสอบทางสถิติ ด้วยการแจกแจงความถี่ของรหัสรวมที่ได้ของอักษรทุกๆตัว ซึ่งจากผลการแจกแจงพบว่า รหัสรวมที่มีความถี่จากการแจกแจงสูงสุด 5 อันดับแรก ซึ่งแสดงลักษณะโครงสร้างส่วนใหญ่ของอักษรตัวนั้นเป็นลักษณะเฉพาะของอักษรแต่ละตัว สามารถใช้แยกแยะตัวอักษรตัวนี้ออกจากอักษรตัวอื่นๆได้

4. การสร้างพจนานุกรม (ฐานความรู้) ของการรู้จำตัวอักษรแต่ละตัว โดยการเก็บรหัสรวมที่ใช้แทนอักษรแต่ละตัว พร้อมทั้งเก็บรหัสแอสกี้ของอักษรตัวนั้น และทำการทดสอบหาและเก็บลักษณะอักษรดังกล่าวแล้วทุกๆตัวอักษรที่กำหนดไว้ทั้ง อักษรคัดลายมือภาษาไทยและภาษาอังกฤษไว้รวมกันในแฟ้มข้อมูลใหม่ในรูปของฐานความรู้ของการรู้จำอักษร

5. การทดสอบพจนานุกรม (ฐานความรู้) ของการรู้จำอักษรที่สร้างขึ้น ด้วยตัวอักษรคัดลายมือที่เขียนขึ้นใหม่ โดยหากอักษรที่เขียนขึ้นใหม่มีลักษณะที่ทดสอบแล้ว ไม่มีลักษณะ (รหัสรวม) ที่ตรงในฐานความรู้ของการรู้จำจะให้แสดงภาพของอักษรตัวนั้นทางจอภาพ เพื่อให้ผู้ใช้ได้ตัดสินใจหรือป้อนความรู้ใหม่ และนำไปเก็บรวมไว้กับฐานความรู้ที่มีอยู่เดิม หากอักษรที่เขียนขึ้นใหม่มีลักษณะตรงกับฐานความรู้ที่มีอยู่ก็ให้แสดงภาพของรหัสแอสกี้ของอักษรตัวนั้นออกมา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากกระบวนการของการรู้จำตัวอักษรของงานวิจัยในวิทยานิพนธ์ที่ได้กล่าวแล้วนั้น สามารถชี้ให้เห็นลักษณะเด่นของวิธีการของการรู้จำ โดยเป็นกระบวนการรู้จำตัวอักษรด้วยการพิจารณาลักษณะโครงสร้างตัวอักษร จากการพิจารณารหัสรวมซึ่งได้จากการพิจารณารหัสในส่วนที่เป็นเนื้อตัวอักษรและส่วนที่เป็นพื้นตัวอักษร โดยพิจารณารหัสเด่นที่ใช้แทนอักษรแต่ละตัว จากการทดลองแจกแจงความถี่รหัสรวมที่เกิดในแต่ละตัวอักษร ซึ่งได้ผลสรุปเป็นรหัสรวมที่มีความถี่สูงสุด 5 อันดับแรกใช้แทนอักษรแต่ละตัว แล้วเก็บลักษณะแทนดังกล่าวรวมกันเป็นพจนานุกรมของการรู้จำอักษร และลักษณะเด่นของงานวิจัยนี้อีกคือ ในส่วนของการจัดทำพจนานุกรมนั้นจะมีลักษณะเป็นพจนานุกรมที่มีความยืดหยุ่น สามารถให้เครื่องไมโครคอมพิวเตอร์เรียนรู้อักษรในลักษณะใหม่ๆได้เอง โดยอัตโนมัติและจะเก็บรหัสแทนที่ได้จากการเรียนรู้ใหม่เพิ่มเติม รวมกับพจนานุกรมที่มีอยู่ก่อนได้โดยตรงเป็นการเพิ่มประสิทธิภาพการรู้จำตัวอักษรให้กับโปรแกรมการรู้จำตัวอักษรต่อๆไปอีก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
การแยกตัวอักษร

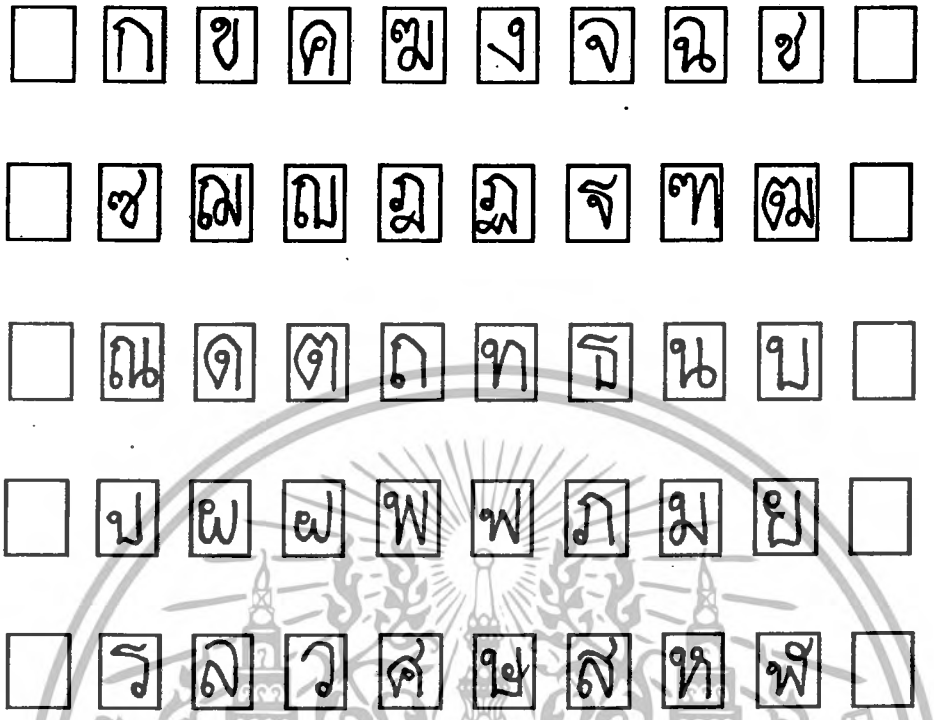
กระบวนการที่สำคัญอย่างหนึ่งของการรู้จำตัวอักษร ซึ่งมีความสำคัญไม่น้อยกว่ากระบวนการอื่นๆ ก็คือกระบวนการในการแยกตัวอักษร (Segmentation) ซึ่งเป็นกระบวนการในการแยกภาพอักษรออกจากข้อความรวมหรือประโยค ออกมาทีละตัวอักษรเพื่อนำไปเข้าสู่กระบวนการรู้จำอักษรซึ่งสามารถประมวลผลได้ครั้งละหนึ่งตัวอักษรเท่านั้น ซึ่งจากข้อมูลที่ใช้ในงานจริงจะมีลักษณะเป็นหน้ากระดาษ หรือเป็นบรรทัด ดังนั้นจึงต้องทำการแยกตัวอักษรเหล่านั้นออกมาเป็นข้อมูลอักษรเดี่ยวๆเสียก่อน

จากการศึกษาลักษณะการเขียนอักษรภาษาไทยนั้น สามารถสรุปได้ว่า

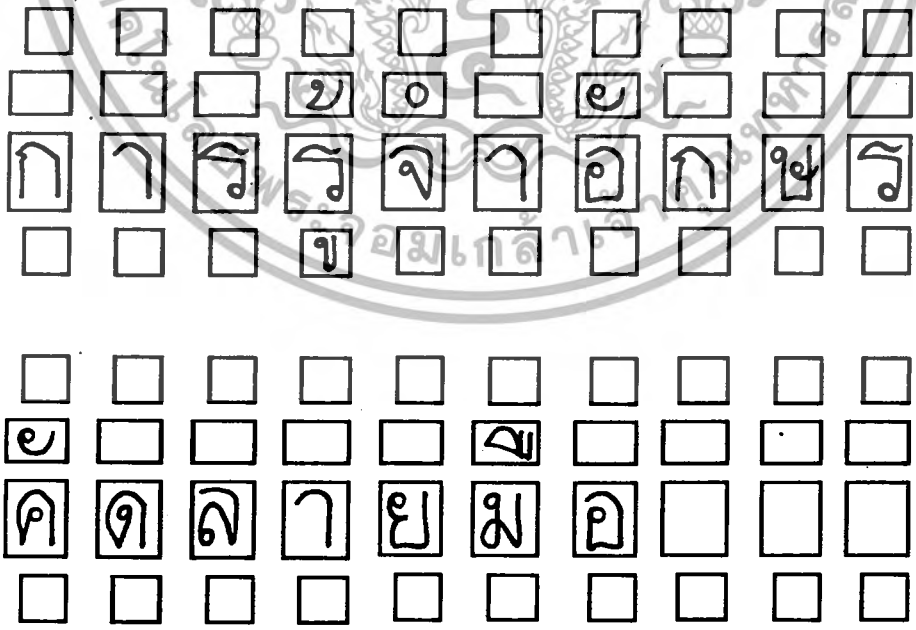
- การเขียนสามารถแบ่งออกได้เป็น 4 ระดับคือ
  1. ระดับที่ 1 เป็นระดับบนสุด ประกอบไปด้วย วรรณยุกต์ และตัวการันต์
  2. ระดับที่ 2 เป็นระดับกลางบน ประกอบด้วย สระระดับบน
  3. ระดับที่ 3 เป็นระดับกลางล่าง ประกอบด้วย พยัญชนะ และสระระดับกลาง
  4. ระดับที่ 4 เป็นระดับล่างสุด ประกอบไปด้วย สระระดับล่าง
- ลักษณะการเขียนที่ทำให้เกิดปัญหาในการรู้จำอักษร
  1. อักษรเขียนติดกัน หรือตัดกัน ทำให้เกิดความสับสนไม่ถูกต้อง
  2. อักษรที่เขียนแยกกัน เช่น แ หรือ ะ เป็นต้น

จากลักษณะดังกล่าวเพื่อให้การรู้จำตัวอักษรของเครื่องไมโครคอมพิวเตอร์ มีประสิทธิภาพมากขึ้น จึงได้กำหนดรูปแบบกระดาษเพื่อเป็นแนวทางในการป้อนข้อมูลจากการเขียนอักษรออกเป็น 2 รูปแบบคือ รูปแบบของกระดาษเพื่อการสร้างพจนานุกรม (ฐานความรู้) ของการรู้จำอักษร ดังแสดงในรูปที่ 3-1 โดยมีลักษณะเป็นช่องสี่เหลี่ยม 1 ระดับวางเรียงกันโดยเว้นระยะเป็นช่องว่างระหว่างสี่เหลี่ยมด้วย ใช้สำหรับเขียนอักษรทีละตัวภายในช่องสี่เหลี่ยมตามลำดับรหัสแอสกี้ของอักษรที่กำหนดไว้ เพื่อใช้เป็นต้นแบบในการเรียนรู้ตัวอักษรแต่ละลักษณะ และรูปแบบของกระดาษเพื่อใช้ทดสอบพจนานุกรม (ฐานความรู้) ของการรู้จำอักษร หรือใช้งานจริง ดังแสดงในรูปที่ 3-2 ซึ่งมีลักษณะเป็นช่องลักษณะเป็นช่องสี่เหลี่ยม 4 ระดับโดยมีช่องว่างระหว่างแถวในแนวตั้งและช่องว่างระหว่างแถวในแนวนอน ใช้สำหรับเขียนอักษรตามรูปแบบของอักษรไทยโดยต้องเขียนอักษรให้อยู่ภายในช่องสี่เหลี่ยมเท่านั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3-1 รูปแบบของกระดาษเพื่อสร้างพจนานุกรม (ฐานความรู้)



รูปที่ 3-2 รูปแบบของกระดาษเพื่อใช้ทดสอบพจนานุกรม (ฐานความรู้)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

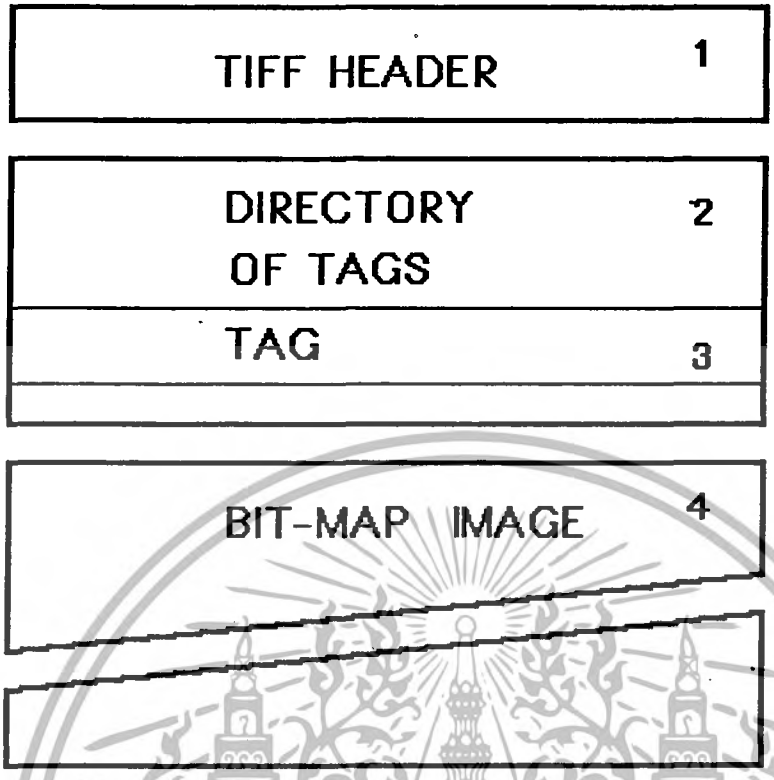
การแยกตัวอักษรออกจากรูปประโยคทีละตัวอักษรนั้น ได้ทำการศึกษาและทำการแยกตัวอักษรได้หลายวิธีการด้วยกัน แต่ด้วยในงานวิจัยนี้ไม่ได้เน้นเทคนิควิธีการในการแยกตัวอักษรเป็นเพียงส่วนในการแยกตัวอักษร เพื่อนำข้อมูลอักษรแต่ละตัวเพื่อเข้าสู่การประมวลผลในส่วนของการรู้จำตัวอักษรต่อไปเท่านั้น จึงใช้กระดาษที่มีรูปแบบเฉพาะเป็นแนวทางในการเขียนอักษรโดยไม่ได้เขียนบนกระดาษที่มีช่องสี่เหลี่ยม แต่จะใช้กระดาษเป็นแนวร่างเพื่อเขียนอักษรดังกล่าว ซึ่งเส้นของช่องสี่เหลี่ยมจะไม่ปรากฏเป็นข้อมูลภาพในการประมวลผล และแยกตัวอักษรโดยการใช้ช่องว่างระหว่างตัวอักษรเป็นหลัก ทั้งในแนวตั้งและในแนวนอน ซึ่งจากวิธีการดังกล่าวแล้วทำให้การแยกแยะตัวอักษรไม่เป็นปัญหาในการประมวลผล และยังใช้เวลาในการประมวลผลน้อยกว่ามากเมื่อเปรียบเทียบกับเวลาที่ใช้ในการประมวลผลทั้งหมด เวลาส่วนใหญ่ของการประมวลผลจึงเป็นเฉพาะส่วนของการประมวลผล เพื่อการรู้จำรูปแบบตัวอักษรจริงๆ โดยมีแนวทางในการแยกตัวอักษรดังต่อไปนี้

3.1 การแปลงข้อมูลภาพจากเครื่องตรวจกวาดภาพ

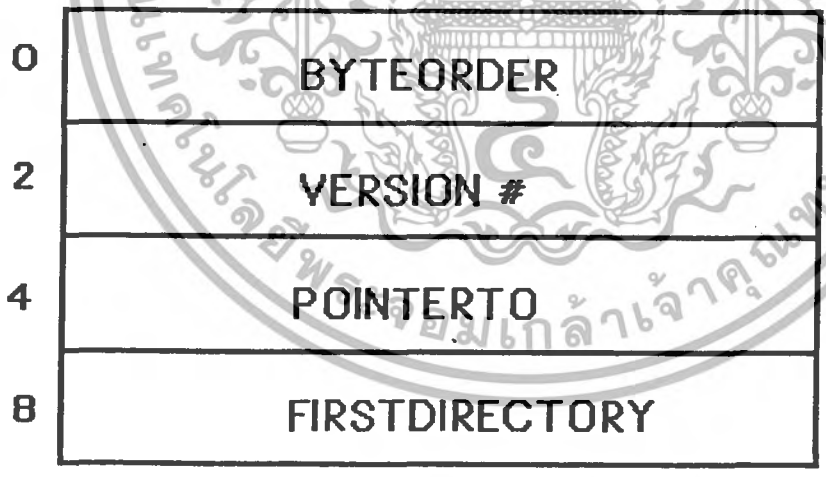
ข้อมูลภาพที่ได้จากเครื่องตรวจกวาดภาพ (Image Scanner) ของ Hewlett Packard (Scanjet Scanner) ซึ่งมีลักษณะข้อมูลภาพที่เก็บในแฟ้มข้อมูลมีลักษณะที่แตกต่างจากเครื่องตรวจกวาดภาพชนิดอื่นๆ โดยมีลักษณะการเก็บข้อมูลในแฟ้มข้อมูลในรูปของ TIFF file ซึ่งสามารถอธิบายโครงสร้างทั่วๆไปของ TIFF file เพื่อให้เห็นภาพแฟ้มข้อมูลโดยง่าย โดยสามารถแบ่งโครงสร้างออกได้เป็น 4 ส่วนด้วยกัน ดังแสดงในรูปที่ 3-3

3.1.1 Details of Header Section เป็นส่วนแรกที่ปรากฏอยู่ใน Tiff file และเป็นโครงสร้างที่สำคัญ ซึ่งจะบอกรายละเอียดของข้อมูลเบื้องต้นในลักษณะของลำดับสองไบท์ โดยแสดงรายละเอียดของรูปแบบการเก็บข้อมูล (BYTE ORDER) ตัวเลขแสดงรุ่นของข้อมูล TIFF Version (VERSION#) และข้อมูลแสดงตำแหน่งแรกของการแสดงรายละเอียดของข้อมูลเข้า (POINTERTO) ดังแสดงในรูปที่ 3-4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3-3 แสดงรายละเอียดโครงสร้างของ TIFF file

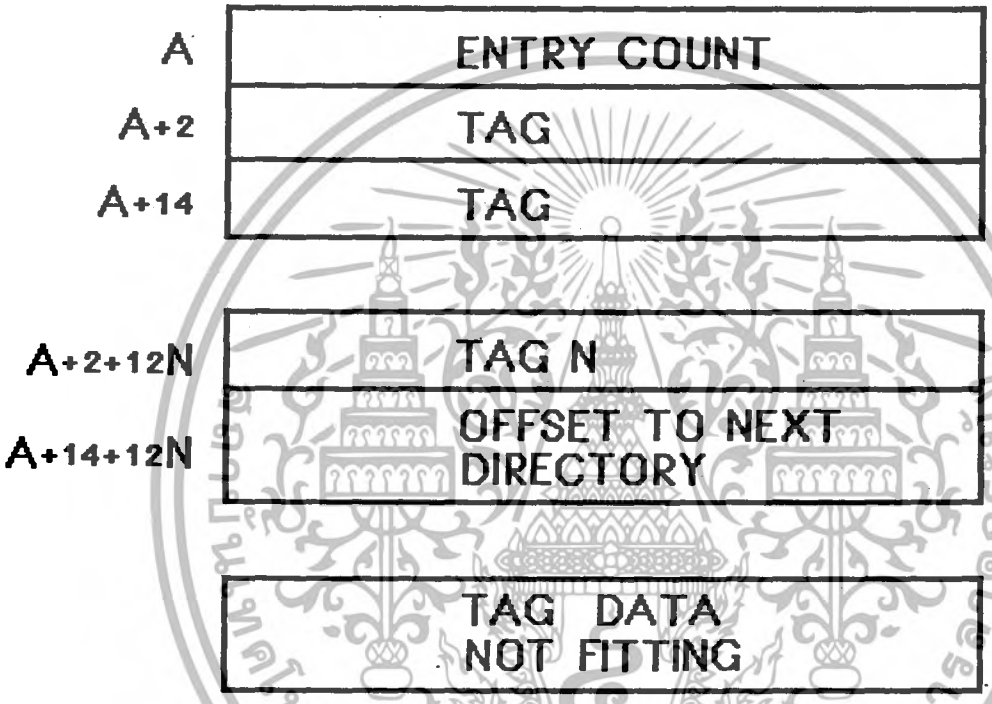


รูปที่ 3-4 แสดงรายละเอียด TIFF Header Section

3.1.2 Details of the Directory เป็นส่วนของข้อมูลที่แสดงหัวข้อรายละเอียดของข้อมูล สามารถแบ่งออกได้เป็น 3 ส่วนด้วยกันคือ ส่วนที่หนึ่งบอกด้วยสองไบท์แรกของส่วน Directory บอกรายละเอียดเกี่ยวกับจำนวนของ Tag (รายละเอียดหลักของข้อมูลภาพ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ส่วนที่สองเป็นส่วนที่แสดงรายละเอียดของข้อมูลในแต่ละ Tag ซึ่งประกอบด้วยข้อมูลแสดงรายละเอียดของแต่ละ Tag จำนวน 12 byte เป็นส่วนของการบอกรายละเอียดข้อมูลของภาพ เช่น ความกว้างของภาพ ความยาวของภาพ เทคนิคการเก็บภาพ ลักษณะการเก็บข้อมูลภาพ เป็นต้น ตามจำนวนของ Tag ซึ่งในแต่ละ Tag จะบอกรายละเอียดของภาพหนึ่งอย่างตามรูปแบบการเก็บของข้อมูลภาพ และส่วนที่สามเป็นส่วนท้ายที่แสดงค่า offset ซึ่งจะชี้ต่อไปยังส่วนของรายละเอียดอื่นๆ ซึ่งถ้าไม่มีส่วนของรายละเอียดถัดไป ก็จะแสดงค่าเป็นศูนย์ ดังแสดงในรูปที่ 3-5



รูปที่ 3-5 แสดงรายละเอียดของ Directory Section

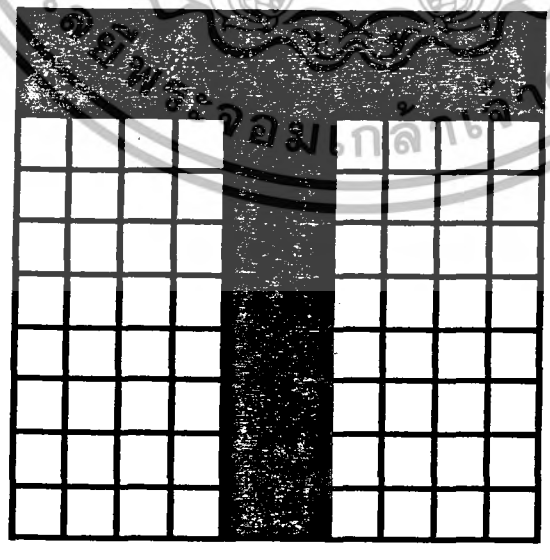
3.1.3 Organization of each Tag Entry องค์ประกอบโครงสร้างของแฟ้มข้อมูลในส่วนของแต่ละ Tag จะประกอบด้วยข้อมูลยาว 12 byte สองไบท์แรกแสดง Tag ตำแหน่งอ้างอิงของข้อมูลหลัก (TAGTYPE) ส่วนสองไบท์ถัดไป(ไบท์ที่ 3 และ 4) แสดงชนิดของข่าวสาร (DATATYPE) เช่นถ้าสองไบท์นั้นมีค่าเป็น 282 (11AH) จะบอกรายละเอียดของความละเอียดในการสแกนตามแกน X ของข้อมูลภาพ ส่วนอีกสี่ไบท์ถัดไป(ไบท์ที่ 5,6,7 และ 8) จะแสดงค่าความยาวของข้อมูล (LENGTH) ส่วนอีกสี่ไบท์ถัดไป(ไบท์ที่ 9,10,11 และ 12) จะแสดงค่าของข้อมูล (VALUE/OFFSET) เช่น ถ้าในส่วนของ DATATYPE มีค่าเป็น 100H และส่วนของ VALUE มีค่าเป็น OAH ก็จะได้ขนาดความกว้างของข้อมูลภาพมีขนาด 10 จุดภาพ เป็นต้น ดังแสดงในรูปที่ 3-6

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

A	TAGTYPE
A+2	DATATYPE
A+4	LENGTH
A+8	VALUEOR
	POINTER

รูปที่ 3-6 แสดงรายละเอียดข้อมูลแต่ละ Tag

3.1.4 Bit-Map Image เป็นส่วนเฉพาะข้อมูลที่มีการจัดเก็บตามลักษณะที่ได้กำหนดไว้ในรูปของข้อมูลแบบ binary ตามในส่วนของ Directory Section ที่ได้กล่าวแล้ว ซึ่งจากโครงสร้างที่ได้กล่าวแล้วนั้น จะทำให้เข้าใจโครงสร้างลักษณะของข้อมูลต่างๆ การจัดเก็บตลอดจนถึงการอ่านข้อมูลออกมาใช้งานอย่างถูกต้อง เพื่อความเข้าใจที่ชัดเจนยิ่งขึ้นจะได้พิจารณาโครงสร้างอย่างหนึ่งซึ่งได้จากเครื่องตรวจกวาดภาพ (Scanjet Hewlett Packard) ในรูปแบบของ 8086 ขนาด 10*10 จุดภาพ ในลักษณะที่เป็น Black and White TIFF Image เพิ่มเติมดังในรูปที่ 3-7 และสามารถแจกแจงการพิจารณาโครงสร้างได้ดังแสดงในตารางที่ 3-1



รูปที่ 3-7 แสดงภาพตัวอย่างในการพิจารณาโครงสร้างของข้อมูลภาพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

HEADER

offset	value	
000	49	Note byte order is intel format.
001	49	
002	2A	Version# 42 ,note Nite byte reversal
003	00	
004	08	Pointer to first directory
005	00	
006	00	
007	00	

DIRECTORY

offset	value	
008	0F	15 tags in this directory
009	00	

	Offset	Tag	Type	Length	Value/Offset	
Tag 1	00A	FF00	0300	0100 0000	0100 0000	(Subfile type)
Tag 2	016	0001	0300	0100 0000	0A00 0000	(Image width 10 pixels)
Tag 3	022	0101	0300	0100 0000	0A00 0000	(Image length 10 pixels)
Tag 4	02E	0201	0300	0100 0000	0100 0000	(1 Bit per sample)
Tag 5	03A	0301	0300	0100 0000	0100 0000	(Pack data compress)
Tag 6	046	0601	0300	0100 0000	0000 0000	(Black is 1)
Tag 7	052	0A01	0300	0100 0000	0100 0000	(Filled MSB to LSB)
Tag 8	05E	1101	0300	0400 0000	D200 0000	(Offset to raster data)
Tag 9	06A	1201	0300	0100 0000	0100 0000	(Normal orientation)
Tag10	076	FF00	1501	0100 0000	0100 0000	(1 Sample/pixel)
Tag11	082	1801	0300	0100 0000	0000 0000	(Min value is 0)
Tag12	08E	1901	0300	0100 0000	0100 0000	(Max value is 1)
Tag13	09A	1A01	0500	0100 0000	C200 0000	(Offset to X resolution)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
Tag14  0A6  1B01  0500  0100 0000  CA00 0000 (Offset to Y resolution)
Tag15  0B2  1C01  0300  0100 0000  0100 0000 (Planar config is normal)
   End of directory
        BE  0000 0000   (no more directories)

  Extended tag information (as needed)
        C2  2C01 0000            0100 0000 (300 dpi X resolution)
        CA  2C01 0000            0100 0000 (300 dpi Y resolution)
 Raster data (as needed)
        D2  FFC0    (10 black pixels)
        D4  FFC0    (10 black pixels)
        D6  0C00    (2 black pixels)
        D8  0C00    (2 black pixels)
        DA  0C00    (2 black pixels)
        DC  0C00    (2 black pixels)
        DE  0C00    (2 black pixels)
        E0  0C00    (2 black pixels)
        E2  0C00    (2 black pixels)
        E4  0C00    (2 black pixels)
```

ตารางที่ 3-1 แสดงการพิจารณาโครงสร้างของแฟ้มข้อมูล

และสามารถเขียนสรุปเป็นตารางแสดงโครงสร้างทั่วไปของแฟ้มข้อมูล ที่แสดงลักษณะโครงสร้างที่เกิดในการสแกนเก็บภาพในลักษณะต่างๆ ดังในตารางที่ 3-2

	Tag #	Tag Name	Value
*	0FFH	Subfile Type	Value : 1
*	100H	Image Width	Value : 10
*	101H	Image length	Value : 10
*	102H	Bits Per Sample	Value : 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

*	103H	Data compression	Value : 1
*	106H	Photometric Interp	Value : 1
	107H	Thresholding	Value :
	108H	Cell Width	Value :
	109H	Cell Length	Value :
*	10AH	Fill Order	Value : 1
	10DH	Document Name	Value :
	10EH	Image Description	Ascii :
	10FH	Scanner Make	Ascii :
	110H	Scanner Model	Ascii :
*	111H	Strip Offset	Value : Offset
*	112H	Orientation	Value : 1
*	115H	Sample Per Pixel	Value : 1
	116H	RowPerStrip	Value :
	117H	StripByteCount	Value : Offset
*	118H	Min Sample Value	Value : 0
*	119H	Max Sample Value	Value : 1
*	11AH	X Resolution	Value : 300
*	11BH	Y Resolution	Value : 300
*	11CH	Planar Configuration	Value : 1
	11DH	Page Name	Ascii :
	11EH	X position	Value :
	11FH	Y position	Value :
	120H	FreeOffset	Value : Offset
	121H	FreeByteCounts	Value : Offset
	122H	GrayResponseUnit	Value :
	123H	GrayResponseCurve	Value : Offset
	124H	Group3Options	Value :
	125H	Group4Options	Value :
	128H	ResolutionUnit	Value :

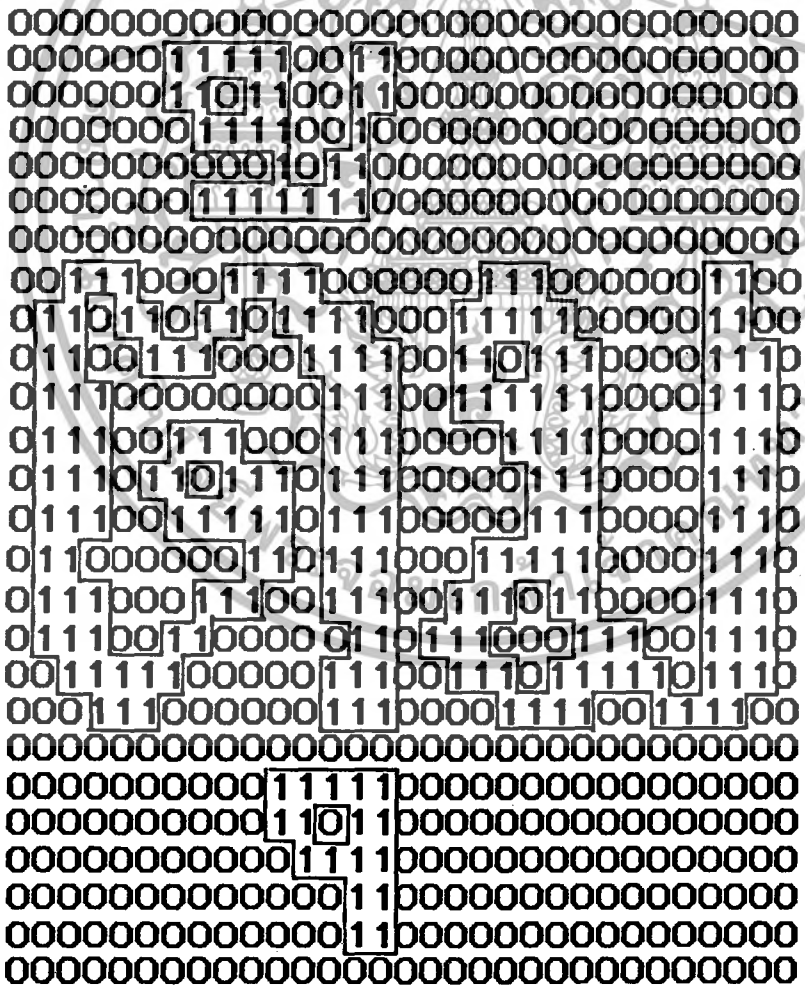
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

129H	PageNumber	Value :
12CH	ColorResponseUnit	Value :
12DH	ColorResponseCurve	Value : Offset

หมายเหตุ * ลักษณะข้อมูลที่จำเป็นในการอธิบายข้อมูล

ตารางที่ 3-2 แสดงโครงสร้างทั่วไปของแฟ้มข้อมูล

เมื่อรับข้อมูลภาพจากเครื่องตรวดกวาดภาพ (Image Scanner) จะทำการแปลงข้อมูลจากแฟ้มข้อมูลที่ได้ โดยพิจารณาตามโครงสร้างแฟ้มข้อมูลที่ได้ทำการพิจารณา และแทนข้อมูลให้อยู่ในรูปของรหัสแทนข้อมูล โดย 1 จะแทนในส่วนที่เป็นเนื้อตัวอักษรและ 0 จะแทนในส่วนที่เป็นพื้นตัวอักษร ดังในรูปที่ 3-8



รูปที่ 3-8 แสดงข้อมูลที่ได้จากการแปลงข้อมูลจากเครื่องตรวจกวาดภาพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

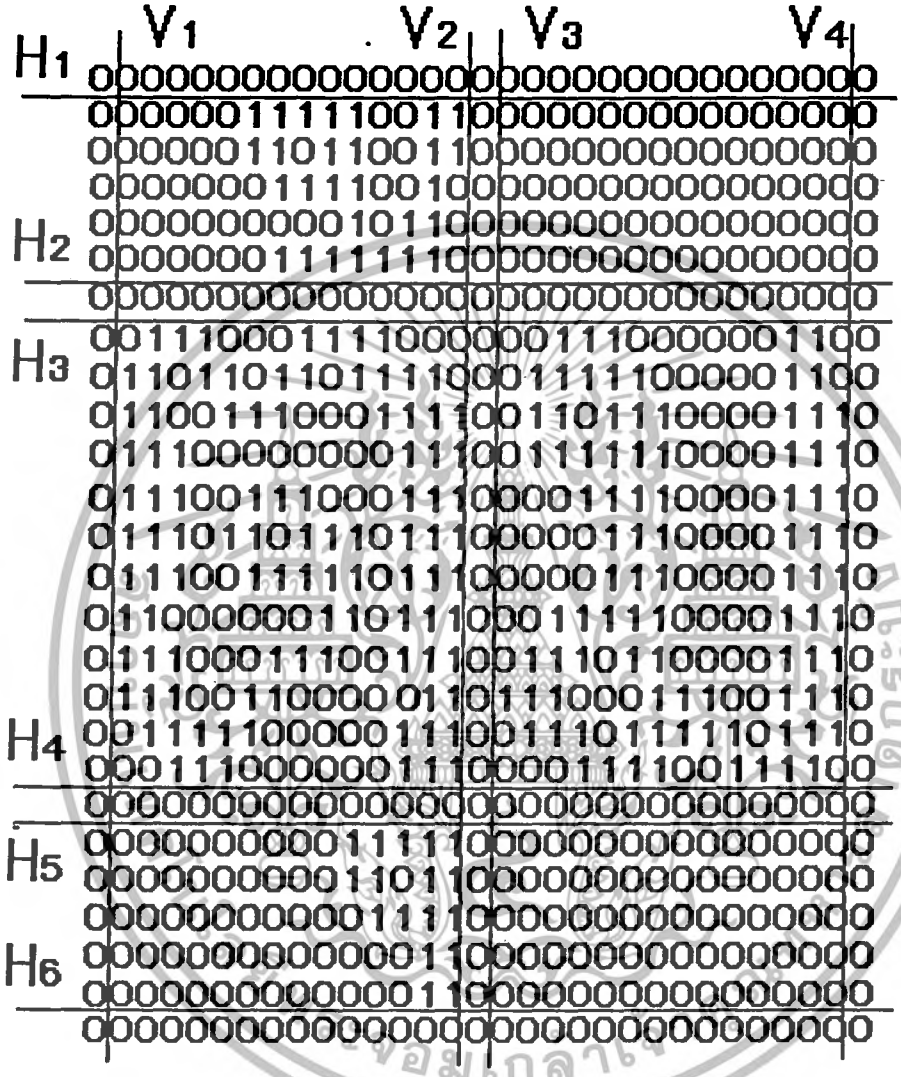
3.2 การแยกแยะภาพอักษรหนึ่งหน้ากระดาษออกเป็นบรรทัด

เมื่อรับข้อมูลภาพที่ได้จากเครื่องตรวจกวาดภาพ (Image Scanner) แล้ว ต้องทำการเปลี่ยนข้อมูลที่อยู่ในรูปของ TIFF file ให้อยู่ในรูปแบบที่ใช้ในการประมวลผล โดยต้องอ่านข้อมูลหลักจากโครงสร้างของแฟ้มข้อมูล TIFF file ที่ได้กล่าวไว้แล้วในหัวข้อ 3.1 แล้วจึงดำเนินแปลงข้อมูลให้อยู่ในรูปรหัส 0 และ 1 เมื่อได้ข้อมูลที่เป็นรหัสแทนทั้งในส่วนที่เนื้อตัวอักษรซึ่งแทนด้วยรหัส 1 และส่วนที่เป็นพื้นอักษรซึ่งแทนด้วยรหัส 0 ทั้งหนึ่งหน้ากระดาษแล้ว จึงทำการสแกนหาช่องว่างระหว่างบรรทัด หรือสแกนหารหัส 0 ที่ต่อเนื่องกันตลอดความกว้างของข้อมูลภาพนั้นเอง เพื่อหาระยะขอบเขตด้านบนและขอบเขตด้านล่างของแต่ละบรรทัด ซึ่งเป็นการแยกแยะตัวอักษรออกเป็นแต่ละบรรทัด

3.3 การแยกแยะภาพอักษรหนึ่งบรรทัดออกเป็นตัวอักษร

เมื่อได้รหัสข้อมูลภาพที่แทนด้วยรหัส 0 และ 1 ในแต่ละบรรทัดแล้ว จึงดำเนินการพิจารณาแยกแยะตัวอักษรต่อไป โดยใช้การพิจารณาหาช่องว่างระหว่างอักษร หรือพิจารณาหารหัสที่เป็น 0 ที่มีการต่อเนื่องตลอดระยะทางระหว่าง ระยะขอบเขตด้านบนและขอบเขตระยะด้านล่างของแต่ละบรรทัด(สแกนตามแนวตั้ง) เพื่อหาขอบเขตระยะด้านหน้าและขอบเขตระยะด้านหลังของแต่ละตัวอักษร(สแกนตามแนวนอน) เป็นขอบเขตของแต่ละตัวอักษรเพื่อใช้ในการประมวลผลต่อไป และดำเนินการหาขอบเขตของอักษรแต่ละตัวจนหมดระยะของบรรทัด ดังแสดงในรูปที่ 3-9 จะได้ระยะ V1,V2,V3 และ V4 เป็นขอบเขตระยะด้านหน้าและขอบเขตระยะด้านหลังของตัวอักษรตามลำดับ และจะได้ระยะ H1,H2,H3,H4,H5 และ H6 เป็นขอบเขตระยะด้านบนและขอบเขตระยะด้านล่างของตัวอักษรตามลำดับ นั่นจะได้ว่าขอบเขตของตัวอักษร "ต" ประกอบไปด้วยระยะ H3,V1,H4 และ V2 ตามลำดับ เป็นต้น และสามารถแยกอักษรออกจากรูปของประโยคข้อความได้ตามลำดับ ดังแสดงในรูปที่ 3-10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



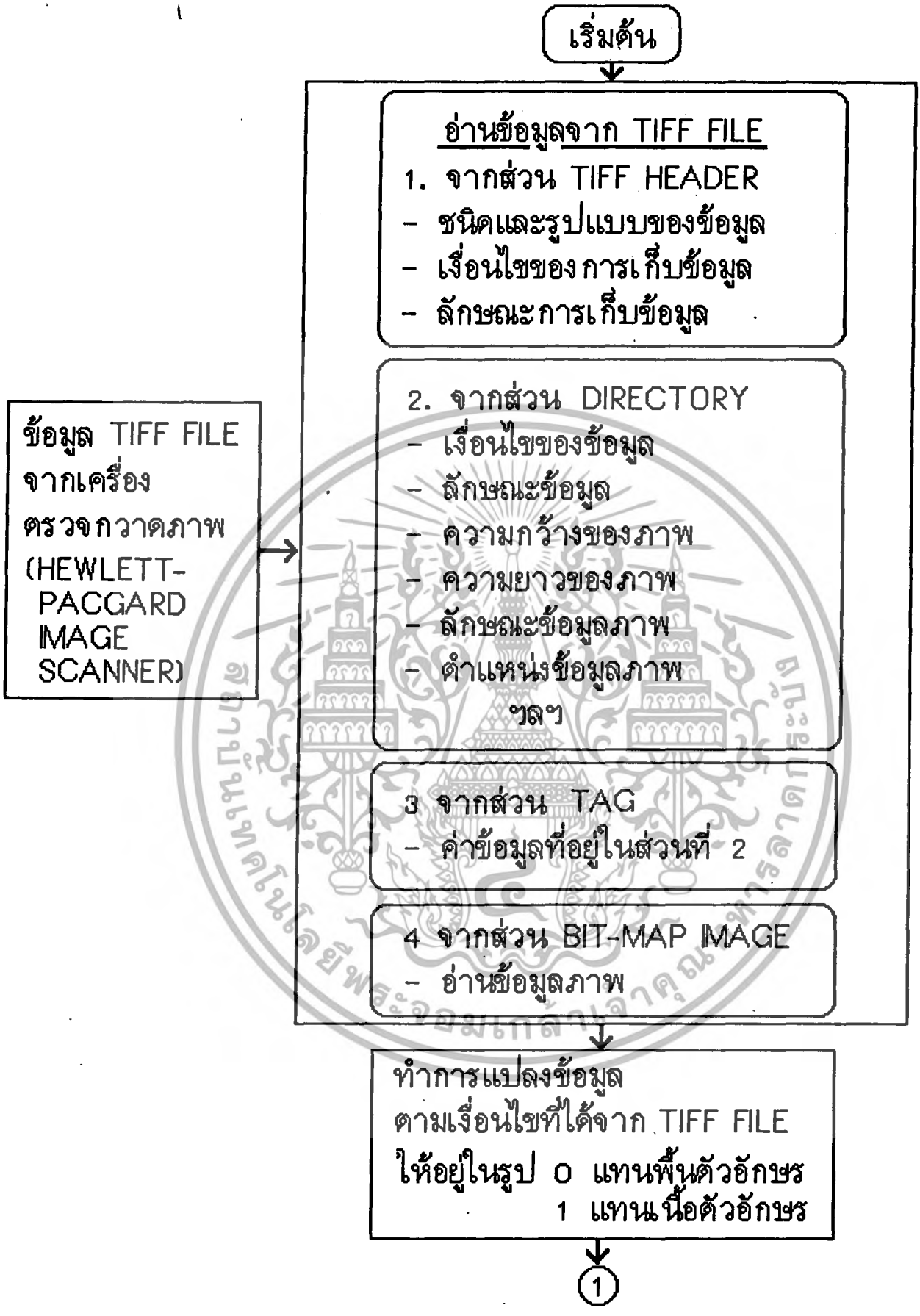
รูปที่ 3-9 แสดงการแยกตัวอักษรออกกันทีละตัวอักษร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 3-10 แสดงผลของการแยกตัวอักษร

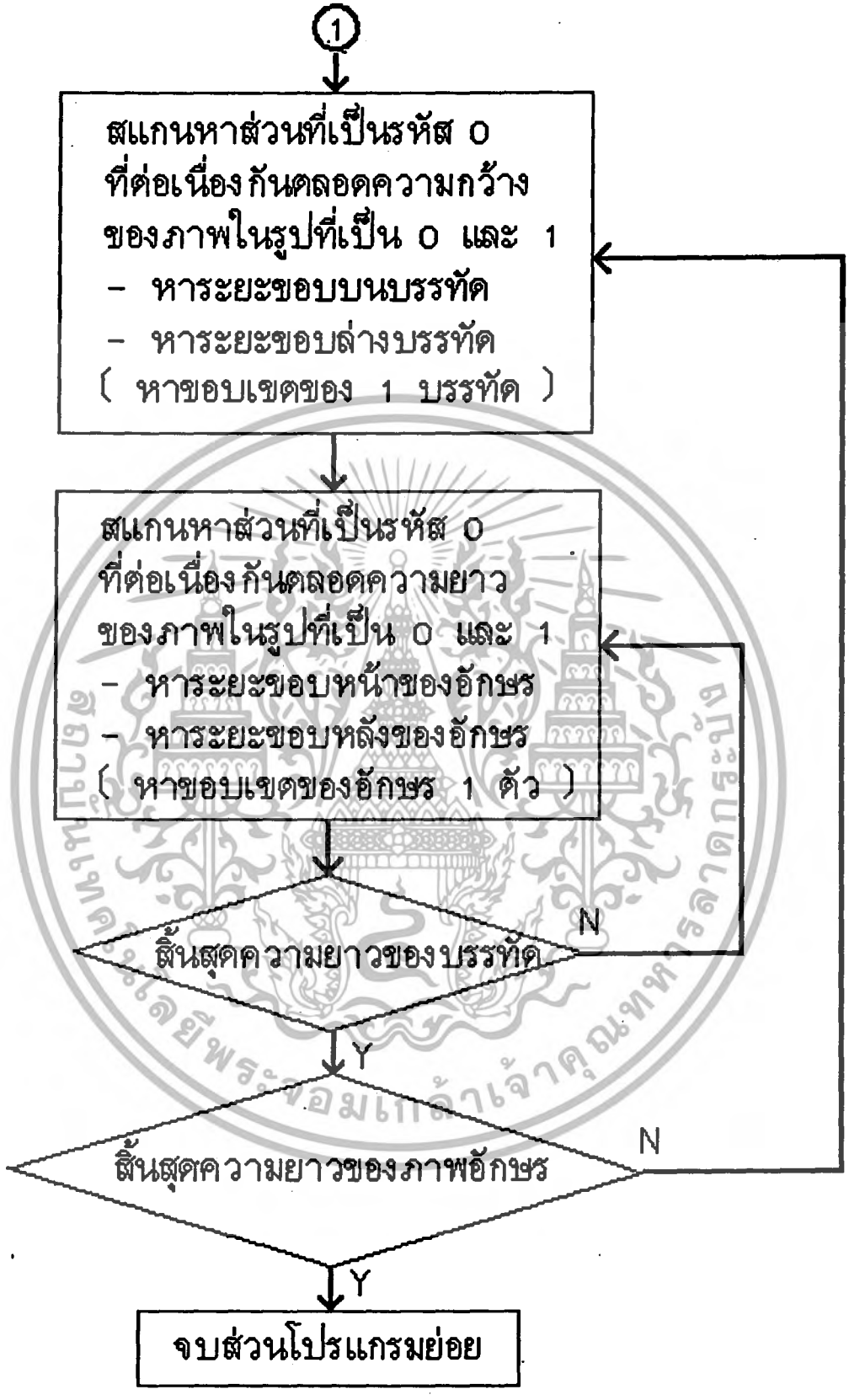
ในส่วนของกระบวนการในการเตรียมข้อมูลเบื้องต้น ตั้งแต่การเปลี่ยนรหัสข้อมูลภาพให้อยู่ในรูปของรหัสแทน 0 และ 1 การแยกแยะตัวอักษรออกจากข้อมูลภาพหนึ่งหน้ากระดาษนั้นสามารถเขียนแสดงในรูปของโฟล์ชาร์ต ดังแสดงในรูปที่ 3-11 ก และ ข

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3-11 (ก) แสดงโฟล์ชาร์ตของการแยกแยะตัวอักษร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



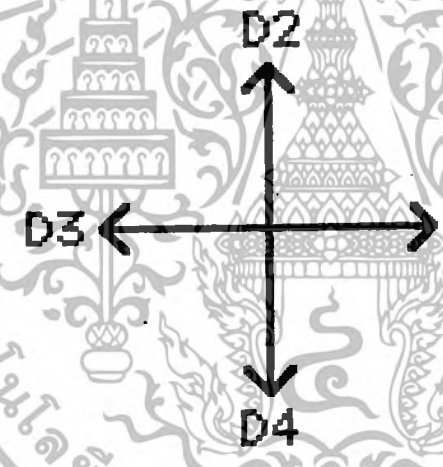
รูปที่ 3-11 (ข) แสดงโฟล์ชาร์ตของการแยกแยะตัวอักษร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ทฤษฎีการรู้จำตัวอักษร

ขั้นตอนในการทำให้เครื่องคอมพิวเตอร์ได้เกิดการรู้จำอักษรลัดลายมือ รู้จักลักษณะรูปแบบของตัวอักษรตลอดจนถึง การจัดทำพจนานุกรม(ฐานความรู้) ของการรู้จำตัวอักษรนั้น อยู่บนพื้นฐานของทฤษฎีของการรู้จำตัวอักษร โดยการวิเคราะห์โครงสร้างของอักษร ด้วยวิธีการ Feature Concentration ซึ่งมีรายละเอียดและสามารถแบ่งออกได้เป็นขั้นตอนใหญ่ๆได้ 3 ขั้นตอนด้วยกันคือ การสร้างรหัสเบื้องต้น (Initial Feature Extraction) การปรับรหัสพื้นตัวอักษร (Unification) การสร้างรหัสรวม (Concentration) โดยในแต่ละขั้นตอนจะต้องทำการพิจารณาทุกๆจุดภาพตามการพิจารณาในแต่ละขั้นตอน โดยดำเนินการพิจารณาตามทิศทางที่กำหนดไว้ ดังแสดงในรูปที่ 4-1

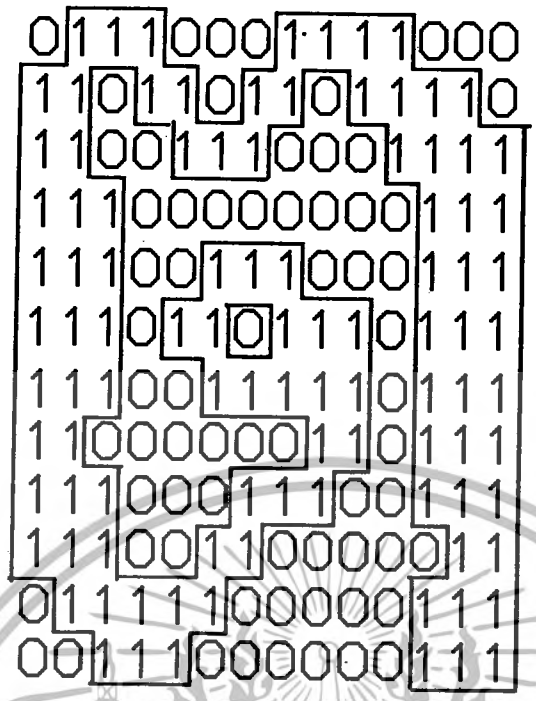


รูปที่ 4-1 แสดงทิศทางในการเข้าหาข้อมูลทั้ง 4 ทิศทาง

4.1 การสร้างรหัสเบื้องต้น (Initial Feature Extraction)

ในส่วนของข้อมูลภาพของตัวอักษรที่ได้ จากแฟ้มข้อมูลของเครื่องตรวจกวาดภาพ (Image Scanner) ซึ่งจะมีลักษณะข้อมูลเป็นไปตามเงื่อนไข และชนิดของการเก็บภาพ ดังนั้นจึงต้องมีกระบวนการในการเตรียมข้อมูลก่อนการประมวลผล (Pre-Processing) เพื่อแปลงข้อมูลจากแฟ้มข้อมูลเดิมให้อยู่ในรูปรหัสที่สามารถนำไปใช้ประมวลผลต่อ ในกระบวนการของการรู้จำได้ต่อไป โดยในส่วนของเนื้อตัวอักษรจะแทนด้วย 1 และในส่วนของพื้นตัวอักษรจะแทนด้วย 0 ดังแสดงในรูปที่ 4-2 แล้วพิจารณาสร้างรหัสแทน โดยแยกออกเป็น 2 ส่วนคือ

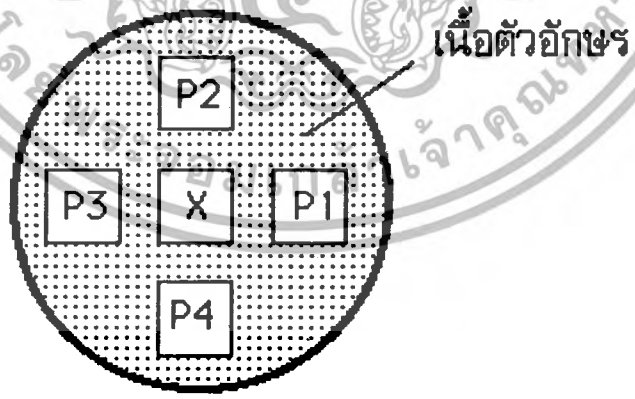
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4-2 แสดงตัวอย่างข้อมูล 1 ตัวอักษร

4.1.1 รหัสเนื้อตัวอักษร (P-Code)

เนื้อตัวอักษรซึ่งถูกแทนด้วย 1 ในรูปที่ 4-2 จะถูกแปลงต่อแทนด้วยรหัส P โดยการพิจารณาเฉพาะจุดที่แทนด้วยรหัสที่เป็น 1 ตามวิธีการ ดังแสดงในรูปที่ 4-3

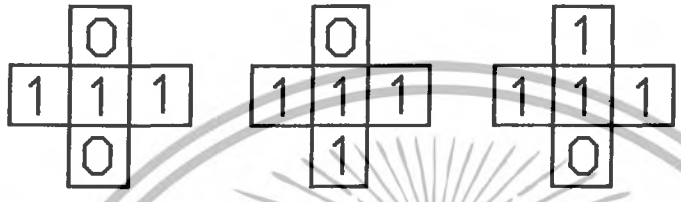


รูปที่ 4-3 Initial Feature Extraction โดยพิจารณาเนื้อตัวอักษร

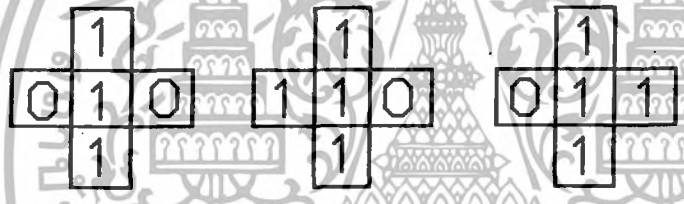
ถ้าให้จุด X เป็นจุดที่แทนด้วยรหัส 1 ซึ่งเรากำลังสนใจที่จะแปลงรหัส และให้จุดอื่นใดที่อยู่ติดกับจุดที่กำลังสนใจเป็น P1,P2,P3,P4 ซึ่งอยู่โดยรอบในทิศทางทั้ง 4 ตามลำดับ โดยเราสนใจกรณี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

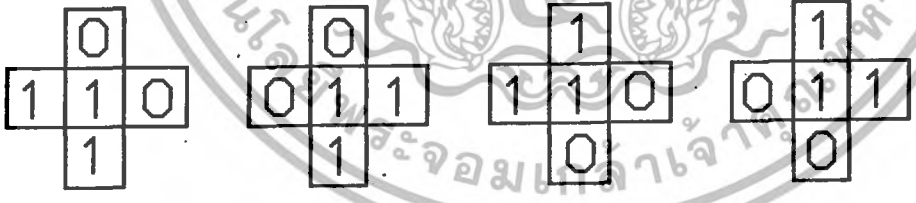
ที่ X ใดๆเป็นรหัส 1 เสมอ จะพบว่ามีโอกาสเป็นไปได้หลายๆกรณีด้วยกัน และสามารถแบ่งออกตามลักษณะการวางตัวของแนวเส้น (แนวทางเดินของรหัส 1) ได้เป็น 4 ลักษณะ ดังแสดงในรูปที่ 4-4 (ก) (ข) (ค) และ (ง) โดยรูป ก จะแสดงการวางตัวในแนวนอน รูป ข แสดงการวางตัวในแนวตั้ง รูป ค แสดงการวางตัวในแนวทแยง และรูป ง แสดงการวางตัวในลักษณะที่ถูกล้อมรอบด้วยรหัสที่เป็น 1 ในทิศทางทั้ง 4



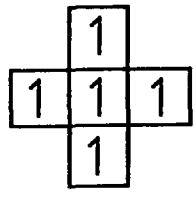
รูปที่ 4-4 (ก) แสดงการวางตัวในแนวนอน



รูปที่ 4-4 (ข) แสดงการวางตัวในแนวตั้ง



รูปที่ 4-4 (ค) แสดงการวางตัวในแนวทแยง



รูปที่ 4-4 (ง) แสดงการวางตัวในลักษณะที่ถูกล้อมรอบด้วย 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

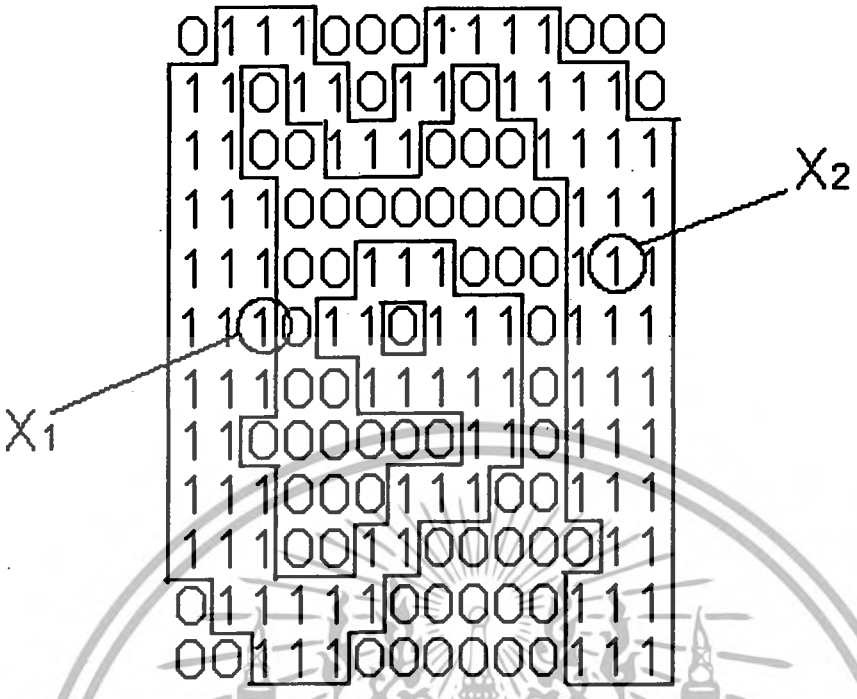
จากรูปที่ 4-4 จะเห็นได้ว่าแนวทางการวางตัวของรหัส 1 นั้นมีด้วยกัน 4 กรณีซึ่งสามารถเขียนอธิบายในรูปของคณิตศาสตร์ได้ทั้ง 4 รูปแบบและได้กำหนดรหัสแทนการวางตัวของรหัส 1 ออกเป็น 4 ลักษณะคือ ถ้าวางตัวในแนวนอนให้แทนด้วย H ถ้าวางตัวในแนวตั้งให้แทนด้วย V ถ้าวางตัวในแนวทแยงใดๆให้แทนด้วย S และถ้าวางตัวในลักษณะที่ถูกล้อมรอบด้วย 1 ให้แทนด้วย I ซึ่งสามารถเขียนแสดงได้ในตารางที่ 4-1

เงื่อนไข	รหัสแทน	ลักษณะเส้น
{(1-P2*P4)*P1*P3} = 1	H	วางตัวในแนวนอน
{(1-P1*P3)*P2*P4} = 1	V	วางตัวในแนวตั้ง
{(1-P1*P3)*(1-P2*P4)} = 1	S	วางตัวแนวทแยง
{P1*P2*P3P4} = 1	I	ถูกล้อมรอบภายใน

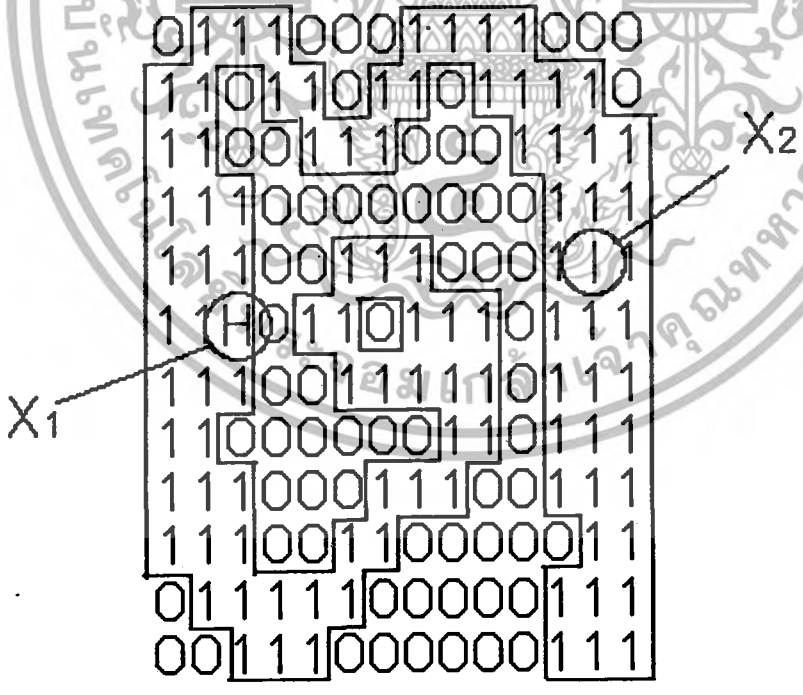
ตารางที่ 4-1 การกำหนดค่ารหัส P

จากรูปที่ 4-2 ถ้าเราแยกการพิจารณาโดยใช้เงื่อนไขตามในตารางที่ 4-1 จะสามารถเขียนแสดงในรูปที่ 4-5 (ก) (ข) โดยหากพิจารณาที่ตำแหน่งของ X1 จะได้ค่าของจุดที่อยู่ติดกับจุด X1 เป็น P1 P2 P3 และ P4 ตามทิศทางในการพิจารณาดังในรูปที่ 4-3 โดยจะมีค่าเป็น 0 1 1 และ 1 ตามลำดับ ซึ่งเมื่อนำมาคำนวณเพื่อหารหัสแทนตามในตารางที่ 4-1 จะสอดคล้องกับสมการในบรรทัดที่ 1 โดยเมื่อแทนค่าแล้วจะทำให้สมการ {(1-P2*P4)*P1*P3}=1 ได้เป็นจริงเพียงสมการเดียว จึงแทนรหัสที่ตำแหน่ง X1 ด้วยรหัส H และในทำนองเดียวกันเมื่อพิจารณาที่ตำแหน่ง X2 จะได้ค่าของจุดที่อยู่ติดกับจุด X2 เป็น P1 P2 P3 และ P4 ตามทิศทางในการพิจารณาดังในรูปที่ 4-3 โดยจะมีค่าเป็น 1 1 1 และ 1 ตามลำดับ ซึ่งเมื่อนำมาคำนวณเพื่อหารหัสแทนตามในตารางที่ 4-1 จะสอดคล้องกับสมการในบรรทัดที่ 4 โดยเมื่อแทนค่าแล้วจะทำให้สมการ {P1*P2*P3P4} 1 เป็นจริงเพียงสมการเดียวเช่นกันจึงแทนรหัสที่ตำแหน่ง X1 ด้วยรหัส I

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหาและต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4-5 (ก) แสดงการพิจารณารหัสเบื้องต้นจากเนื้อตัวอักษร

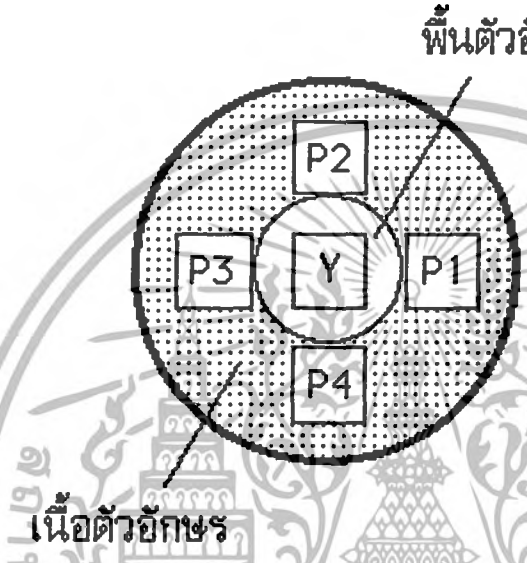


รูปที่ 4-5 (ข) แสดงการแทนรหัสเบื้องต้นจากเนื้อตัวอักษร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.2 รหัสพื้นตัวอักษร (Q-Code)

พื้นตัวอักษรซึ่งถูกแทนด้วย 0 ในรูปที่ 4-2 จะถูกแปลงด้วยอาศัยรหัส Q โดยการพิจารณาเฉพาะจุดที่แทนด้วยรหัสที่ 0 ตามวิธีการดังในรูปที่ 4-6



รูปที่ 4-6 Initial Feature Extraction โดยพิจารณาพื้นตัวอักษร

ถ้าให้จุด Y เป็นจุดที่แทนด้วยรหัส 0 ซึ่งเป็นส่วนของพื้นอักษรที่เรากำลังพิจารณาที่จะแปลงรหัส และให้พิจารณาลักษณะการวางตัวของจุด Y ว่าถูกล้อมรอบด้วยเส้นตรงหรือจุดที่เป็นเนื้อตัวอักษร ซึ่งแทนด้วยรหัส 1 หรือไม่ ในทิศทางทั้ง 4 จากจุดที่กำลังพิจารณา ซึ่งกำหนดให้เป็น P1,P2,P3 และ P4 ตามลำดับ ถ้าพบรหัสแทนที่เป็น 1 ด้านใดก็ให้แทนค่าของ P ในทิศทางนั้นด้วย 1 ถ้าหากไม่พบรหัสแทนที่เป็น 1 ก็ให้แทนค่าของ P ในทิศทางนั้นด้วย 0 โดยพิจารณาทั้ง 4 ทิศทางตามลำดับ จะพบว่ามีโอกาสเป็นไปได้ที่จะมีการวางตัวของรหัสแทนที่เป็น 1 จากทิศทางทั้ง 4 ด้วยกัน 16 กรณี ซึ่งสามารถเขียนแสดงในตารางที่ 4-2 และแทนลักษณะการวางตัวในลักษณะต่างๆด้วยรหัสแทน Q0,Q1,...,QF ตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รหัส	ลักษณะแทน	รหัส	ลักษณะแทน
Q0		Q8	
Q1		Q9	
Q2		QA	
Q3		QB	
Q4		QC	
Q5		QD	
Q6		QE	
Q7		QF	

ตารางที่ 4-2 การกำหนดค่ารหัส Q

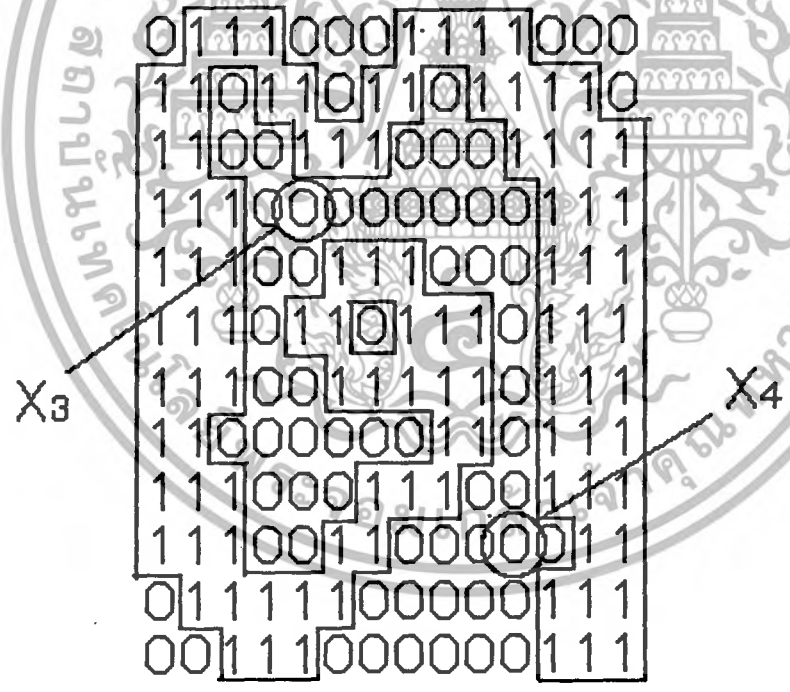
จากตารางที่ 4-2 จะเห็นได้ว่าโอกาสในการเกิดตามเงื่อนไขการพิจารณาส่วนที่เป็นพื้นตัวอักษรจะเป็นไปได้ด้วยกัน 16 กรณี ซึ่งสามารถอธิบายการเกิดลักษณะต่างๆได้ ด้วยสมการทางคณิตศาสตร์ โดยการกำหนดค่าน้ำหนักถ่วง (Weight) ให้แต่ละแนวทางของการพิจารณาโดยให้แนวทางด้านล่างมีน้ำหนักเป็น 1 แนวทางด้านซ้ายมีน้ำหนักเป็น 2 แนวทางด้านบนมีน้ำหนักเป็นเป็น 4 และแนวทางด้านขวามีน้ำหนักเป็นเป็น 8 ของจุดพื้นตัวอักษรที่มีรหัสแทนเป็น 0 ซึ่งเป็นจุดที่กำลังพิจารณา แล้วนำค่าที่ได้จากการพิจารณาการวางตัวของจุดพื้นตัวอักษรล้อมรอบด้วยรหัส 1 (P1,P2,P3,P4) หรือไม่ ทั้ง 4 ทิศทางไปแทนค่าในสมการที่ 4-1 เพื่อหาค่า K ซึ่งจะเป็นรหัสที่ใช้แทนในการพิจารณาสร้างรหัสเบื้องต้นด้วยการพิจารณาพื้นตัวอักษร จะได้ค่าที่เป็นไปได้ทั้งหมด 16 กรณี ซึ่งสามารถเขียนแสดงด้วยเลขฐานสิบหาตั้งแต่ 0,1,2,3,...,F แต่จากการทดลองเมื่อข้อมูลผ่านกระบวนการแยกแยะตัวอักษรแล้ว จะทำให้รหัสพื้นตัวอักษรที่ได้มีเพียง 9 กรณี ยกเว้นกรณีของ 0,1,2,4 และ 8 ด้วยข้อมูลที่ผ่านขั้นตอนของการแยกแยะอักษรแล้วนั้น จะต้องแสดงส่วนที่เป็นเนื้อตัวอักษรติดด้านทั้ง 4 ของขอบเขตอักษรภาพแต่ละตัว ดังแสดงในรูปที่ 4-2 รหัสพื้นตัวอักษรจึงไม่เกิดกรณีที่ถูกล้อมรอบด้วยเส้นตรงหรือจุดที่มีรหัสเป็น 1 เพียงด้านเดียว หรือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ไม่ถูกล้อมรอบด้วยจุดที่มีรหัสเป็น 1 เลย ซึ่งสามารถแสดงการหารหัสพื้นตัวอักษรได้ดังแสดงในรูปที่ 4-7 (ก) และ (ข)

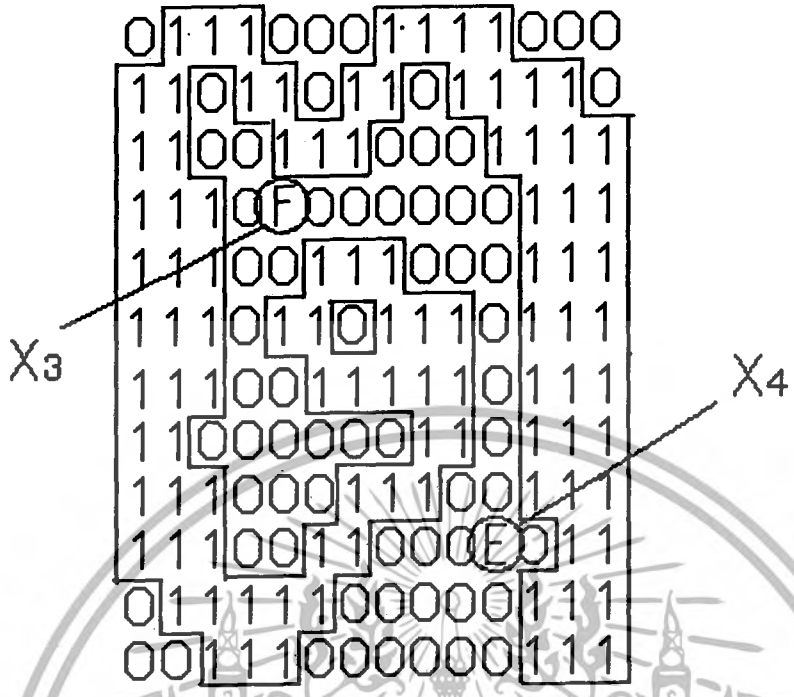
$$K = \{ (8*P1)+(4*P2)+(2*P3)+(1*P4) \} \quad ... \quad 4\text{-}1$$

จากรูปหากพิจารณาที่จุด X3 และพิจารณาตามแนวทางทั้ง 4 เพื่อหาการถูกล้อมรอบด้วยเส้นที่มีรหัสเป็น 1 จะได้ค่าของ P1,P2,P3,P4 มีค่าเป็น 1,1,1,1 ตามลำดับ และนำมาแทนค่าในสมการที่ 4-1 จะได้ค่า K = 15 ซึ่งแทนด้วยรหัส F จะได้รหัสแทนในตำแหน่ง X3 นั้นด้วย F และในทำนองเดียวหากพิจารณาที่จุด X4 จะได้ค่าของ P1,P2,P3,P4 มีค่าเป็น 1,1,1,0 ตามลำดับ และนำมาแทนค่าในสมการที่ 4-1 จะได้ค่า K = 14 ซึ่งแทนด้วยรหัส E จะได้รหัสแทนในตำแหน่ง X4 นั้นด้วย E



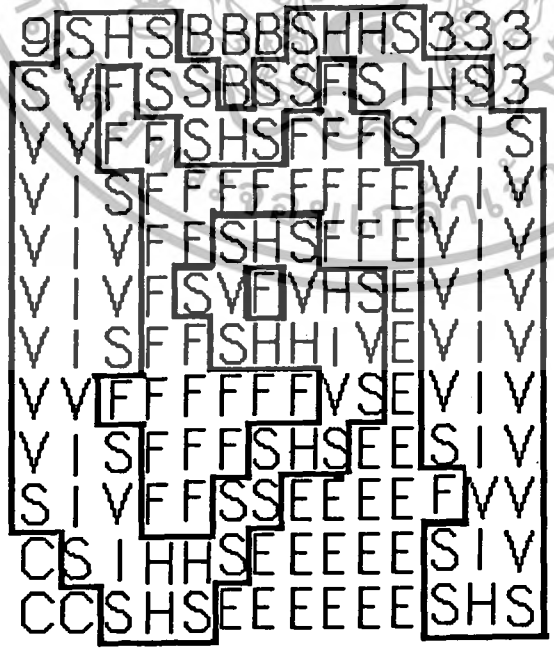
รูปที่ 4-7 (ก) แสดงการพิจารณารหัสเบื้องต้นจากพื้นตัวอักษร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4-7 (ข) แสดงการแทนรหัสเบื้องต้นจากพื้นตัวอักษร

จากหัวข้อที่ 4.1.1 และ 4.1.2 เมื่อทำการพิจารณาทุกจุดภาพทั้งในส่วนที่เป็นพื้นตัวอักษรและส่วนที่เป็นเนื้อตัวอักษร หรือเมื่อแทนค่าของรหัส P และรหัส Q ของทุกจุดภาพแล้วจะจะสามารถแสดงดังในรูปที่ 4-8



รูปที่ 4-8 แสดงข้อมูลที่ผ่านขั้นตอนของการกำหนดรหัสเบื้องต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 การปรับรหัสพื้นตัวอักษร (Unification)

จากขั้นตอนของการสร้างรหัสเบื้องต้น เป็นการพิจารณาแทนรหัสส่วนที่เป็นเนื้อตัวอักษรและส่วนที่เป็นพื้นตัวอักษร แต่ปัญหาที่เกิดจากการทดลองนั้น ลักษณะของข้อมูลอักษรที่ได้จากกระบวนการเตรียมข้อมูล(การเขียนและเครื่องตรวจกวาดภาพ) อยู่ในลักษณะที่ไม่สมบูรณ์มีลักษณะเว้า แว่ง ไม่แสดงลักษณะโดยรวมของตัวอักษรบริเวณนั้นๆ ดังแสดงในรูปที่ 4-9 จากรูปเมื่อพิจารณาจุด X5 จะเป็นได้ว่ารหัสที่ได้คือ F แต่ลักษณะที่แสดงไม่ได้มีลักษณะที่ถูกล้อมปิดด้วยจุดที่มีรหัสเป็น 1 (เส้นวงปิด) ในตำแหน่ง X5 จึงไม่แสดงลักษณะที่ถูกต้อง ซึ่งควรจะเป็นในลักษณะที่ถูกปิดด้วยเส้น หรือจุดที่มีรหัสเป็น 1 เพียงสามด้านบนและเปิดด้านล่างในลักษณะแบบรหัส E หากพิจารณาที่จุด X6,X7 ซึ่งทั้งสองจุดนั้นแสดงลักษณะที่ไม่ถูกต้อง สาเหตุอาจเนื่องมาจากการผิดพลาดจากกระบวนการเตรียมข้อมูล เช่น ลักษณะการเขียน กระดาษ การเก็บภาพ เป็นต้น ทำให้เกิดลักษณะที่เว้า แว่ง แสดงรหัสของข้อมูลไม่สอดคล้องกับลักษณะอักษรที่ถูกต้อง

รูปที่ 4-9 แสดงข้อมูลที่มีลักษณะไม่สมบูรณ์

จากรูปลักษณะดังกล่าว จึงต้องมีการเพิ่มกระบวนการในการปรับส่วนที่เป็นพื้นตัวอักษร(ส่วนที่เป็นรหัส Q) เพื่อแสดงลักษณะของโครงสร้างโดยรวมของบริเวณเดียวกัน และปรับรหัสพื้นตัวอักษรที่เกิดจากข้อมูลที่ไม่สมบูรณ์ (ผิดพลาดจากกระบวนการเตรียมข้อมูล) ซึ่งอาศัย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คุณสมบัติทางเรขาคณิตด้วยการนำค่ารหัส Q ที่อยู่ติดกันมาพิจารณาปรับรหัสตามกฎเกณฑ์ของการปรับรหัส Q ในตารางที่ 4-3 โดยการพิจารณาจุดที่สนใจให้อยู่ในตารางแถวที่ 1 และให้จุดรอบข้างทั้ง 4 ทิศทางมีลักษณะในตารางแถวที่ 2 ซึ่งถ้าตรงกับลักษณะใดลักษณะหนึ่งก็ให้เปลี่ยนรหัสในตารางแถวที่ 1 เป็นรหัสในตารางแถวที่ 3 แทน ซึ่งเป็นผลจากการทดลองจากอักษรญี่ปุ่นและตัวเลขอารบิค สามารถดัดแปลงมาใช้กับเงื่อนไขของอักษรไทย เพื่อลดรหัสพื้นในตารางแถวที่ 4 โดยออกมาได้ดังแสดงในตารางที่ 4-3

รหัสแถวที่ 1	รหัสแถวที่ 2	รหัสแถวที่ 3	อักษรที่สอดคล้อง
■	■ ■ ■ ■ ■ ■ ■	■ ■ ■ ■ ■ ■ ■	ก ภ ด ต ศ จ ฉ ธ ฏ ฐ ช ซ ๒ ฐ ฯ ๔ ๕ ๘ แ ั
■	■ หรือ ■	■	แ
■	■ หรือ ■	■	ั
■	■ หรือ ■	■	แ
■	■ หรือ ■	■	ั

ตารางที่ 4-3 กฎการปรับรหัสพื้นตัวอักษร

จากหัวข้อที่ 4-1 การสร้างรหัสเบื้องต้นจะได้ข้อมูลในรูปที่ 4-8 เมื่อทำการปรับรหัสพื้นอักษร ตามเงื่อนไขจะได้ผลดังแสดงในรูปที่ 4-10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 4-10 แสดงข้อมูลที่ผ่านขั้นตอน การปรับรหัสพื้นตัวอักษร

4.3 การสร้างรหัสรวม (Concentration Code)

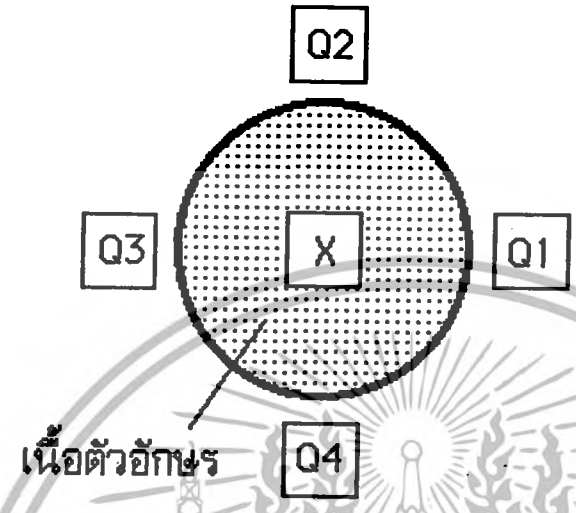
เมื่อผ่านกระบวนในการพิจารณารหัสแทนส่วนที่เป็นเนื้อตัวอักษร และส่วนที่เป็นพื้นตัวอักษร และการปรับรหัสส่วนที่เป็นพื้นตัวอักษร กระบวนการต่อไปเป็นกระบวนการในการนำเอารหัสเบื้องต้นที่ได้ไปอธิบายและพิจารณาเก็บเป็นลักษณะเด่นเพื่อใช้แทนอักษรตัวนั้นๆ ในการนำเอารหัสเบื้องต้นไปอธิบาย อาจทำได้โดยการใช้ส่วนที่เป็นเฉพาะรหัสเนื้อตัวอักษร หรือเฉพาะส่วนที่เป็นรหัสพื้นตัวอักษรส่วนใดส่วนหนึ่งไปอธิบาย ซึ่งทำให้ลักษณะข้อมูลที่ได้ไม่ชัดเจนในการอธิบายถึงลักษณะอักษรบริเวณที่ใช้รหัสดังกล่าวไปแทน ดังนั้นในส่วนของนำเอารหัสเบื้องต้นไปอธิบายจึงได้นำเสนอใช้วิธีการสร้างรหัสรวมใหม่ ซึ่งเกิดจากการพิจารณารหัสร่วมกันทั้งในส่วนที่เป็นเนื้อตัวอักษรและส่วนที่เป็นพื้นตัวอักษร โดยการนำรหัสในจุดที่พิจารณาและจุดที่อยู่รอบๆข้างทั้ง 4 ทิศทางมารวมกันเป็นรหัสรวม ซึ่งสามารถอธิบายลักษณะของอักษรบริเวณนั้นๆได้ชัดเจนกว่า โดยสามารถแบ่งการพิจารณาสร้างรหัสรวมได้เป็น 2 ส่วนคือ

4.3.1 รหัสรวมที่ใช้เนื้อตัวอักษรเป็นหลัก

โดยการพิจารณาใช้ส่วนที่เป็นเนื้อตัวอักษรเป็นหลัก ซึ่งเป็นลักษณะของรหัส P แล้วพิจารณาหารหัส Q ในทิศทางที่อยู่โดยรอบทั้ง 4 ทิศทางตามลำดับ ดังในรูปที่ 4-11 จากรูปให้ X เป็นส่วนที่เป็นเนื้อตัวอักษรและ Q1,Q2,Q3,Q4 เป็นรหัสพื้นตัวอักษรตำแหน่งแรกของการพิจารณา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

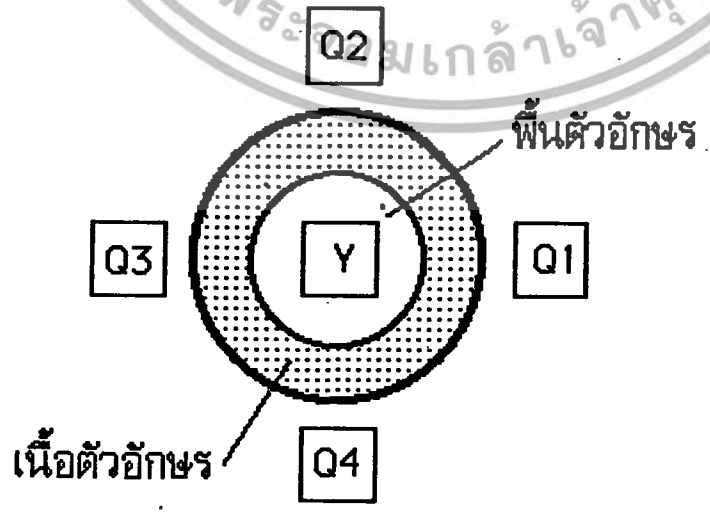
ในแนวทางโดยรอบทั้ง 4 ทิศทาง ซึ่งถ้าการพิจารณาในแนวทางใดๆ ไม่พบรหัสแทนในส่วนที่เป็นพื้นตัวอักษร ก็จะให้ค่าในส่วนที่แทน Q ใดๆนั้นด้วยศูนย์



รูปที่ 4-11 Concentration โดยการพิจารณาเนื้อตัวอักษร

4.3.2 รหัสรวมที่ใช้พื้นตัวอักษรเป็นหลัก

โดยการพิจารณาใช้ส่วนที่เป็นพื้นตัวอักษรเป็นหลัก ซึ่งเป็นลักษณะของรหัส Q แล้วพิจารณาหารหัส Q ที่อยู่ถัดไปจากรหัส P ที่อยู่ถัดจากจุดที่พิจารณาออกไปทั้ง 4 ทิศทางตามลำดับ ดังรูปที่ 4-12 จากรูปกำหนดให้ Y เป็นรหัส Q ส่วนพื้นตัวอักษรที่กำลังพิจารณา และให้ Q1,Q2, Q3,Q4 เป็นส่วนของพื้นตัวอักษรในตำแหน่งแรก ที่พิจารณาโดยผ่านส่วนที่เป็นเนื้อตัวอักษรออกไป หากการพิจารณาไม่พบส่วนที่เป็นพื้นตัวอักษรตามเงื่อนไขที่กำหนดก็จะให้ค่าของ Q ใดๆนั้นด้วยศูนย์

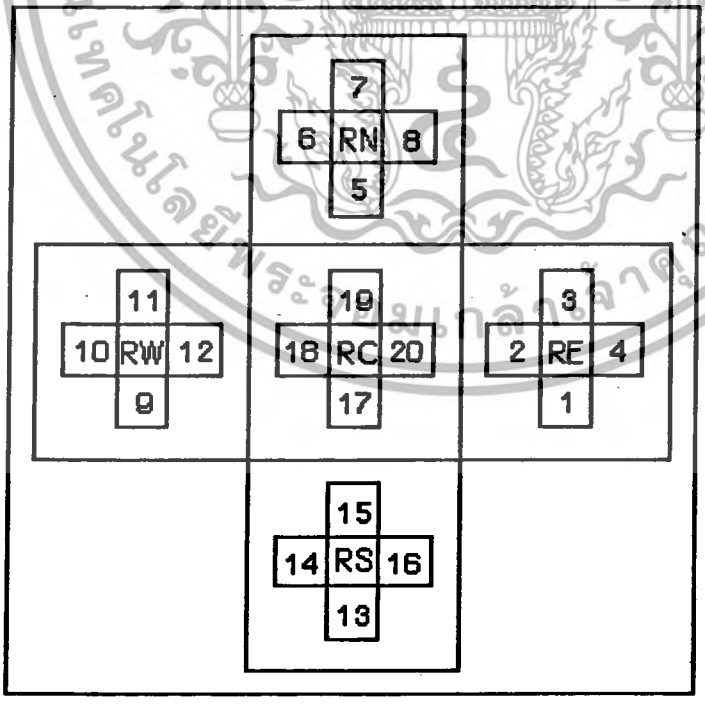


รูปที่ 4-12 Concentration โดยการพิจารณาพื้นตัวอักษร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3.3 การจัดโครงสร้างของรหัสรวม

จากการพิจารณาสร้างรหัสรวมในหัวข้อที่ 4.3 จะได้รหัสรวมที่แสดงโครงสร้างเป็น RE,RN,RW,RS,RC โดยค่าของ RE เป็นค่าของ Q1 ซึ่งมีขนาด 4 bits, RN เป็นค่าของ Q2 ซึ่งมีขนาด 4 bits,RW เป็นค่าของ Q3 ซึ่งมีขนาด 4 bits,RS เป็นค่าของ Q4 ซึ่งมีขนาด 4 bits และ RC เป็นค่าของ X เมื่อพิจารณาในลักษณะรูปที่ 4-11 และ RC เป็นค่าของ Y เมื่อพิจารณาในลักษณะรูปที่ 4-12 ซึ่งเป็นส่วนรหัสพื้นตัวอักษร 16 ตัวและส่วนรหัสเนื้อตัวอักษร 4 ตัว รวมเป็น 20 ตัว ซึ่งต้องใช้ข้อมูลแสดงความแตกต่างถึง 20 แบบ จึงเกิดความไม่สะดวกในการอธิบายด้วยเลขฐาน 2 ซึ่งต้องใช้ถึง 5 bits แต่จากการพิจารณาส่วนของการสร้างรหัสแทนพบว่าส่วนของรหัสพื้นตัวอักษรจะไม่เกิดขึ้น 5 ตัวอันเนื่องมาจากกระบวนการแยกแยะตัวอักษร ซึ่งได้แก่รหัส Q0,Q1,Q2,Q4,Q8 จึงได้พิจารณาแทนส่วนของรหัสเนื้อตัวอักษร 4 ตัวด้วยรหัสพื้นตัวอักษรที่ไม่เกิด โดยกำหนดให้ H,V,S,I แทนด้วยค่า Q8,Q4,Q2,Q1 ตามลำดับ ก็จะสามารถบอกความแตกต่างของข้อมูลได้เพียง 16 แบบ ซึ่งสามารถแสดงด้วยเลขฐาน 2 เพียง 4 bits จึงได้โครงสร้างของรหัสรวมจากทิศทางทั้ง 5 รวมเป็น 20 bits ดังแสดงในรูปที่ 4-13 (ก) และ (ข)



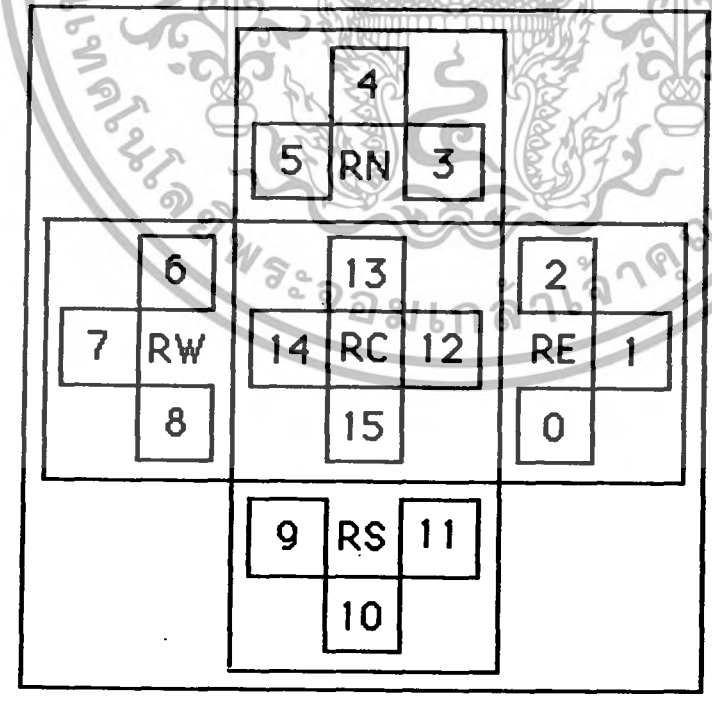
รูปที่ 4-13 (ก) แสดงการพิจารณาโครงสร้างของรหัสรวม 20 บิต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1 2 3 4	5 6 7 8	9 10 11 12	13 14 15 16	17 18 19 20
E N W S	E N W S	E N W S	E N W S	E N W S
RE	RN	RW	RS	RC

รูปที่ 4-13 (ข) แสดงการแจกแจงโครงสร้างของรหัสรวม 20 บิต

แต่จากการพิจารณาสร้างรหัสรวมดังกล่าวจะเห็นได้ว่า ในขั้นตอนของการพิจารณารหัสรวมนั้น ส่วนของการพิจารณาหารหัส Q1,Q2,Q3,Q4 นั้นจะต้องพิจารณาผ่านส่วนที่เป็นเนื้อตัวอักษรก่อนเสมอ หรือหมายความว่าบิตข้อมูลของ Q1,Q2,Q3,Q4 ด้านในที่อยู่ใกล้กับจุดที่กำลังพิจารณาจะมีค่าเป็น 1 เสมอเมื่อพบรหัส Q ใดๆซึ่งไม่เป็นศูนย์ ดังนั้นจึงสามารถ ที่จะลดขนาดของ bit ข้อมูลรหัสรวมลงได้อีก 4 bits เพื่อเป็นการประหยัดพื้นที่ในการเก็บข้อมูลโดยเก็บเป็นรหัสรวมเพียง 16 bits หรือใช้เพียง 2 bytes สามารถลดลงถึง 1 byte ซึ่งเป็นการลดขนาดข้อมูลและให้ความหมายของข้อมูลเหมือนเดิม จึงมีโครงสร้างของข้อมูลรหัสรวมดังแสดงในรูปที่ 4-14 (ก) และ (ข)



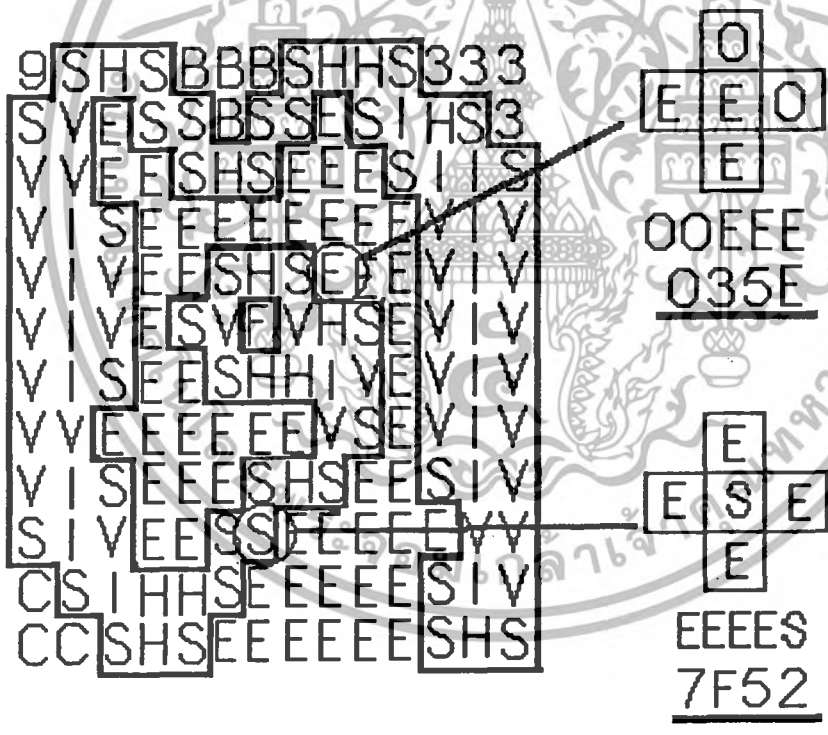
รูปที่ 4-14 (ก) แสดงการพิจารณาลดโครงสร้างของรหัสรวม 16 บิต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

0 1 2	3 4 5	6 7 8	9 10 11	12 13 14 15
S E N	E N W	N W S	W S E	E N W S
RE	RN	RW	RS	RC

รูปที่ 4-14 (ข) แสดงการแจกแจงโครงสร้างของรหัสรวม 16 บิต

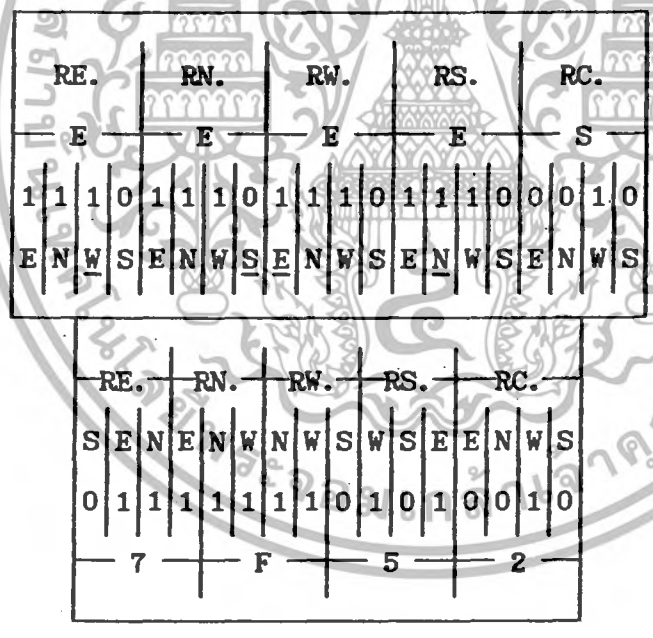
จากข้อมูลที่ได้จากการปรับรหัสพื้นตัวอักษร ในรูปที่ 4-10 สามารถแสดงการหารหัสรวมได้ดังแสดงในรูปที่ 4-15



รูปที่ 4-15 แสดงการหารหัสรวม (Concentrated Word)

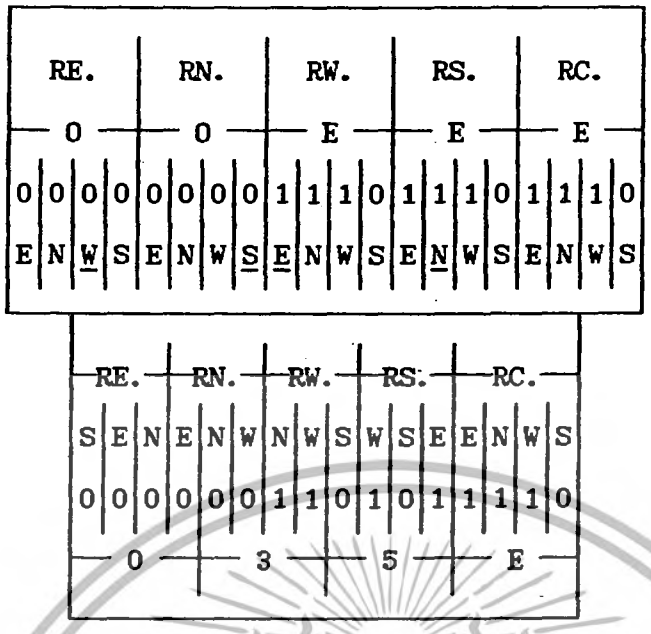
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ถ้าหากพิจารณาส่วนที่เป็นเนื้อตัวอักษร ในรูปที่ 4-10 บริเวณตำแหน่ง X8 จะได้ค่า Q ใดๆโดยจะได้ค่า Q1,Q2,Q3,Q4 และ X เป็น E,E,E,E และ S ตามลำดับ หรือเขียนแสดงตามโครงสร้างของรหัสรวมตามในรูปที่ 4-14 (ก) และ (ข) จะได้รหัสรวมในรูปของรหัสเนื้อตัวอักษร และรหัสพื้นตัวอักษรเป็น EEEES (1110,1110,1110,1110,0010) โดยจัดเรียงตามลำดับการพิจารณา แล้วจัดรหัสรวมใหม่ตามการพิจารณารหัสรวมเป็น 7F52(011,111,110,101,0010) แล้วดังแสดงในรูปที่ 4-15 และรูปที่ 4-16 (ก) และในทำนองเดียวกันหากพิจารณาในส่วนที่เป็นพื้นตัวอักษรดังในรูปที่ 4-10 บริเวณตำแหน่ง X9 จะได้ค่า Q ใดๆ โดยจะได้ค่า Q1,Q2,Q3,Q4 และ Y เป็น 0,0,E,E และ E ตามลำดับ หรือเขียนแสดงตามโครงสร้างของรหัสรวมตามในรูปที่ 4-14 (ก) และ (ข) จะได้รหัสรวมในรูปของรหัสเนื้อตัวอักษรและรหัสพื้นตัวอักษรเป็น 00EEE (0000,0000,1110,1110,1110) โดยจัดเรียงตามลำดับการพิจารณาและจะได้รหัสรวมใหม่ตามการพิจารณารหัสรวมเป็น 035E (000,000,110,101,1110) แล้วดังแสดงในรูปที่ 4-15 และรูปที่ 4-16 (ข)



รูปที่ 4-16 (ก) แสดงการลดขนาดของรหัสรวม EEEES

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4-16 (ข) แสดงการลดขนาดของรหัสรวม OOEEE

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

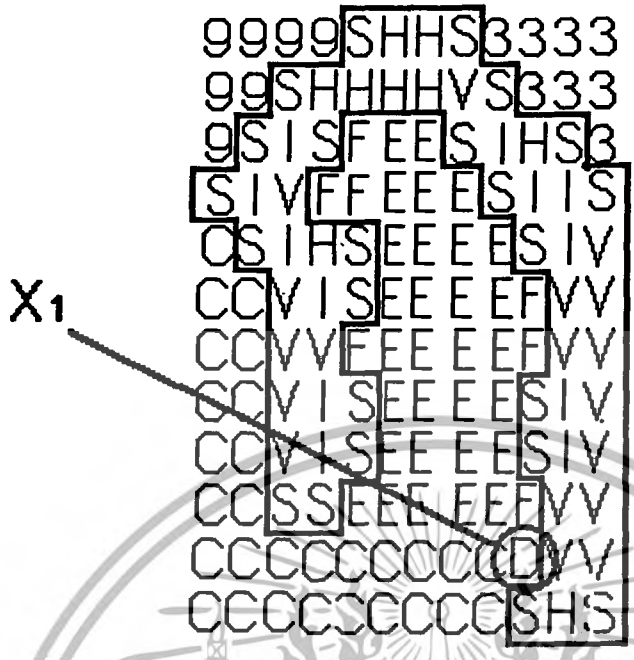
บทที่ 5
การปรับปรุงการปรับรหัสพื้นตัวอักษรเพิ่มเติมและการสร้างพจนานุกรม

จากการศึกษาทฤษฎีต่างๆที่เกี่ยวข้อง และทำการเขียนโปรแกรมคอมพิวเตอร์ตามขั้นตอนของการทำงาน เพื่อให้คอมพิวเตอร์สามารถรู้จำอักษรคัดลายมือภาษาไทยและภาษาอังกฤษ ซึ่งมีขั้นตอนในการทำงานอยู่ด้วยกัน 3 ส่วนหลักๆคือ Pre-processing เป็นส่วนของการเตรียมข้อมูลเบื้องต้นเพื่อเข้าสู่การประมวลผล ในขั้นตอนนี้เริ่มต้นตั้งแต่การเก็บข้อมูลจากเครื่องตรวจกวาดภาพ ทำการแปลงข้อมูลให้อยู่ในรูป 0-1 เพื่อใช้ในการประมวลผล และทำการจัดแบ่งอักษรภาพออกจากกัน (แยกอักษรแต่ละตัวออกจากกัน) ซึ่งเรียกกันว่า Segmentation ส่วนการทำงานขั้นต่อไปเป็นขั้นตอน Processing ซึ่งมีลำดับการทำงานแบ่งออกได้เป็น 3 ส่วนย่อย ได้แก่ 1.การให้รหัสพื้นฐาน (Initial Feature Extraction) 2.การปรับรหัสพื้นตัวอักษร (Unification) และ 3.การหารหัสรวม (Concentration) และส่วนการทำงานขั้นตอนสุดท้ายเป็นขั้นตอน Post-processing เป็นขั้นตอนในการพิจารณาหาลักษณะเด่นของอักษรแต่ละตัว เพื่อใช้รหัสแทนอักษรตัวนั้นๆ ซึ่งบทสรุปของการพิจารณารหัสรวมที่ใช้แทนนั้นได้จากการทดลองแจกแจงความถี่ของรหัสรวม (Concentrate Code) ของอักษรแต่ละตัวทั้งหมด โดยได้ข้อสังเกตว่ารหัสรวมที่มีความถี่สูงสุด 5 อันดับแรกนั้นจะมีลักษณะที่แตกต่างกันของแต่ละอักษร ใช้เป็นลักษณะเด่นแทนลักษณะของอักษรตัวนั้นได้ และจัดเก็บลักษณะเด่นเหล่านี้ให้อยู่รวมกันเพื่อเป็นพจนานุกรม (ฐานความรู้) ของการรู้จำตัวอักษร และใช้ทดสอบการรู้จำอักษรได้ต่อไป

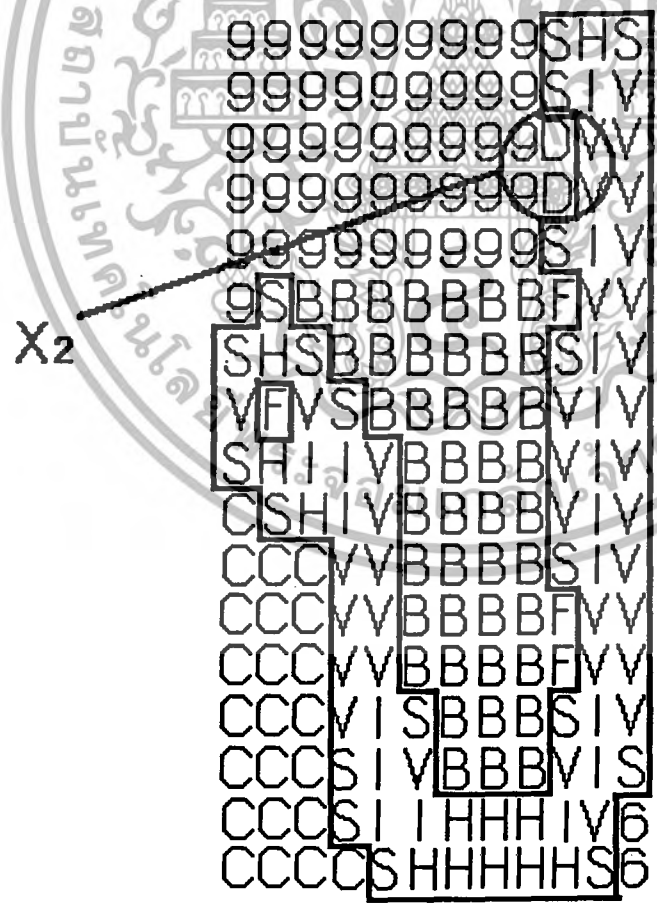
5.1 การปรับรหัสพื้นตัวอักษรเพิ่มเติม

จากการศึกษาและทดลองตามวิธีการในการรู้จำตัวอักษรดังกล่าวแล้วนั้น ยังปรากฏลักษณะของข้อมูลที่ผ่านกระบวนการในการปรับรหัสพื้นตัวอักษร ตามเงื่อนไขในตารางที่ 4-3 ยังมีลักษณะที่ไม่แสดงรหัสโครงสร้างโดยรวมของตัวอักษรอย่างชัดเจน เนื่องจากกระบวนการการรู้จำตัวอักษรในขั้นตอนของการปรับรหัสพื้นตัวอักษรนั้นยังไม่ครอบคลุมลักษณะอักษรที่เข้ามาทดสอบ ดังจะเห็นได้จากรูปที่ 5-1 (ก) และ (ข) จึงต้องทำการเพิ่มเงื่อนไขของการปรับรหัสพื้นตัวอักษรให้มากขึ้น เพื่อให้ครอบคลุมลักษณะอื่นๆโดยเฉพาะอักษรไทยอีก ดังแสดงในตารางที่ 5-1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5-1 (ก) แสดงข้อมูลที่มีโครงสร้างโดยรวมไม่ชัดเจนของอักษร "ก"



รูปที่ 5-1 (ข) แสดงข้อมูลที่มีโครงสร้างโดยรวมไม่ชัดเจนของอักษร "ป"

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

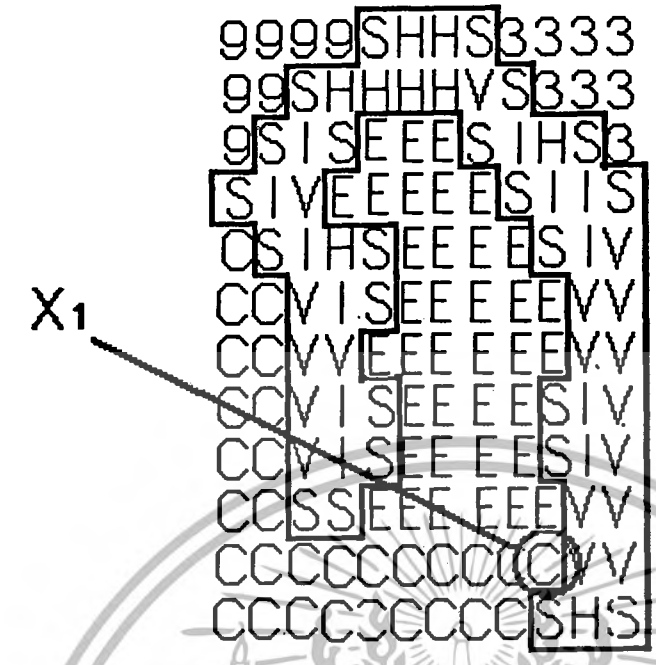
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รหัสแถวที่ 1	รหัสแถวที่ 2	รหัสแถวที่ 3	อักษรที่สอดคล้อง
			ก ถ ป ภ ฝ ฟ ฤ
			ธ ฐ ส ศ ษ ๔ ๕
			ล ส จ ฉ

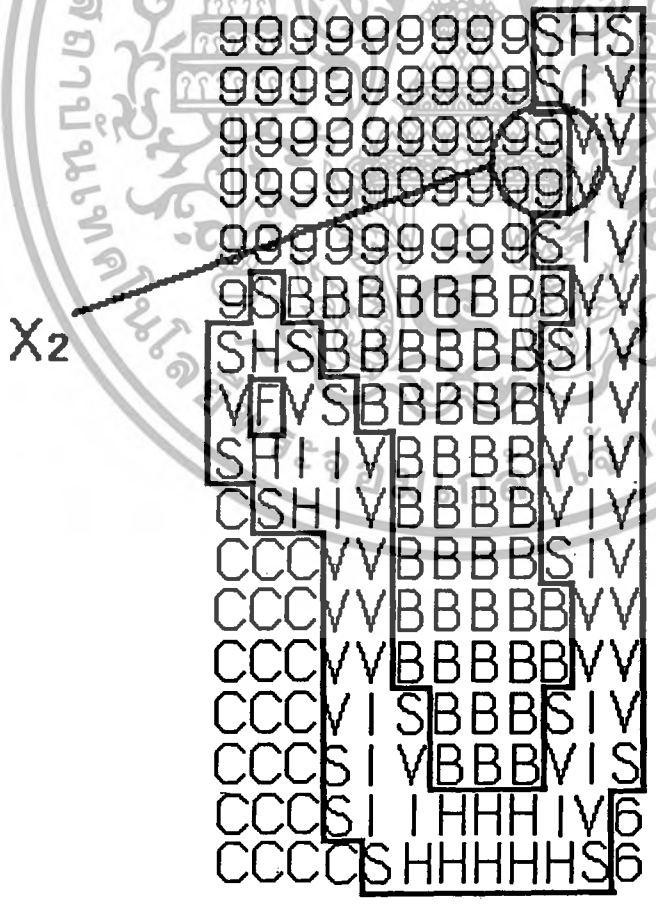
ตารางที่ 5-1 แสดงการปรับรหัสพื้นตัวอักษรเพิ่มเติม

จากรูปที่ 5-1 (ก) จะเห็นได้ว่ารหัสพื้นตัวอักษรที่ผ่านขั้นตอนของการปรับรหัสพื้นตัวอักษรในตำแหน่ง X1 ซึ่งได้แก่รหัส D นั้นไม่แสดงโครงสร้างของรหัสโครงสร้างโดยรวมที่ถูกต้องด้วยตำแหน่งดังกล่าวแสดงอยู่ในบริเวณส่วนเปิดด้านล่างของอักษร "ก" และจากรูป 5-1 (ข) ในตำแหน่ง X2 ซึ่งได้แก่รหัส D นั้นไม่แสดงโครงสร้างของรหัสโครงสร้างโดยรวมที่ถูกต้องด้วยตำแหน่งดังกล่าวแสดงอยู่ในบริเวณส่วนเปิดด้านบนเหนือส่วนหัวของอักษร "ป" ดังนั้นรหัส D ที่ได้เกิดจากความคลาดเคลื่อนของการเขียนอักษรและการสแกนเก็บภาพ ทำให้เกิดลักษณะเว้าแว่งไม่แสดงรหัสที่ถูกต้อง จึงต้องเพิ่มเงื่อนไขของการปรับรหัสพื้นตัวอักษรจากตารางที่ 4-3 เพื่อแสดงรหัสโครงสร้างโดยรวมและยังเป็นการลดรหัสที่ไม่เกี่ยวข้องออก ลดเงื่อนไขของการพิจารณารหัสที่เกิดจากความคลาดเคลื่อนในการเขียนและการสแกนเก็บภาพ ดังแสดงผลในรูป 5-2 (ก) และ (ข)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5-2 (ก) แสดงข้อมูลที่ผ่านการปรับรหัสพื้นอักษรของอักษร "ก"



รูปที่ 5-2 (ข) แสดงข้อมูลที่ผ่านการปรับรหัสพื้นอักษรของอักษร "ป"

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2 การสร้างพจนานุกรมการรู้จำอักษร

จากกระบวนการในการรู้จำตัวอักษรทั้ง 3 ขั้นตอน จนกระทั่งได้รหัสรวมซึ่งแสดงโครงสร้างของอักษรของทุกจุดภาพ จึงได้พิจารณาหาลักษณะที่จะใช้แทนอักษรตัวนั้นๆ โดยการใช้วิธีการทางสถิติ ด้วยการแจกแจงความถี่ของรหัสรวมของทุกจุดภาพในแต่ละตัวอักษร ซึ่งสามารถแสดงผลการแจกแจงความถี่ของอักษร "ก" ดังในตารางที่ 5-2 (ก) อักษร "บ" ดังในตารางที่ 5-2 (ข) อักษร "B" ดังในตารางที่ 5-2 (ค) และอักษร "D" ดังในตารางที่ 5-2 (ง)

รหัส-ความถี่	รหัส-ความถี่	รหัส-ความถี่	รหัส-ความถี่	รหัส-ความถี่
060e-225	1e0e-79	020e-75	700c-36	7a01-27
00d3-27	0741-25	7a04-25	8751-21	0353-21
0701-19	701c-18	0764-18	260e-18	a60e-18
7e54-16	860e-15	840e-15	8752-14	800e-12
8019-11	80d1-10	7a14-10	7e51-8	0363-9
805e-9	7011-9	7e51-8	3e0e-8	be0e-8
808e-7	0343-7	820e-6	80de-6	0704-6
6019-6	80d8-6	a20e-6	220e-6	0303-6
1f04-5	065e-5	7e52-4	7a02-4	0f44-3
7091-3	00c3-3	00e3-3	7014-3	0083-3
8758-3	7012-2	60d2-2	7018-2	8701-2
0762-2	a744-2	7092-2	0f06-2	70d1-2
0f42-2	1f52-2	7f52-2	0747-2	a701-2
80d2-2	7a12-2	84d2-2	7c51-2	780c-1
7c58-1	7212-1	70d4-1	7094-1	7e58-1
7052-1	60d1-1	8059-1	3f04-1	3f02-1
2742-1	2702-1	2701-1	0754-1	0752-1
825e-1	8352-1	8358-1	0706-1	848e-1
84d1-1	bf04-1	bf02-1	90d2-1	8702-1

ตารางที่ 5-2 (ก) แสดงการแจกแจงความถี่ของรหัสรวมของอักษร "ก"

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รหัส-ความถี่	รหัส-ความถี่	รหัส-ความถี่	รหัส-ความถี่	รหัส-ความถี่
025b-86	024b-63	a25b-60	d00c-51	dc0c-50
a24b-40	d601-29	c00c-25	03db-24	01c1-24
d604-23	01c4-22	a1c4-22	03cb-16	da14-14
156b-14	00db-12	1606-12	164b-12	360e-11
b65b-10	0dc4-9	00cv-8	d781-8	80db-6
0dc7-6	700c-6	005b-6	a35b-6	b64b-6
d651-5	05e7-5	05e4-5	d658-5	80cb-4
004b-4	f011-4	e014-4	c019-4	7c0c-4
da12-4	1658-4	d81c-4	c01f-4	d674-4
3741-4	c091-4	00fb-4	a34b-4	c0f8-4
1648-3	d7f3-3	1651-3	c0f1-3	7601-3
f014-3	0de4-3	0de2-3	d652-2	c011-2
1c0c-2	d641-2	d701-2	05e2-2	300c-2
d7f1-2	81c2-2	c05f-2	80fb-2	808b-2
0f06-2	d401-2	c092-2	0306-2	c211-2
3758-2	0241-2	0206-2	100c-2	1601-2
3748-2	3602-2	3742-2	1641-2	3c0c-2
600c-2	c394-2	c391-2	000c-1	f012-1
01e3-1	dfd2-1	de58-1	de11-1	de0e-1
dc51-1	dc18-1	d7f4-1	03fb-1	d774-1
d672-1	d642-1	0dc2-1	d488-1	d484-1
d482-1	d481-1	0e06-1	0e42-1	d474-1
d212-1	145b-1	15c2-1	d094-1	d092-1

ตารางที่ 5-2 (ข) แสดงการแจกแจงความถี่ของรหัสรวมของอักษร "บ"

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รหัส-ความถี่	รหัส-ความถี่	รหัส-ความถี่	รหัส-ความถี่	รหัส-ความถี่
82ff-66	1ecf-54	e2c1-41	e01d-38	03c3-21
07c1-21	e2c4-16	e00d-16	e284-16	87c1-14
3ecf-13	9ecf-13	8078-12	9ef8-11	1e06-10
87f1-10	3c48-9	9ef1-7	87e1-7	07c4-7
82c1-6	0043-5	27c1-5	3ec1-5	0606-5
3ec8-5	87e4-5	0406-5	87e2-4	801d-4
9ff2-4	02c3-4	0063-4	87f4-3	22c1-3
8071-3	2048-3	86ff-3	3fc1-3	0206-3
0686-3	06c3-3	e281-3	06cf-3	8038-3
8041-3	0786-3	8048-3	800d-2	8084-2
27c2-2	86f1-2	8478-2	07c2-2	0073-2
9ff8-2	3fc2-2	9ff1-2	87c2-2	83f2-2
03e3-1	fef2-1	fec2-1	fa82-1	e2f2-1
e2b4-1	e2b2-1	07c3-1	9ff4-1	07f2-1
07f4-1	87f2-1	86c4-1	1fc2-1	2001-1
200c-1	200d-1	86c2-1	2202-1	2281-1
86c1-1	26c2-1	26c8-1	8442-1	83f1-1
3801-1	380c-1	3812-1	3818-1	3a02-1
3a81-1	3a92-1	3c42-1	8072-1	8032-1
8031-1	8010-1	8008-1	8001-1	3fc8-1

ตารางที่ 5-2 (ค) แสดงการแจกแจงความถี่ของรหัสรวมของอักษร "B"

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รหัส-ความถี่	รหัส-ความถี่	รหัส-ความถี่	รหัส-ความถี่	รหัส-ความถี่
e00c-61	824f-59	024f-56	224f-47	e201-45
064f-30	07c1-26	03c3-21	1e06-19	e204-18
0786-17	3e48-16	864f-16	0606-15	020f-14
8087-13	3fc1-13	87f1-13	87c1-12	0043-10
264f-9	fa14-8	220f-8	0073-8	27c1-8
07c4-7	3fc2-6	8071-6	2201-62	87f2-5
27c2-4	8018-40	200c-4	87c4-3	3a02-3
87f4-3	3a01-3	87c2-3	2208-3	fe04-3
3e01-3	e214-3	8008-3	8001-3	83f8-2
3e41-2	fa12-2	fa04-2	83f2-2	f80c-2
07c2-2	e272-2	3e08-2	8011-2	3e42-2
87f8-2	3a08-2	1fc4-2	e212-2	e211-2
1fc2-1	1f86-1	07f4-1	07f2-1	03f3-1
fe02-1	27c4-1	fa02-1	e271-1	e01c-1
87c8-1	3e02-1	8478-1	8472-1	83f1-1
8072-1	8012-1	8000-1	3fc4-1	3fc8-1

ตารางที่ 5-2 (ง) แสดงการแจกแจงความถี่ของรหัสรวมของอักษร "D"

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางแจกแจงความถี่ที่ 5-2 (ก) (ข) (ค) และ (ง) โดยได้ใช้วิธีการแจกแจงความถี่ของรหัสรวมกับอักษรหลายๆตัว จะสังเกตได้ว่ารหัสรวมที่มีความถี่จากการนับสูงจะแสดงลักษณะส่วนใหญ่หรือแทนลักษณะส่วนใหญ่ของอักษรตัวนั้นๆ และจากการทดลองแจกแจงความถี่ของอักษรคัดลายมือทั้งภาษาไทยและภาษาอังกฤษ พบว่าถ้าหากใช้รหัสรวมที่มีความถี่สูง ซึ่งเป็นลักษณะเฉพาะแทนอักษรแต่ละตัว เป็นเงื่อนไขในการแยกแยะอักษร สามารถแยกพิจารณาได้ เช่น ถ้าพิจารณาเฉพาะรหัสรวมที่มีความถี่สูงสุด 1 อันดับ จะสามารถแยกแยะอักษรคัดลายมือภาษาไทยออกเป็นกลุ่มอย่างคร่าวๆได้ดังแสดงในตารางที่ 5-3

รหัสรวมที่พิจารณา	กลุ่มอักษร
060e	ก - ถ
760f	ช - ซ
030e	ค - ด - ต
025b	ฆ - ม
120b	บ - ป
158e	ท - ท
a05b	ผ - ฝ
160e	พ - ฟ
84dc	ฮ - ฮ
831b	ฌ - ญ - ฒ

ตารางที่ 5-3 แสดงการแยกแยะอักษรด้วยรหัสรวมที่มีความถี่สูงสุด

ลักษณะการแยกแยะด้วยรหัสรวมที่มีความถี่สูงสุดด้วยรหัสเดียว นั้นไม่สามารถแยกอักษรออกจากกันได้อย่างชัดเจน เนื่องมาจากลักษณะทางโครงสร้างของอักษรในภาษาไทยนั้นมีลักษณะที่ใกล้เคียงกันมาก และเพื่อให้การพิจารณาแยกแยะอักษรออกจากกันด้วยรหัสรวมได้อย่างชัดเจน จึงต้องทำการพิจารณารหัสรวมที่มีความถี่สูงในอันดับรองลงไป และจากการทดลองพิจารณารหัสรวมของอักษรคัดลายมือที่ไม่สามารถแยกด้วยรหัสรวมที่มีความถี่สูงสุด จะปรากฏว่าผลการแจกแจงความถี่ในช่วงที่น่าสนใจ ดังแสดงในตารางที่ 5-4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อักษรชุด	ลำดับรหัสรวมที่ได้จากการแจกแจงความถี่ในทางสถิติ				
ม	225b	d00c	220b	d01c	7c7c
ฆ	225b	d00c	220b	d47c	365b
บ	120b	024b	a25b	d00c	dc0c
ป	120b	024b	a25b	d00c	a059
ฌ	831b	c60e	c6fe	c28e	d751
ญ	831b	c60e	de1e	d2ae	8581
ฒ	831b	d75e	d751	7f51	05c1
ฏ	84dc	601a	98fc	90fc	849c
ฏ	84dc	601a	98fc	18dc	7091
ผ	a05b	948e	81db	d081	81d1
ฝ	a05b	948e	8059	8079	160e
พ	160e	d58e	835b	625b	01c3
ฟ	160e	d58e	835b	625b	8059

ตารางที่ 5-4 แสดงการแจกแจงความถี่รหัสรวมของอักษร

จากผลการทดลองแจกแจงความถี่รหัสรวม ที่เกิดขึ้นกับอักษรคัดลายมือแต่ละตัว จะได้ข้อสรุปว่า เราสามารถแยกแยะอักษรคัดลายมือภาษาไทยทุกๆตัวอักษรออกจากกัน ด้วยการพิจารณารหัสรวมที่มีความถี่สูงสุด 5 อันดับแรกของอักษรแต่ละตัวเป็นเงื่อนไขในการแยกอักษรแต่ละตัวออกจากกัน ซึ่งสามารถแสดงในตารางที่ 5-5

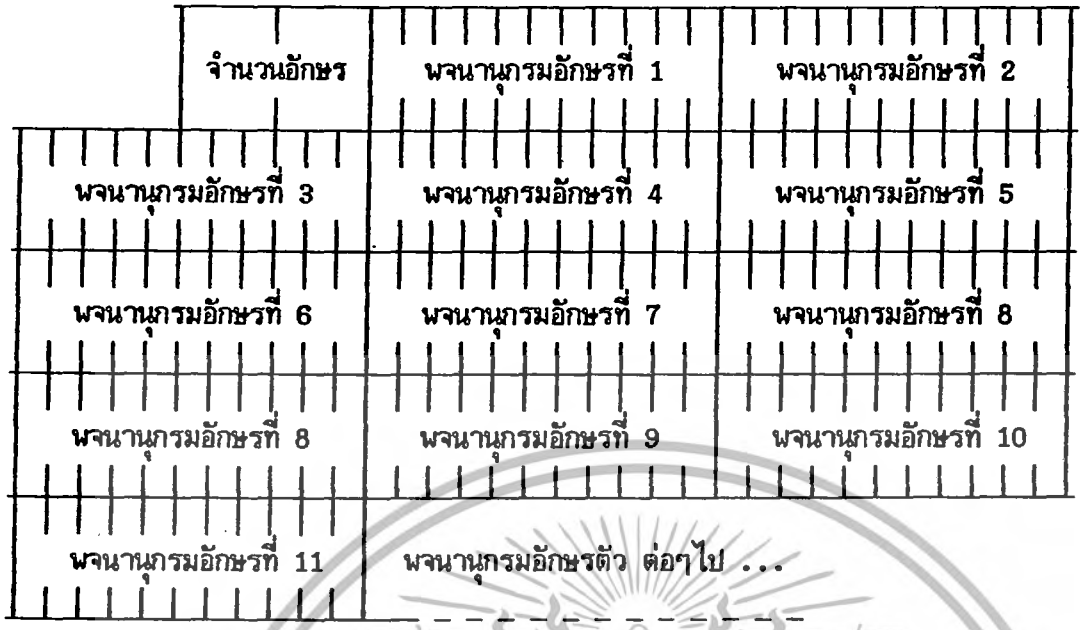
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอักษร	รหัสโครงสร้างรวม (Concentrate Code)				
	ลำดับที่ 1	ลำดับที่ 2	ลำดับที่ 3	ลำดับที่ 4	ลำดับที่ 5
ก	060e	1e0e	020e	700c	7a01
ด	030e	1a0c	6019	7a18	709c
ป	025b	024b	a25b	d00c	a059
ร	1606	300c	1206	301c	0606
ษ	01c3	1606	1586	807b	0586
◌	1c0c	100c	8019	00f3	fc11
◌	1c0c	100c	d49f	00fb	c019
◌	3c0c	8079	300c	f09f	3e11
◌	3c0c	c079	c019	d09f	1c0c
A	9c8e	e059	8059	e009	03d3
C	00c3	8069	8019	1093	1083
U	000b	220b	05c8	c018	05c2

ตารางที่ 5-5 แสดงตัวอย่างรหัสแทนอักษรบางตัวในพจนานุกรม

ในการออกแบบพจนานุกรมมีหลักการว่า พจนานุกรม(ฐานความรู้)ของการรู้จำที่ได้ต้องสามารถเขียน อ่านข้อมูลอย่างถูกต้องและรวดเร็ว และจะต้องครอบคลุมตัวอักษรที่ได้กำหนดและเก็บข้อมูลในฐานความรู้ (พจนานุกรม) ของการรู้จำของทุกตัวที่กำหนดทั้งภาษาไทยและภาษาอังกฤษ จึงได้ออกแบบพจนานุกรมโดยมีโครงสร้าง ดังแสดงในรูปที่ 5-4 ให้มีการเก็บลักษณะแทนอักษรแต่ละตัวในรูปของรหัสรวม 5 ตัวๆละ 2 byte รวมเป็น 10 byte แล้วตามด้วยรหัสแอสกี้ของอักษรตัวนั้นอีก 1 byte รวมเป็น 11 byte สำหรับหนึ่งตัวอักษร โดยจะเว้นที่ 2 byte แรกของแม่ข้อมูลพจนานุกรมนี้ ไว้เก็บค่าจำนวนของข้อมูลอักษรที่ได้จัดเก็บไว้ในพจนานุกรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5-3 แสดงโครงสร้างของพจนานุกรม

เพื่อให้การค้นหาพจนานุกรมเป็นไปอย่างมีประสิทธิภาพ สามารถค้นหารหัสแทนแต่ละอักษรได้อย่างรวดเร็ว จึงได้ให้มีการจัดเก็บรหัสแทนอักษรแต่ละตัวในพจนานุกรม ตามลำดับการใช้งานของอักษรแต่ละตัว โดยอ้างอิงรายงานวิจัยของคุณสฤษดิ์ ปัตตะโชติ เรื่องการกำหนดแป้นพิมพ์แบบปรับปรุงใหม่ 4 พ.ย. 2509 ซึ่งได้นำเสนอการกำหนดตำแหน่งการวางตัวอักษรแต่ละตัวบนแป้นพิมพ์คีตที่ได้ปรับปรุงขึ้นใหม่ โดยอาศัยค่าของความถี่ในการใช้งานของอักษรแต่ละตัว ซึ่งสามารถแสดงดังตารางที่ 5-6

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลำดับที่	ตัวอักษร	จำนวน	ลำดับที่	ตัวอักษร	จำนวน	ลำดับที่	ตัวอักษร	จำนวน
1	า	73.92	2	น	61.08	3	ร	49.82
4	◌ั	45.00	5	ก	44.14	6	ง	43.52
7	◌ะ	41.26	8	อ	39.00	9	เ◌	37.08
10	ม	32.48	11	ย	29.02	12	◌ี	26.82
13	◌ิ	26.08	14	ว	25.80	15	ด	23.96
16	ล	23.80	17	ท	23.64	18	◌ะ	21.68
19	ป	20.74	20	◌ื	19.70	21	ห	19.62
22	ต	18.84	23	ส	17.90	24	บ	17.76
25	จ	16.50	26	ค	15.32	27	◌ไ	14.80
28	แ◌	13.82	29	ข	13.46	30	◌ใ	12.74
31	พ	11.88	32	◌ำ	11.46	33	◌่	10.70
34	ช	10.26	35	◌ี	9.20	36	◌้	7.78
37	◌ุ	7.68	38	◌ู	5.40	39	ผ	5.20
40	โ◌	5.70	41	◌ะ	4.72	42	ศ	4.58
43	ถ	4.32	44	◌็	3.72	45	ษ	2.88
46	ณ	2.68	47	ญ	2.66	48	ซ	2.30
49	ภ	2.22	50	ฮ	2.10	51	ๆ	1.96
52	ฝ	1.36	53	ฐ	1.28	54	ฎ	0.96
55	ฟ	0.80	56	ฉ	0.76	57	ฤ	0.66
58	"	0.60	59	0	0.54	60	2	0.50
61	ฏ	0.48	62	4	0.46	63	.	0.40
64	5	0.36	65	1	0.32	66	ฒ 3 ฯ	0.30
67	,	0.18	68	6	0.16	69	ฆ 7	0.14
70	ฬ 9	0.12	71	◌๊ 8 ◌๋	0.10	72	( )	0.08
73	-	0.06	74	ฑ %	0.04	75	ฌ ฮ /	0.02

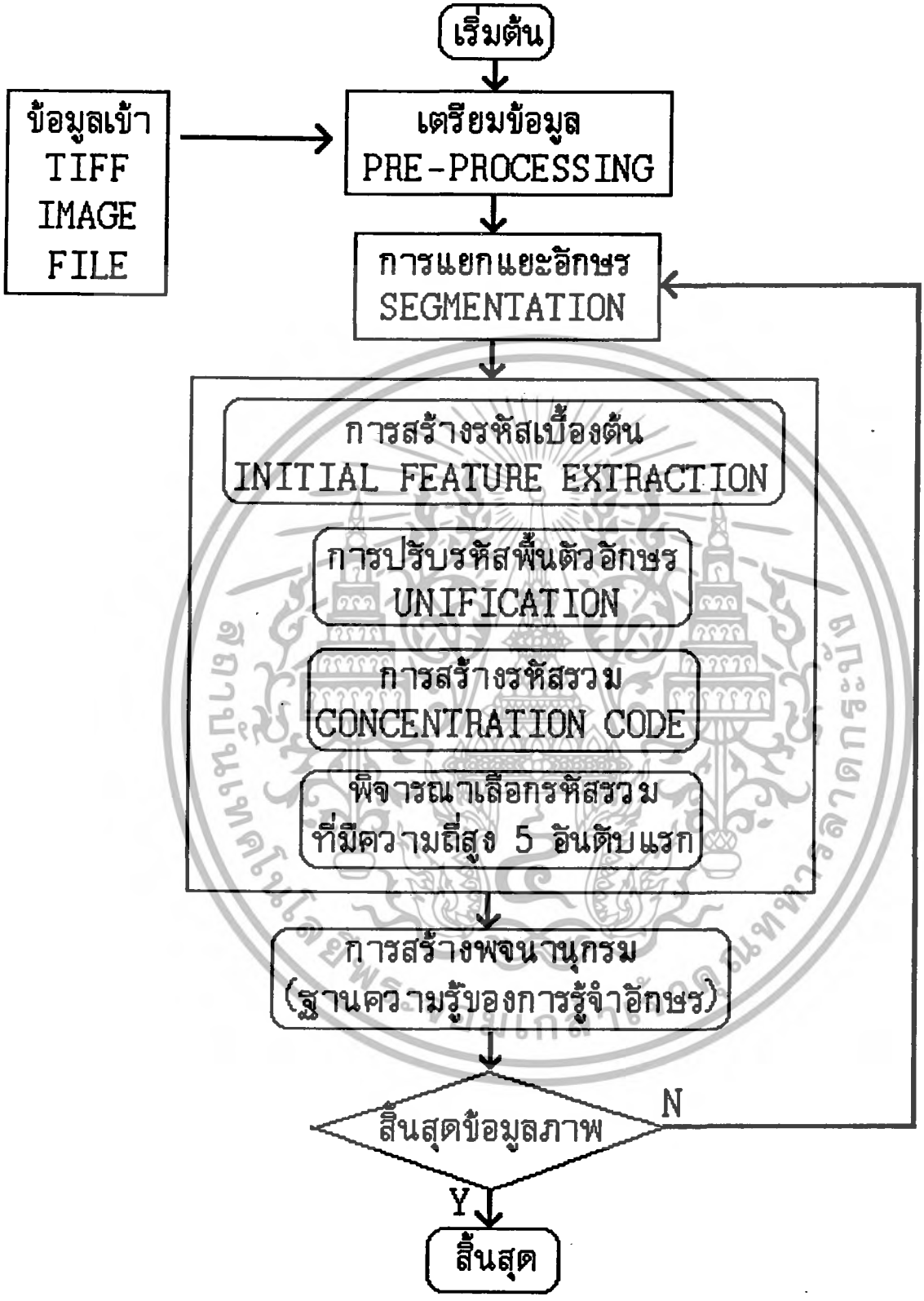
ตารางที่ 5-6 ตารางแจกแจงความถี่ที่คุณสฤษดิ์ ปัตตะโชติใช้ในโครงการวิจัย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

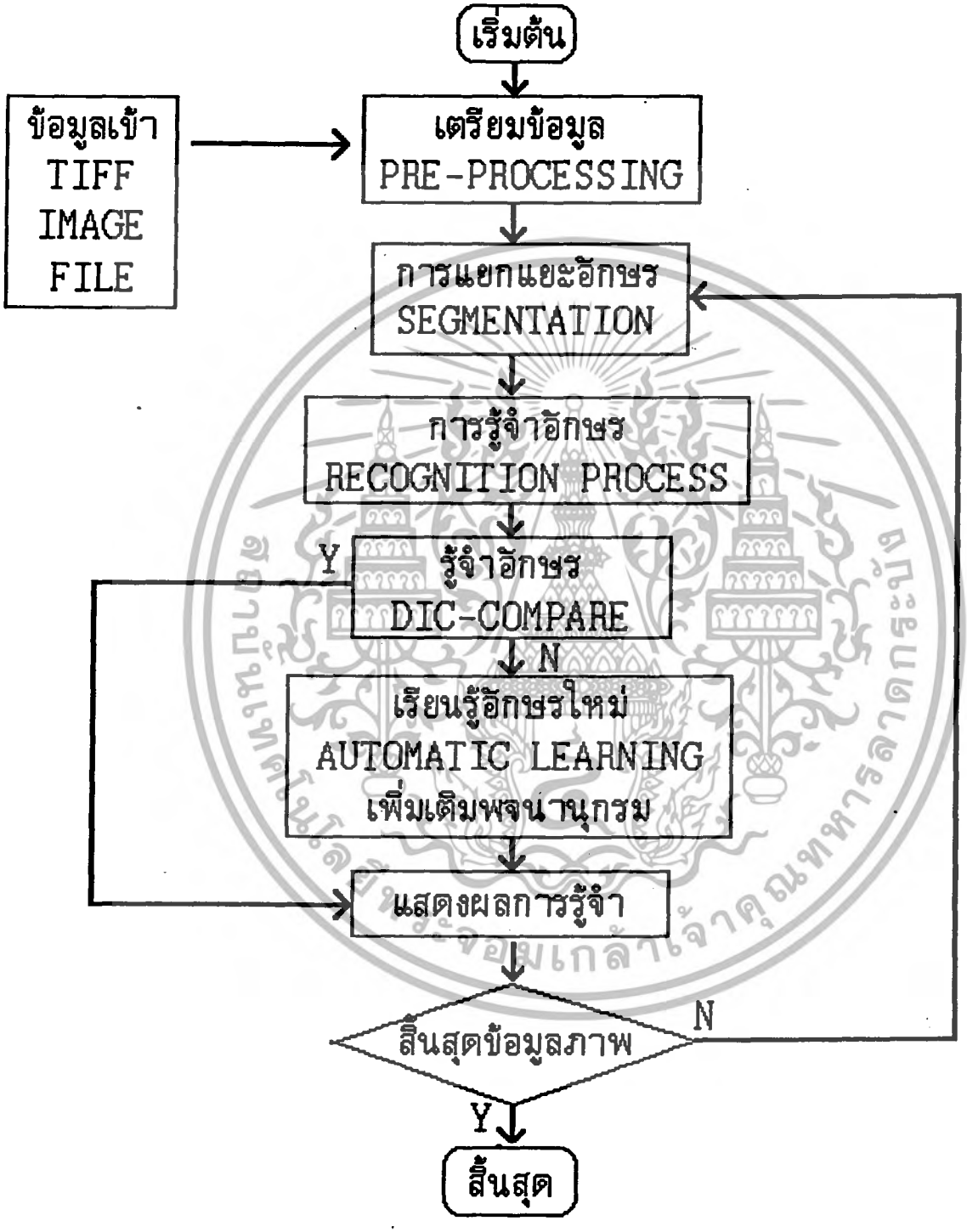
จากโครงสร้างของพจนานุกรมในรูปที่ 5-3 และตารางแจกแจงความถี่การใช้งานอักษรในตารางที่ 5-6 จะได้โครงสร้างพจนานุกรมที่มีลำดับการจัดเก็บและการค้นหาตามตัวอักษรที่มีความถี่ในการใช้งานมากตามลำดับ และต้องจัดทำพจนานุกรมการเรียนรู้จากอักษรคัดลายมืออักษรละหลายๆครั้ง เพื่อให้เกิดความหลากหลายในการรู้จำที่นำมาใช้เป็นต้นแบบในการเรียนรู้เบื้องต้น แล้วจึงจัดการทดสอบกับอักษรที่เขียนมาครั้งใหม่ โดยถ้าส่วนของโปรแกรมสามารถรู้จำอักษรที่เข้ามาได้ก็จะส่งรหัสแอสกี้ของอักษรตัวนั้นออกมาพร้อมทั้งแสดงอักษรที่ได้ แต่หากโปรแกรมไม่สามารถรู้จำอักษรดังกล่าวได้ก็จะแสดงภาพของอักษรออกมาให้ผู้ใช้ได้ตัดสินใจ เพื่อให้ป้อนความรู้ใหม่ และจัดเก็บรวมต่อท้ายส่วนของพจนานุกรม (ฐานความรู้ของการรู้จำ) เดิมที่มีอยู่ก่อน และสามารถสรุปขั้นตอนต่างๆ ดังโฟล์ชาร์ตในรูปที่ 5-4 (ก) และ (ข)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5-4 (ก) โฟล์ชาร์ตแสดงการทำงานของการรู้จำอักษรคัดลายมือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5-4 (ข) โฟล์ชาร์ตแสดงโปรแกรมการรู้จำอักษรคัดลายมือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 6
โปรแกรมการรู้จำอักษรคัดลายมือ

โปรแกรมการรู้จำอักษรที่สอนให้เครื่องไมโครคอมพิวเตอร์ ได้เรียนรู้และรู้จำอักษรต่างๆนั้น ได้มีการพัฒนา ปรับปรุงกันอย่างกว้างขวาง ทั้งในรูปแบบของอักษรภาษาไทยและอักษรภาษาอังกฤษ ทั้งที่เป็นอักษรคัดลายมือตัวบรรจงและอักษรตัวพิมพ์ ด้วยเทคนิควิธีการต่างๆกันไป โปรแกรมการรู้จำอักษรคัดลายมือด้วยวิธีการพิจารณารหัสรวม เป็นอีกโปรแกรมหนึ่งซึ่งได้พัฒนาให้สามารถสอนเครื่องไมโครคอมพิวเตอร์ ให้รู้จำอักษรคัดลายมือทั้งภาษาไทยและภาษาอังกฤษ โดยมีรายละเอียดของโปรแกรมดังนี้

ลักษณะเฉพาะของโปรแกรม

โปรแกรมการรู้จำอักษรคัดลายมือ ถูกเขียนขึ้นด้วยโปรแกรมภาษาซีเวอร์ชั่น 2 โดยในแต่ละโปรแกรมจะมีรายละเอียดของโปรแกรมย่อยต่อไปนี้

- ส่วนการจัดการข้อมูลเบื้องต้นจากเครื่องตรวจกวาดภาพ
- ส่วนการจัดการรูปแบบเฉพาะของ TIFF FILE และแปลงข้อมูล
- ส่วนการแบ่งแยกอักษรออกมาพิจารณาเฉพาะส่วน
- ส่วนการสร้างรหัสเบื้องต้น
- ส่วนการปรับรหัสพื้นตัวอักษร
- ส่วนการสร้างรหัสรวม
- ส่วนการพิจารณาลักษณะเด่น(รหัสรวม)
- ส่วนการสร้างฐานความรู้ของการรู้จำอักษร

อุปกรณ์ในการใช้งาน

1. เครื่องไมโครคอมพิวเตอร์ ซึ่งมีหน่วยความจำไม่น้อยกว่า 640 Kbyte
2. เครื่องตรวจกวาดภาพ (HEWLETT PACKARD SCANNER)
3. โปรแกรมการรู้จำอักษรคัดลายมือ
4. โปรแกรมการจัดการรหัสภาษาไทย สมอ. (THAITIP.COM)

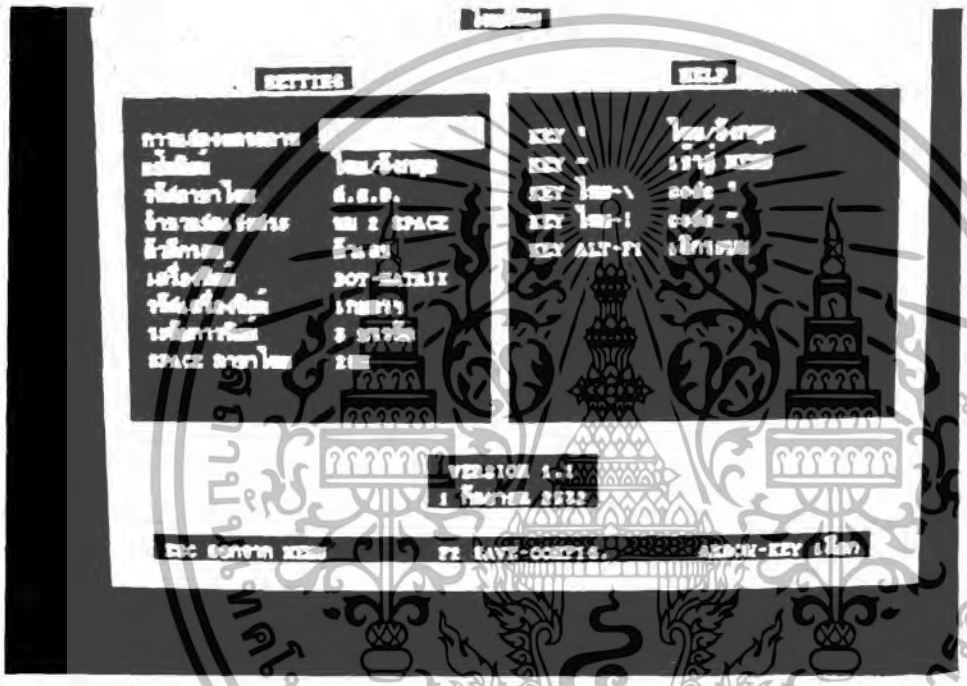
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนการใช้งาน

1. หลังจาก BOOT เครื่องไมโครคอมพิวเตอร์ด้วย DOS ให้ run program thai driver เพื่อการจัดการระบบภาษาไทย ด้วย THAITIP.COM

   ตัวอย่าง... A> THAITIP <ENTER>

   จัดการเปลี่ยนการแสดงผลรหัสภาษาไทยให้ รหัสสมอ. ดังในรูปที่ 6-1



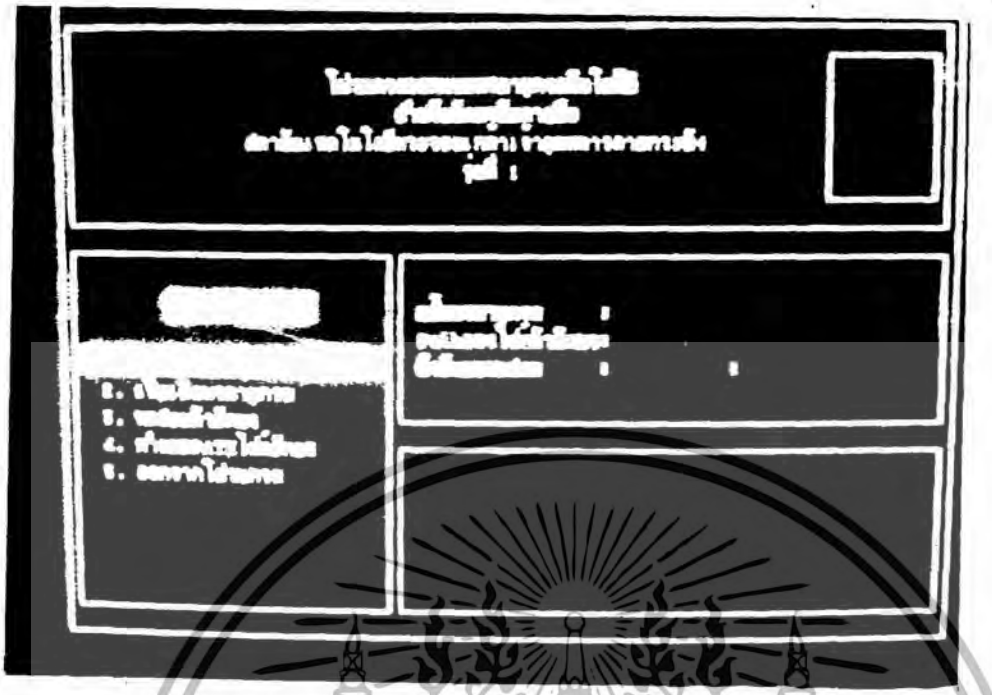
รูปที่ 6-1 การจัดการระบบการแสดงผลภาษาไทยด้วยโปรแกรม THAITIP

2. run โปรแกรมการเรียนรู้อักษรคัดลายมือ

   ตัวอย่าง... A> PATTEST <ENTER>

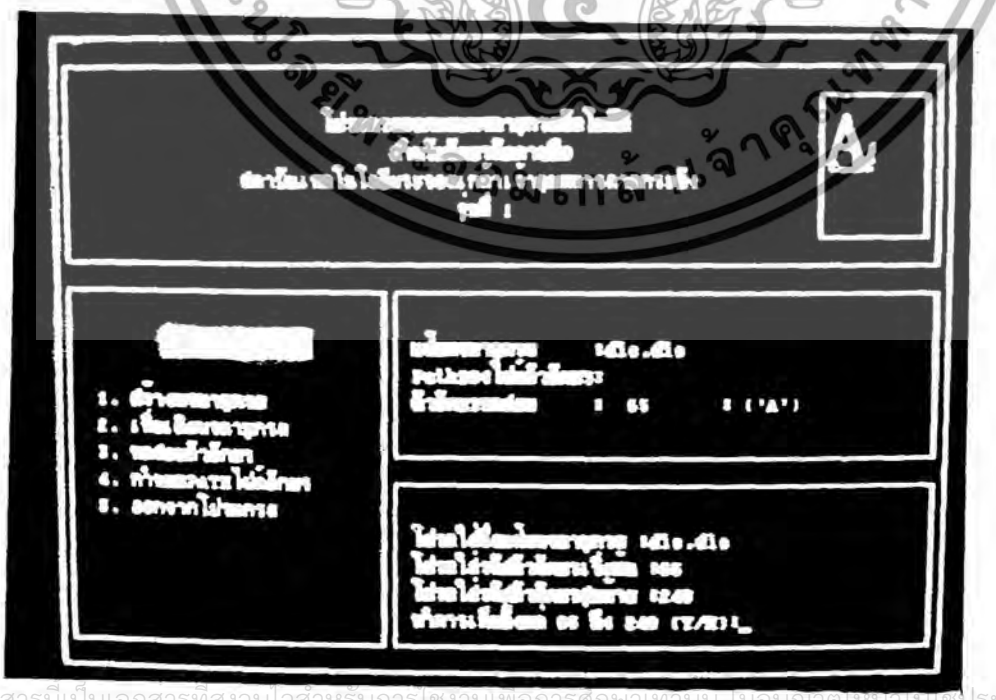
   จะได้ผลแสดงดังรูปที่ 6-2 โดยจะมี menu สำหรับการใช้งานอยู่ 5 รายการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



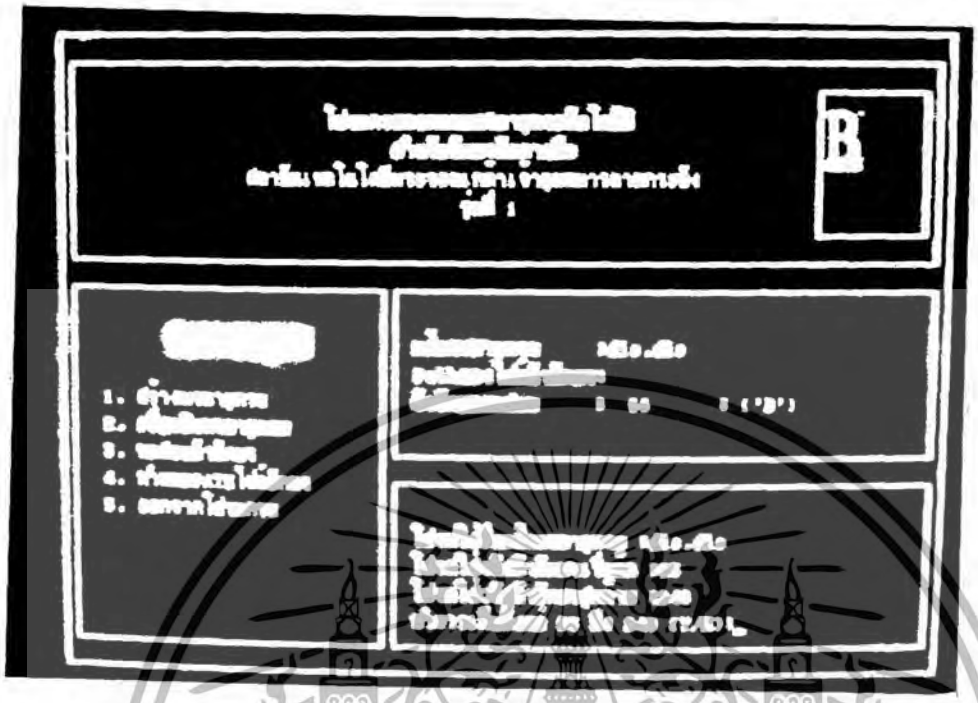
รูปที่ 6-2 แสดงรายละเอียดหน้าจอภาพของโปรแกรมการรู้จำ

1. การสร้างพจนานุกรม เป็นการสร้างฐานความรู้ของการรู้จำอักษรคัดลายมือเบื้องแรกก่อนการนำไปใช้งานต่อไป ดังในรูปที่ 6-3
2. การเพิ่มเติมพจนานุกรม เป็นการเพิ่มเติมฐานความรู้ ให้มีข้อมูลในการพิจารณาอักษรมากยิ่งขึ้นก่อนการใช้งาน ดังในรูปที่ 6-4



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 6-3 การสร้างพจนานุกรม



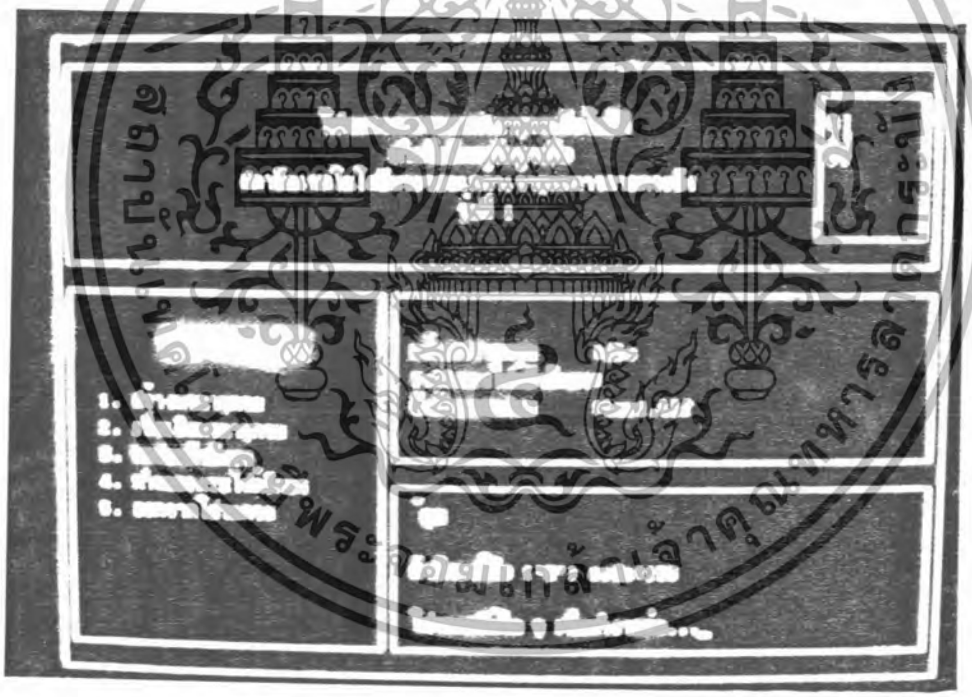
รูปที่ 6-4 การเพิ่มเติมพจนานุกรม

3. การทดสอบตัวอักษร เป็นการนำอักษรใดๆมาทดสอบการรู้จำอักษร ดังแสดงในรูปที่ 6-5

รูปที่ 6-5 การทดสอบตัวอักษร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ถ้าโปรแกรมการรู้จำอักษรคัดลายมือ สามารถบอกลักษณะของอักษรที่เข้ามาได้ จะแสดงผลการทดสอบด้วยลักษณะข้อมูลที่อ่านเข้ามา และบอกรหัสแอสกี้ของอักษรตัวนั้น ดังแสดงในรูปที่ 6-6
- ถ้าโปรแกรมการรู้จำอักษรคัดลายมือ ไม่สามารถบอกลักษณะของอักษรที่เข้ามาได้ จะแสดงผลการทดสอบด้วยลักษณะข้อมูลที่อ่านเข้ามา และถามหาลักษณะข้อมูลอักษรตัวนั้นๆ ว่าจะให้กำหนดเป็นอักษรใด โดยผู้ใช้จะต้องบอก ด้วยการป้อนอักษรให้กับเครื่องไมโครคอมพิวเตอร์ เป็นการเรียนรู้โดยอัตโนมัติ ดังในรูปที่ 6-7



รูปที่ 6-6 แสดงการทดสอบอักษร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 6-7 แสดงการเรียนรู้ด้วยเครื่องไมโครคอมพิวเตอร์

4. การกำหนด PATH ไฟล์ข้อมูล เป็นการกำหนด PATH ของข้อมูล tiff file เพื่อการอ่านมาประมวลผลการรู้จำอักษร
5. การออกจากโปรแกรมการรู้จำอักษรคัดลายมือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 7

ผลการทดลองโปรแกรมการรู้จำอักษร

จากผลการศึกษาและทดลองโปรแกรมการรู้จำอักษร โดยการพิจารณารหัส Concen -trate กับอักษรคัดลายมือภาษาไทย (พยัญชนะ 43 ตัว สระ 12 ตัว วรรณยุกต์ 4 ตัว อักขระพิเศษ 5 ตัว ตัวเลขไทย 10 ตัว) และอักษรคัดลายมือภาษาอังกฤษ(อักษรพิมพ์ใหญ่ 26 ตัวอักษร ตัวพิมพ์เล็ก 26 ตัว ตัวเลขอารบิค 10 ตัว) ปรากฏว่ามีความถูกต้องสูงโดยมีปัจจัยในการทดสอบขึ้นอยู่กับความเปลี่ยนแปลง ความไม่แน่นอน ไม่คงที่ของอักษรคัดลายมือของแต่ละบุคคลและจำนวนชุดของอักษรคัดลายมือที่นำมาใช้ทำเป็นฐานความรู้ของการรู้จำอักษร โดยอยู่ใช้ข้อมูลอักษรขนาดระหว่าง 400 - 600 byte บนเครื่องไมโครคอมพิวเตอร์ PC/AT ขนาด 12 bit ประมวลผลด้วยความเร็ว 12 MHz ดังรูปที่ 7-1 ใช้เวลาในการจัดทำฐานความรู้ของการรู้จำอักษร (พจนานุกรมการเรียนรู้) ประมาณ 1วินาทีต่อตัวอักษร และใช้เวลาในการทดสอบการรู้จำอักษรประมาณ 1.2 วินาทีต่อตัวอักษร

รูปที่ 7-1 เครื่องไมโครคอมพิวเตอร์และเครื่องตรวจกวาดภาพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากผลการศึกษาและโปรแกรมในส่วนของ การสร้างพจนานุกรมอัตโนมัติสำหรับการรู้จำอักษรคัดลายมือด้วยวิธีพิจารณาโครงสร้างของอักษร (Feature Concentration) และพิจารณารหัสรวม (Concentrate Code) ที่มีความถี่สูงสุด 5 อันดับแรกเป็นลักษณะแทนอักษรแต่ละตัว และด้วยเป็นโปรแกรมการรู้จำที่ใช้กับอักษรคัดลายมือ ซึ่งมีความไม่แน่นอนของลายเส้น ขนาดและทิศทาง พจนานุกรมที่สร้างไว้จึงไม่สามารถที่จะครอบคลุมลักษณะของอักษรได้ทั้งหมด จึงได้แก้ไขปัญหาโดยการให้เครื่องไมโครคอมพิวเตอร์ สามารถเรียนรู้อักษรตัวต่างๆได้เอง โดยในกรณีที่ป้อนข้อความแล้ว มีอักษรบางตัวมีลักษณะไม่ตรงในพจนานุกรมที่เก็บไว้ ก็ให้มีเรียนรู้ตัวอักษรนั้น โดยโปรแกรมจะถามกลับมายังผู้ใช้งานว่าเป็นตัวอักษรตัวอะไร เป็นการเพิ่มเติมข้อมูลในพจนานุกรม และโปรแกรมยังยืดหยุ่นให้สามารถเลือกใช้พจนานุกรมที่เกิดจากการเรียนรู้ลายมือใดๆก็ได้ เพื่อให้การรู้จำมีประสิทธิภาพมากขึ้น เมื่อเปรียบเทียบกับงานของการรู้จำรูปแบบอักษรตัวพิมพ์ภาษาไทย [27] และงานของการรู้จำอักษรลายมือเขียนภาษาไทยโดยการพิจารณาหัวตัวอักษร [29] ซึ่งมีเงื่อนไขของการเรียนรู้และเรียนรู้จากอักษรในลักษณะที่ต่างกัน จะได้ผลดังต่อไปนี้

การเปรียบเทียบงานด้านการรู้จำอักษร

1. ป้อนข้อมูลด้วยอักษรตัวพิมพ์

ลักษณะที่พิจารณา	1	2	3
ความเร็ว	1.0	*	1.1
ความถูกต้อง	99%	*	98%
เงื่อนไขเฉพาะ	-ไทย-	อักษรคัดลายมือ	

*** หมายเหตุ ***

1.การจดจำรูปแบบตัวอักษรตัวพิมพ์ภาษาไทย [27]

2.การรู้จำอักษรลายมือเขียนภาษาไทยโดยการพิจารณาหัวของตัวอักษร [29]

3.งานวิจัยที่ทำการศึกษา

* ไม่ได้พิจารณาด้วยติดเงื่อนไขของตัวอักษร

ตารางที่ 7-1 เปรียบเทียบในลักษณะของอักษรตัวพิมพ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากการศึกษาเปรียบเทียบ ผลงานหมายเลข 1 กับผลงานหมายเลข 3 จะสังเกตได้อย่างชัดเจนว่าในส่วนของ Pre-process และ Process นั้นมีวิธีการได้มาซึ่งรหัสแทนในรูปของรหัสรวม (Concentrate Code) เหมือนกันทุกประการ คงต่างกันในส่วนของการพิจารณาลักษณะเด่นที่จะนำรหัสรวมไปใช้เป็นลักษณะแทนของอักษรแต่ละตัว ในผลงานหมายเลข 1 นั้นได้พิจารณาใช้การแทนโดยใช้จุดศูนย์ถ่วงของตัวอักษร ทั้ง 9 ลักษณะ และข้อมูลเป้าหมายเป็นลักษณะของอักษรตัวพิมพ์ภาษาไทย ฐานความรู้ของการรู้จำ(พจนานุกรม)จัดทำเป็นแบบรวมเป็นอันเดียวเหมาะกับรูปแบบตัวอักษรที่มีลักษณะค่อนข้างคงรูปร่าง อย่างตัวพิมพ์ซึ่งจะได้ผลของความถูกต้องของการใช้งานสูง ส่วนผลงานหมายเลข 3 เป็นการศึกษาเปรียบเทียบกับวิธีดังกล่าว ได้พิจารณาใช้ความถี่สูงสุด 5 อันดับแรกเป็นลักษณะเด่นรูปแบบ เพื่อให้ใช้กับข้อมูลที่มีความยืดหยุ่นมากขึ้นอย่างลายมือเขียน ในส่วนของการทดสอบด้วยอักษรตัวพิมพ์นั้น ได้ผลเช่นเดียวกับผลงานหมายเลข 1 และสามารถแยกพจนานุกรมการรู้จำออกจากกันตามรูปแบบของอักษร (Fonts) สามารถเพิ่มเติมพจนานุกรม และยังสามารถเรียนรู้อักษรได้เอง ในกรณีที่ข้อมูลตัวอักษรที่ได้จากเครื่อง scanner มีความผิดพลาด ไม่ชัดเจนไม่ตรงกับเงื่อนไขปกติของอักษรที่ได้รับไว้โดยจะแสดงข้อมูลที่ได้ในรูปกราฟิค และการถามกลับมายังผู้ใช้งานว่าเป็นอักษรอะไร หรือจะกำหนดให้เป็นอักษรอะไร และเมื่อได้รับคำตอบก็จะทำการเพิ่มเติมเข้าไปในพจนานุกรมโดยอัตโนมัติด้วย แต่อย่างไรก็ตามถ้าข้อมูลที่ใช้ขนาดเท่ากัน จะเสียเวลาในการประมวลมากกว่าด้วยต้องพิจารณาทุกๆตำแหน่งของอักษร อีกยังต้องการเรียนรู้ที่มากครั้งกว่าเพื่อความถูกต้องในการรู้จำจะได้สูงมากขึ้น

2.ป้อนข้อมูลด้วยอักษรคัดลายมือ

เมื่อศึกษาเปรียบเทียบระหว่างผลงานหมายเลข 2 กับผลงานหมายเลข 3 แล้วจะมีการพิจารณาที่ค่อนข้างยาก ด้วยผลงานแต่ละวิธีนั้นมีเทคนิควิธีการในการเรียนรู้ และรู้จำที่แตกต่างกันมาก โดยผลงานหมายเลข 2 มีวิธีการที่แน่นอนในการพิจารณาอักษรที่เข้ามาทดสอบ โดยจะต้องเป็น อักษรลายมือเขียนภาษาไทยที่มีหัวอักษร (ลักษณะวงปิดที่เกิดจากการเขียน อักษรภาษาไทยในลักษณะเขียนบรรจง) และพิจารณาโดยอาศัย วิธีการตั้งแต่การทำให้อักษรบาง (Thinning) การหาความสัมพันธ์ระหว่างจุดภาพ (จุดต่อ จุดตัด จุดแยก จุดปลาย จุดอิสระ) โดยอาศัย 4-Neighbourhood และ 8-Neighbourhood การหาจุดศูนย์ถ่วง (Center of Gravity) การตรวจสอบขอบของภาพ (Edge Detection) เพื่อหาจำนวนและตำแหน่งของหัว และการปรับรูปแบบอักษรให้ได้ขนาดที่กำหนด (Standard Size) เป็นขั้นตอนในการพิจารณาหาลักษณะเฉพาะของอักษรแต่ละตัว เป็นโปรแกรมที่มีวิธีการรู้จำที่แน่นอน ซึ่งต่างจากผลงานหมายเลข 3 อย่างมาก จึงเปรียบเทียบกันค่อนข้างยาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ลักษณะที่พิจารณา	1	2	3
ความเร็ว	*	1.35	1.1
ความถูกต้อง	*	99%	98%
เงื่อนไขเฉพาะ	อักษรตัวพิมพ์	-ไทย-	-

*** หมายเหตุ ***

1.การจดจำรูปแบบตัวอักษรตัวพิมพ์ภาษาไทย [27]

2.การรู้จำอักษรลายมือเขียนภาษาไทยโดยการพิจารณาหัวของตัวอักษร [29]

3.งานวิจัยที่ทำการศึกษา

* ไม่ได้พิจารณาด้วยติดเงื่อนไขของตัวอักษร

ตารางที่ 7-2 เปรียบเทียบในลักษณะของอักษรคัดลายมือ

ประสิทธิภาพของโปรแกรม " ออกแบบพจนานุกรมอัตโนมัติสำหรับอักษรคัดลายมือภาษาไทยและภาษาอังกฤษ โดยได้ทำการทดสอบเพื่อหาประสิทธิภาพเฉพาะโปรแกรมด้วยเทคนิคของการจัดสร้างพจนานุกรมกับขนาดของข้อมูลที่ใช้ทดสอบ สามารถแสดงเปรียบเทียบการทดสอบใช้งานดังในตารางที่ 7-3

ขนาดแฟ้มข้อมูลอักษรคัดลายมือต่ออักษร 10ตัว	RESOLUTION ของการเก็บข้อมูลภาพอักษร	เวลาในการทำพจนานุกรม (DICTIONARY)	เวลาในการทดสอบพจนานุกรม	
			เรียงตามASCII	สฤษฎ์ บัตตะโชติ
1,125 byte	75 DPI	6 sec	0.8 sec/ch	0.8 sec/ch
1,710 byte	100 DPI	8 sec	1.0 sec/ch	0.9 sec/ch
3,420 byte	150 DPI	10 sec	1.2 sec/ch	1.1 sec/ch

ตารางที่ 7-3 แสดงการเปรียบเทียบด้วยเงื่อนไขการสร้างพจนานุกรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และสามารถทดสอบประสิทธิภาพของการรู้จำอักษรคัดลายมือ จากจำนวนครั้งของการเรียนรู้อักษรแต่ละตัว (ซึ่งเป็นความไม่คงที่ของอักษรคัดลายมือที่ใช้) กับความสามารถในการรู้จำอักษร ซึ่งจะปรากฏผลการรู้จำดังแสดงในตารางที่ 7-4

จำนวนอักษรต้นแบบที่ใช้ทำพจนานุกรม	การทดสอบอักษรชุดเดิม		การทดสอบอักษรชุดใหม่	
	รู้จำอักษร	ไม่รู้จำอักษร	รู้จำอักษร	ไม่รู้จำอักษร
10 char	100 %	-	2 %	98 %
50 char	100 %	-	95 %	5 %
100 char	100 %	-	98 %	2 %

ตารางที่ 7-4 แสดงประสิทธิภาพของการรู้จำกับวิธีการรู้จำที่ใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 8

บทสรุปและข้อเสนอแนะ

จากการศึกษาและทดลอง โปรแกรมในส่วนของการออกแบบจนานุกรมอัตโนมัติสำหรับการรู้จำอักษรคัดลายมือภาษาไทยและภาษาอังกฤษ ด้วยการพิจารณาลักษณะโครงสร้างของอักษรซึ่งแบ่งออกเป็น 3 ส่วนคือการหารหัสเบื้องต้น การปรับรหัสพื้น และการสร้างรหัสรวม แล้วจึงพิจารณาเฉพาะรหัสรวมที่มีความถี่สูงสุด 5 อันดับแรกเป็นลักษณะเด่น เฉพาะของอักษรตัวนั้นเก็บไว้เป็นฐานความรู้ของการรู้จำตัวอักษรจากอักษรภาษาไทยจำนวน 85 ตัวๆละ 100 ครั้ง (พยัญชนะ 44 ตัว สระ 27 ตัว วรรณยุกต์ 4 ตัว เลขไทย 10 ตัว) และอักษรภาษาอังกฤษจำนวน 62 ตัวๆละ 100 ครั้ง (ตัวพิมพ์ใหญ่ 26 ตัว ตัวพิมพ์เล็ก 26 ตัว เลขอารบิค 10 ตัว) โดยเขียนตัวอักษรขนาด 0.5-1.0 ซม.และเก็บภาพด้วยความละเอียด 75-150 dpi (จุดภาพต่อหน่วยนิ้ว) หรือใช้อักษรขนาด 400-600 byteต่อตัวอักษร ปรากฏว่าได้ผลการรู้จำ เวลาในการเรียนรู้(สร้างฐานความรู้ของการรู้จำ) เฉลี่ยประมาณ 1 วินาทีต่อตัวอักษร และเวลาที่ใช้ในการทดสอบการรู้จำตัวอักษรเฉลี่ยประมาณ 1.2 วินาทีต่อตัวอักษร โดยได้ผลการรู้จำถูกต้องประมาณ 98% ต้องเรียนรู้ใหม่อีก 2% บนเครื่องไมโครคอมพิวเตอร์ AT/XT ขนาด 16 bit ความเร็วของการประมวลผล 12 Mbytesต่อวินาที

<u>ข้อเสนอแนะ</u>

จากการศึกษาและทดลอง พบว่ายังสามารถที่จะพัฒนาให้โปรแกรมมีประสิทธิภาพขึ้นได้

1. ฐานความรู้ของการรู้จำตัวอักษร จะต้องมีการเรียนรู้จากอักษรคัดลายมือจำนวนมากๆ เพื่อให้การใช้งานโปรแกรมมีประสิทธิภาพ สามารถรู้จำอักษรได้จำนวนหลายหลายรูปแบบ ซึ่งสามารถที่จะแยกฐานความรู้ของการรู้จำออกเป็น บุคคล กลุ่มบุคคล แยกภาษาไทยและอังกฤษ เพื่อให้ข้อมูลมีจำนวนน้อยลง มีความเร็วในการประมวลผลมากขึ้น

2. การป้อนข้อมูล ลักษณะของการป้อนข้อมูลที่ต้องมีแบบฟอร์มกระดาษเขียนอักษรคัดลายมือยังเป็นข้อจำกัดด้วยต้องอาศัยการแยกตัวอักษรออกจากคำหรือประโยค ซึ่งสามารถพัฒนาได้โดยการเปลี่ยนวิธีการในการแยกอักษรออกจากคำ ก็จะสามารถลดข้อจำกัดในการป้อนข้อมูลได้อีกส่วนหนึ่ง ซึ่งวิธีการในการแยกแยะอักษรนั้นสามารถหาศึกษาได้จากหนังสือ เอกสารวิชาการที่เกี่ยวข้องกับเทคนิคในการรู้จำและแยกแยะตัวอักษรออกจากคำหรือประโยคได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. ประสิทธิภาพในการประมวลผล สามารถเพิ่มประสิทธิภาพของโปรแกรมด้วยการประมวลผลด้วยเครื่องไมโครคอมพิวเตอร์ที่มีความเร็วในการประมวลผลที่สูงขึ้น มีหน่วยความจำสำรองมากขึ้น ซึ่งจะทำให้การประมวลผลมีประสิทธิภาพและตอบสนองการใช้งานได้

4. ในการพิจารณาโครงสร้างของอักษรจากการพิจารณารหัสรวมที่มีความถี่สูงสุด 5 อันดับแรกเป็นลักษณะเด่นนั้น เป็นการหาลักษณะเด่นที่มองโครงสร้างโดยรวม พิจารณาเฉพาะบริเวณที่มีรหัสรวมอยู่เป็นส่วนใหญ่ จึงเหมาะสำหรับการรู้จำอักษรคัดลายมือที่มีลักษณะโครงสร้างที่คล้ายคลึงกัน หรือมีแนวโน้มไปทางเดียวกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

ขอขอบพระคุณ รองศาสตราจารย์ ดร.ชม กิ้มปาน
ได้ให้ความอนุเคราะห์ ให้คำปรึกษาตลอดจนแนะนำการจัดทำวิทยานิพนธ์ฉบับนี้
จนกระทั่งสำเร็จลุล่วงด้วยดี

ขอขอบพระคุณ มูลนิธิเพื่อการศึกษาคอมพิวเตอร์และการสื่อสาร
ได้ให้ความอนุเคราะห์ ให้ทุนอุดหนุนในการศึกษาและวิจัยฯ

ขอขอบพระคุณ คณาจารย์ เจ้าหน้าที่ภาควิชาวิศวกรรมคอมพิวเตอร์

ขอขอบใจเพื่อน-น้อง

และ " ตุ้ม "

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หนังสืออ้างอิง

1. ผศ.ชม กิ้มปาน,"ทฤษฎีการจดจำรูปแบบเบื้องต้น (Pattern Recognition Principles), ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าวิทยาเขตเจ้าคุณทหารลาดกระบัง กรุงเทพฯ, พิมพ์ครั้งที่ 1, กรกฎาคม 2525
2. YASUHIRO YAMADA ,TAKAHIGO KAWATANI ,HIROSHI KANEKO and TOSHIO TSUTSUMIDA ,"Handprinted Numeral Recognition by Feature Concentration Method" , Review of the electrical communication laboratories , vol.26 , No.11-12 , November-December 1978
3. HIRANVANICHAKORN,P., AGUI,T.,ana NAKAJIMA,M.:"A Recognition Method of Thai Characters". The Transactions of the IECE of Japan,vol.E65,no.12,December 1982.
4. KIMPAN,C.,A.LTOH and PROF.K.KAWANISHI,"Recognition of Printed Thai Characters Using a Matching Methods",IEE Procedding ,vol.130 ,Pt.E,No.6,November 1983.
5. HIRANVANICHAKORN,P., AGUI,T.,ana NAKAJIMA,M.:"Recognition of Thai Characters by Using Local Features".The Transaction of the IECE of Japan,vol.E67,no.8,August 1984.
6. KIMPAN,C.,"Printed Thai Characters Recognition Using Topological Properties Method", INT.J.Electronics, 1986,vol.60,No.3,303-329
7. KIMPAN,C.,:"Printed Thai Characters Recognition",Dissertation Of Doctor of Engineering, King Mongkut's Institute of Technology Chaokhun Taharn Ladkrabang,1986.
8. KIMPAN,C. , A.LTOH and PROF.K.KAWANISHI , "Fine Classification of Printed Thai Character Recognition Using the Karhunen-Loeve Expansion",IEE Proceeding , vol.134 ,Pt.E, No.5, September 1987

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

9. SOMSAK WALAIRACHT , CHAI KASEMAMORNKUL AND DR.CHOM KIMPAN , "Handprinted Thai Character Recognition by Feature Concentrate Method" , COMPUTER PROCESSING OF ASIAN LANGUAGE (CPAL) , ASIAN INSTITUTE OF TECHNOLOGY , September 26-28 , 1989

10. สุรสิทธ์ ราตรี ,รศ.ดร.ชม กิ้มปาน ,"การจดจำรูปแบบตัวพิมพ์อักษรภาษาไทยโดยการวิเคราะห์โครงสร้างแบบมีไวยากรณ์", การประชุมวิชาการทางวิศวกรรมไฟฟ้าครั้งที่ 10 ณ จุฬาลงกรณ์มหาวิทยาลัย, 24-25 พฤศจิกายน 2530

11. วัชระ ฉัตรวิริยะ, กิตติพัทธ์ เจียรพงษ์ และ ศจ.ดร.ศรีศักดิ์ จามรมาน, "การแยกแยะตัวพิมพ์ภาษาไทยออกจากภาพตัวอักษรของประโยค" , การประชุมวิชาการทางวิศวกรรมไฟฟ้าครั้งที่ 10 ณ จุฬาลงกรณ์มหาวิทยาลัย, 24-25 พฤศจิกายน 2530

12. อนุชิต จารุวนารัตน์ สุรสิทธ์ ราตรี และรศ.ดร.ชม กิ้มปาน, "การแยกภาพตัวอักษรภาษาไทยออกจากภาพของประโยค", การประชุมวิชาการวิศวกรรมไฟฟ้าครั้งที่ 10, คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย , 24-25 พฤษภาคม 2530

13. ศมิทธ์ เอมสมบัติ, สุรสิทธ์ ราตรี และ รศ.ดร.ชม กิ้มปาน," การจดจำรูปแบบอักษรตัวพิมพ์อักษรภาษาไทยด้วยการกระจายแบบคาร์ฮูเนน-โลบ", การประชุมทางวิชาการและนิทรรศการ การประยุกต์ใช้คอมพิวเตอร์ในทางวิศวกรรม ณ มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่ , 11-12 กุมภาพันธ์ 2531

14. สมศักดิ์ วลัยรัชต์ สุรสิทธ์ ราตรี และ รศ.ดร.ชม กิ้มปาน,"การจดจำรูปแบบอักษรตัวพิมพ์ภาษาไทย", การประชุมทางวิชาการและนิทรรศการ การประยุกต์ใช้คอมพิวเตอร์ในทางวิศวกรรม ณ มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่ , 11-12 กุมภาพันธ์ 2531

15. ประกาศิต ชาติบุรุษ ,ธนา หงษ์สุวรรณ ,วัชระ ฉัตรวิริยะ และ ดร.บุญวัฒน์ อัตชู, " การแยกแยะตัวพิมพ์ภาษาไทยทีละตัวออกจากคำ", การประชุมทางวิชาการและนิทรรศการ การประยุกต์ใช้คอมพิวเตอร์ในทางวิศวกรรม ณ มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่ , 11-12 กุมภาพันธ์ 2531

16. สุรพันธ์ เอื้อไพบูลย์, ดร.บุญวัฒน์ อัตชู และ รศ.มนัส สังวรศิลป์ ,"การจำแนกกลุ่มของตัวอักษรภาษาไทยโดยการพิจารณาหัวของตัวอักษร",บทความประชุมทางวิชาการ "การประยุกต์ใช้คอมพิวเตอร์ในทางวิศวกรรม",มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่ 11-12 กุมภาพันธ์ 2531

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

17. สมศักดิ์ มิตะถา ชาย เกษมอมรกุล และ ดร.ชม กิ้มปาน ,"การออกแบบพจนานุกรมอัตโนมัติสำหรับการจดจำรูปแบบตัวเลขคัดลายมือ" , การประชุมวิชาการวิศวกรรมไฟฟ้า ครั้งที่ 11 ณ คณะเทคโนโลยี สถาบันเทคโนโลยีราชมงคล วิทยาเขตเทเวศน์ , 16-17 ธันวาคม 2531

18. วัชระ ฉัตรวิริยะ และ ดร.ชม กิ้มปาน , "การสร้างพจนานุกรมโดยอัโนมัติสำหรับให้เครื่องไมโครคอมพิวเตอร์สามารถรู้จำอักษร (Automatic design dictionary for Characters Recognition in Microcomputer" , เอกสารประกอบการประชุมใหญ่ทางวิชาการประจำปี 2532 วิศวกรรมสถานแห่งประเทศไทย ในพระบรมราชูปถัมภ์ , กรุงเทพฯ , 26-28 ตุลาคม 2532

19. ชาย เกษมอมรกุล และ ดร.ชม กิ้มปาน , "การรู้จำอักษรคัดลายมือภาษาไทยสำหรับเครื่องไมโครคอมพิวเตอร์โดยการพิจารณาลักษณะเด่นของอักษร(Handprinted Thai Chracter of Microcomputer Using Concentrated Code of Each Character), การประชุมวิชาการวิศวกรรมไฟฟ้าครั้งที่ 12 ณ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์, 16-17 พฤศจิกายน 2532

20. พงษ์เทพ ศิริบวรเกียรติ กิจจา สรณียะภัทร์, "การวิเคราะห์อักษรไทยสำหรับเครื่องอ่านด้วยแสง", ปริญญานิพนธ์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง , ปีการศึกษา 2519

21. วีรพล ชฎาวัฒน์ วิชัย ตั้งชัยพิทักษ์ และ สุชิน สินสุวรรณ ,"การจดจำรูปแบบตัวเลขลายมือเขียนโดยใช้คุณสมบัติทางโทโปโลยี", ปริญญานิพนธ์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง , ปีการศึกษา 2520

22. สำราญ ศรีวัชระพงศ์ กวีลย์ ตันติวัสดากร ,"Pattern Recognition of Thai Number Pattern",ปริญญานิพนธ์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง , ปีการศึกษา 2520

23. วรสิษฐ์ สุวรรณทำ วิชัย ยิ่งยงสมสวัสดิ์,"Character Recognition",ปริญญานิพนธ์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง , ปีการศึกษา 2524

24. วสันต์ ศิริมาลีชัย สิทธิ ธเนศพิชิตชัย ,"Handwriting Thai Character Recognition" ,ปริญญานิพนธ์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง , ปีการศึกษา 2526

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

25. กำจร วิศลดิลก วิเชียร สุภานันท์ และ วิเชียร เสิศธนพงศ์ ,"การพัฒนาซอฟแวร์สำหรับเครื่องอ่านอักษรภาษาไทย",ปริญญานิพนธ์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง , ปีการศึกษา 2527

26. ธนา หงษ์สุวรรณ ประกาศิต ชาติบุรุษ ,"การประยุกต์การจดจำรูปแบบตัวอักษรคัดลายมือ" ,ปริญญานิพนธ์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้า เจ้าคุณทหารลาดกระบัง ,ปีการศึกษา 2530

27. มาฆกษ์ เกียรติก้อง และ ธนุษย์สิทธิ์ คล้ายมงคล , "การจดจำรูปแบบตัวอักษรตัวพิมพ์ภาษาไทย " , ปริญญานิพนธ์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง , ปีการศึกษา 2531

28. ประสาร ตังติสานนท์ , "การจดจำรูปแบบตัวอักษรคัดลายมือภาษาไทย โดยวิธีแยกลักษณะเด่น " ,วิทยานิพนธ์มหาบัณฑิต คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหาลาดกระบัง , ประจำปีการศึกษา 2529

29. สุรนันท์ เอื้อไพบูลย์ ,"การรู้จำอักษรลายมือเขียนภาษาไทยโดยการพิจารณาหัวของตัวอักษร (Recognition of Hand-written Thai Character Considering the Head of Characters)",วิทยานิพนธ์มหาบัณฑิต คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง , ประจำปีการศึกษา 2531

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก

รหัสแอสกี้ของอักขระใน รหัส สมอ.

- อักขระพิเศษ (1-15)

รหัสแอสกี้	อักขระ	รหัสแอสกี้	อักขระ	รหัสแอสกี้	อักขระ
33	!	34	"	35	#
36	$	37	%	38	&
39	'	40	(	41	)
42	*	43	+	44	,
45	-	46	.	47	/

- ตัวเลขอารบิค (2-10)

รหัสแอสกี้	อักขระ	รหัสแอสกี้	อักขระ	รหัสแอสกี้	อักขระ
48	0	49	1	50	2
51	3	52	4	53	5
54	6	55	7	56	8
57	9				

- อักขระพิเศษ (3-7)

รหัสแอสกี้	อักขระ	รหัสแอสกี้	อักขระ	รหัสแอสกี้	อักขระ
58	:	59	;	60	<
61	=	62	>	63	?
64	@				

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รหัสแอสกี้ของอักขระใน รหัส สมอ.

- อักษรภาษาอังกฤษ (4-26)

รหัสแอสกี้	อักขระ	รหัสแอสกี้	อักขระ	รหัสแอสกี้	อักขระ
65	A	66	B	67	C
68	D	69	E	70	F
71	G	72	H	73	I
74	J	75	K	76	L
77	M	78	N	79	O
80	P	81	Q	82	R
83	S	84	T	85	U
86	V	87	W	88	X
89	Y	90	Z		

- อักขระพิเศษ (5-6)

รหัสแอสกี้	อักขระ	รหัสแอสกี้	อักขระ	รหัสแอสกี้	อักขระ
91	[	92	\	93	]
94	^	95	_	96	`

- อักษรภาษาอังกฤษ (6-26)

รหัสแอสกี้	อักขระ	รหัสแอสกี้	อักขระ	รหัสแอสกี้	อักขระ
97	a	98	b	99	c
100	d	101	e	102	f
103	g	104	h	105	i

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รหัสแอสกี้ของอักขระใน รหัส สมอ.

รหัสแอสกี้	อักขระ	รหัสแอสกี้	อักขระ	รหัสแอสกี้	อักขระ
106	j	107	k	108	l
109	m	110	n	111	o
112	p	113	q	114	r
115	s	116	t	117	u
118	v	119	w	120	x
121	y	122	z		

- อักขระพิเศษ (7-3)

รหัสแอสกี้	อักขระ	รหัสแอสกี้	อักขระ	รหัสแอสกี้	อักขระ
123	{	124	|	125	}

- อักษรภาษาไทย (8-46)

รหัสแอสกี้	อักขระ	รหัสแอสกี้	อักขระ	รหัสแอสกี้	อักขระ
161	ก	162	ข	163	
164	ค	165		166	ฆ
167	ง	168	จ	169	ฉ
170	ช	171	ซ	172	ฌ
173	ญ	174	ฎ	175	ฏ
176	ฐ	177	ฑ	178	ฒ
179	ณ	180	ด	181	ต
182	ถ	183	ท	184	ธ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รหัสแอสกี้ของอักขระใน รหัส สมอ.

รหัสแอสกี้	อักขระ	รหัสแอสกี้	อักขระ	รหัสแอสกี้	อักขระ
185	น	186	บ	187	ป
188	ผ	189	ฝ	190	พ
191	ฟ	192	ภ	193	ม
194	ย	195	ร	196	ฤ
197	ล	198	ฦ	199	ว
200	ศ	201	ษ	202	ส
203	ห	204	ฬ	205	อ
206	ฮ				

- อักขระพิเศษ (9-23)

รหัสแอสกี้	อักขระ	รหัสแอสกี้	อักขระ	รหัสแอสกี้	อักขระ
207	ฯ	208	ะ	209	ั
210	า	211	ำ	212	ิ
213	ี	214	ึ	215	ื
216	ุ	217	ู		
224	เ	225	แ	226	โ
227	ใ	228	ไ		
230	ๆ	231	็	232	่
233	้	234	๊	235	๋
236	์				

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รหัสแอสกี้ของอักขระใน รหัส สมอ.

- ตัวเลขไทย (10-10)

รหัสแอสกี้	อักขระ	รหัสแอสกี้	อักขระ	รหัสแอสกี้	อักขระ
240	๐	241	๑	242	๒
243	๓	244	๔	245	๕
246	๖	247	๗	248	๘
249	๙				

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข

การใช้งานโปรแกรมเครื่องตรวจกวาดภาพ
(SCANGAL ... HEWLETT PACKARD SCANNER)

การใช้งานเครื่องตรวจกวาดภาพ (SCANNER HEWLETT PACKARD) จะต้องมีเครื่องมือ ประกอบด้วย 2 ส่วนใหญ่ๆคือ

1. ส่วน Hardware ประกอบด้วย
   - เครื่องตรวจกวาดภาพ (HEWLETT PACKARD SCANNER)
   - เครื่องไมโครคอมพิวเตอร์ ที่มีขนาดหน่วยความจำไม่น้อยกว่า 640 Kbyte Harddisk ขนาดไม่น้อยกว่า 10 Mbyte
2. ส่วน Software ประกอบด้วย
   - โปรแกรม Scanning Gallery Setup Disc
   - โปรแกรม Scanning Gallery Build Disc
   - โปรแกรม Scanning Gallery Utilities Disc
   - โปรแกรม Scanning Gallery Font Disc
   - โปรแกรม Scanning Gallery Application Disc 1
   - โปรแกรม Scanning Gallery Application Disc 2
   - โปรแกรม Scanjet Software Disc

ขั้นตอนการเตรียมโปรแกรมเพื่อใช้งาน

1. ต้องจัดการเตรียมชุดโปรแกรมของการตรวจกวาดภาพ บรรจุลงไปใน Harddisk (Install Program Scanner) ไว้ใน sub-directory ที่ชื่อว่า /SCANGAL โดยต้องกำหนดข้อกำหนดต่างๆของอุปกรณ์ (ลักษณะเครื่องไมโครคอมพิวเตอร์ ชนิดของจอภาพ(Monitor) แป้นพิมพ์(Key-board) เครื่องพิมพ์(Printer) เครื่องจุดภาพ(Plotter) อุปกรณ์เลื่อนตำแหน่ง(Mouse) ช่องทางการติดต่อ(Port)) ให้ตรงกับการใช้งานจริง
2. จัดการ Install Utility เพื่อจัดส่วนโปรแกรม SJDRIVER.SYS ให้กับเครื่องไมโครคอมพิวเตอร์ แล้วทำการ Reboot เครื่อง เตรียมการใช้งานโปรแกรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนการใช้งานโปรแกรม

1. ต้อง Boot เครื่องไมโครคอมพิวเตอร์โดยผ่านส่วนของ SJDRIVER.SYS เสมอ เป็นการจัดการเพื่อการเตรียมการใช้งานโปรแกรม SCANGAL
2. ใช้โปรแกรมโดยเข้าสู่ Sub-directory ของโปรแกรมแล้วเรียกชื่อโปรแกรม Scan ตัวอย่าง เช่น ... A> \SCANGAL\SCANGAL <ENTER>
   โปรแกรมจะแสดง ชื่อโปรแกรม ปี ค.ศ. บริษัทที่ผลิต ในรูปที่ ข-1 และเข้าสู่ส่วนของโปรแกรม ในรูปที่ ข-2

HEWLETT
PACKARD

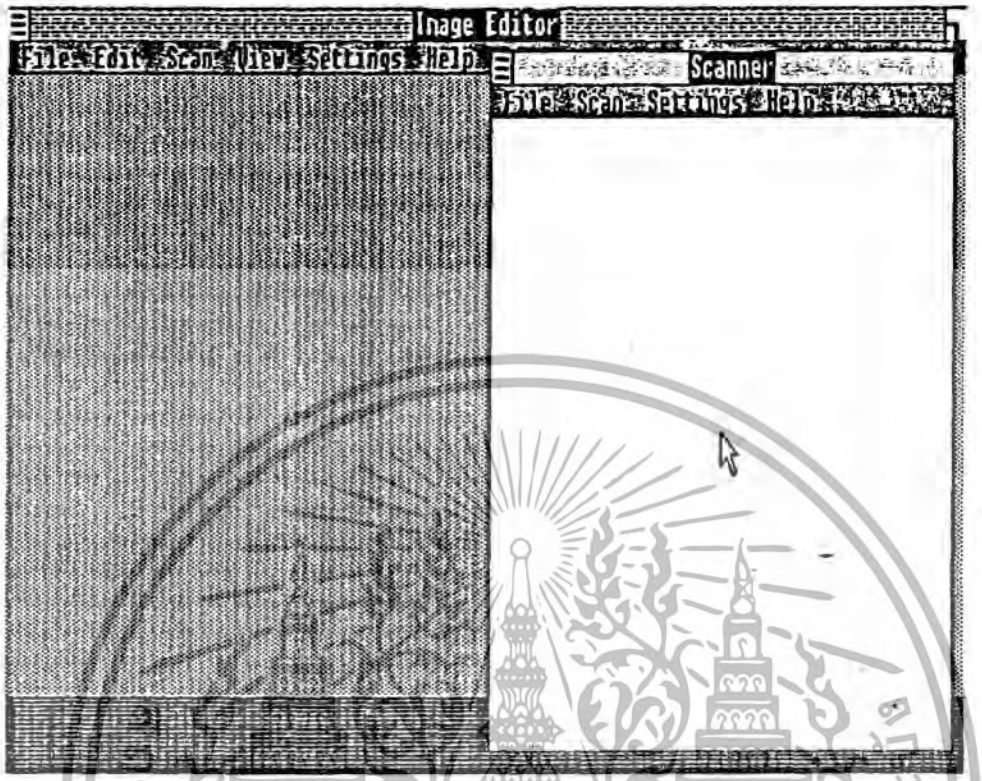
Microsoft(R) Windows
For the HP Vectra Personal Computer

A.01.03 2645

Copyright(c) 1986, Hewlett-Packard Co.
Copyright(c) 1985,1986, Microsoft Corp.

รูปที่ ข-1 แสดงชื่อโปรแกรม ปี ค.ศ. บริษัทที่ผลิต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ ข-2 แสดง window ของการ edit และ scan

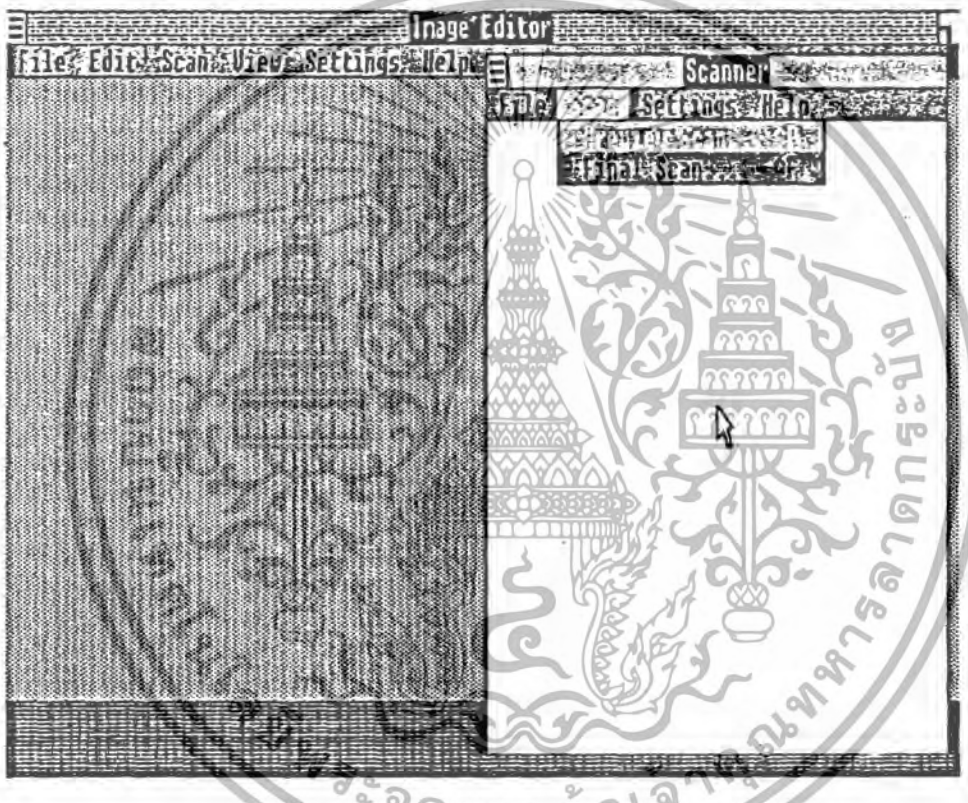
การออกแบบพจนานุกรม
การรู้จำอักษรคัดลายมือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ ข-3 ตัวอย่างลายมือเขียนที่ใช้

3. วางกระดาษที่มีข้อความ ตัวอย่างลายมือเขียนในรูปที่ ข-3 ที่ต้องการ scan ที่เครื่อง HP.Scanner โดยวางให้มุมบนขวาของข้อความอยู่ใกล้กับหัวลูกศรสีส้มที่เครื่อง Scanner แล้ววางแผ่นฝาปิดทับกระดาษ

4. Preview scan โดยกด Ctrl-P หรือใช้ Alt-S แล้วเลือกมาที่ Preview scan กด <ENTER> จะได้ดังรูปที่ ข-4 ต้องทำการกำหนดเงื่อนไขของการ scan

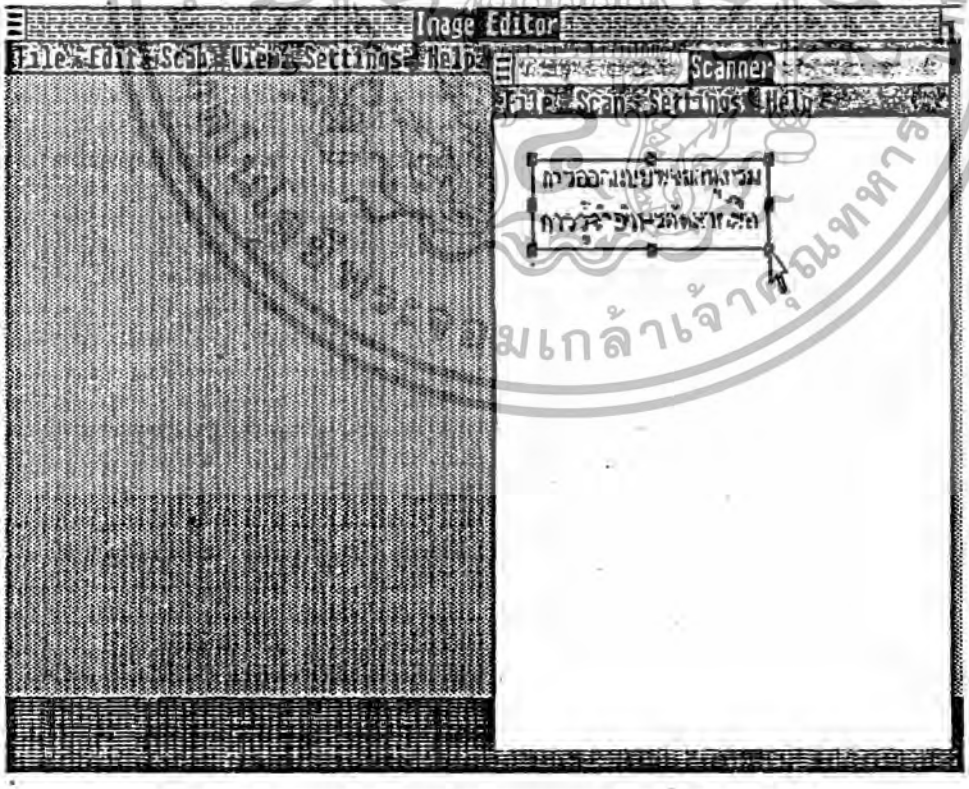


รูปที่ ข-4 แสดงการทำ Previrw scan

เครื่องจะถามเงื่อนไขในการทำ Preview scan ในรูปที่ ข-5 เมื่อตกลงเครื่อง scanner จะทำการ scan และนำข้อมูลทั้งหมดมาแสดงบนจอภาพในส่วน scan window หลังจากนั้นจะทำการเก็บเฉพาะส่วนข้อมูลโดยการใช้ Ctrl กับ ลูกศร ทั้ง 4 ทิศทาง แล้ววาดขอบเขตที่ต้องการเก็บ ในรูปที่ ข-6

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

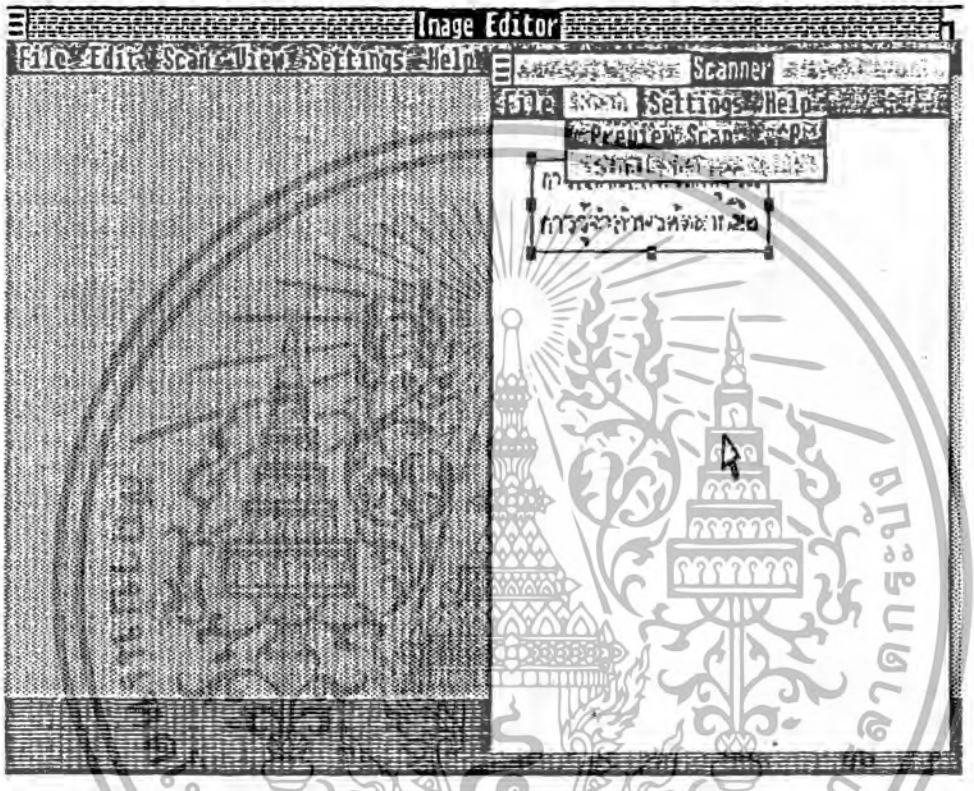
รูปที่ ข-5 การกำหนดเงื่อนไขของ Preview scan



รูปที่ ข-6 แสดงการกำหนดขอบเขตข้อมูลที่ต้องการเก็บ

เอกสารนี้เป็นเอกสาร[illegible]ให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

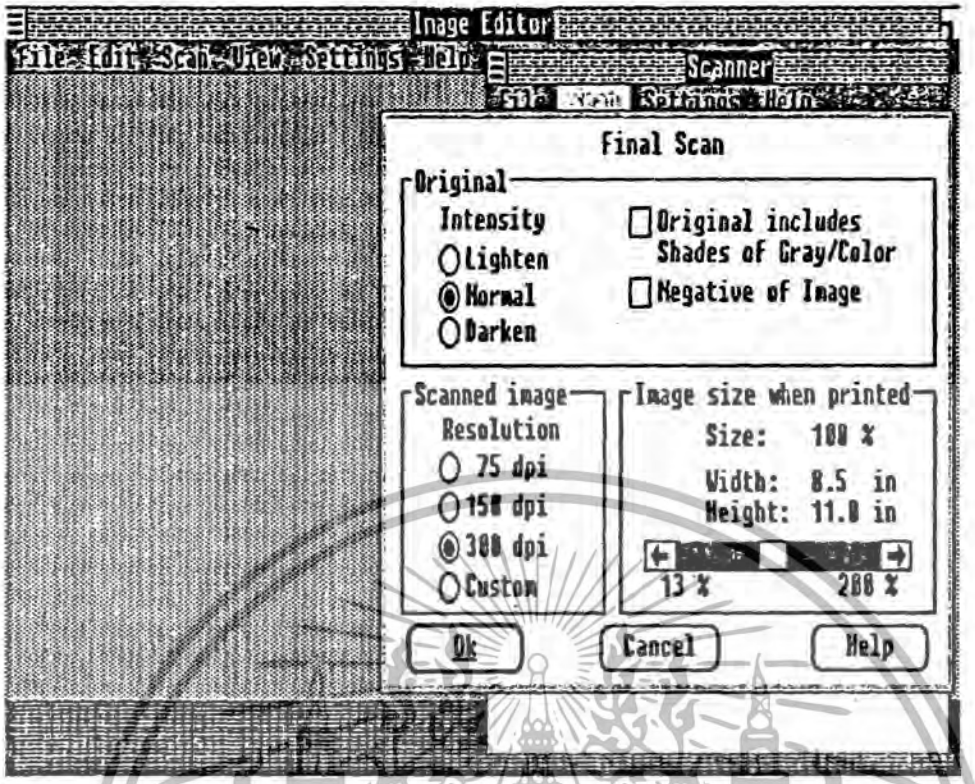
5. Final Scan เป็นส่วนของการ scan เก็บข้อมูลลง file หลังจากกำหนดขอบเขตข้อมูลแล้ว กด Ctrl-F หรือ Alt-S แล้วเลือกมาที่ Final Scan ในรูปที่ ข-7



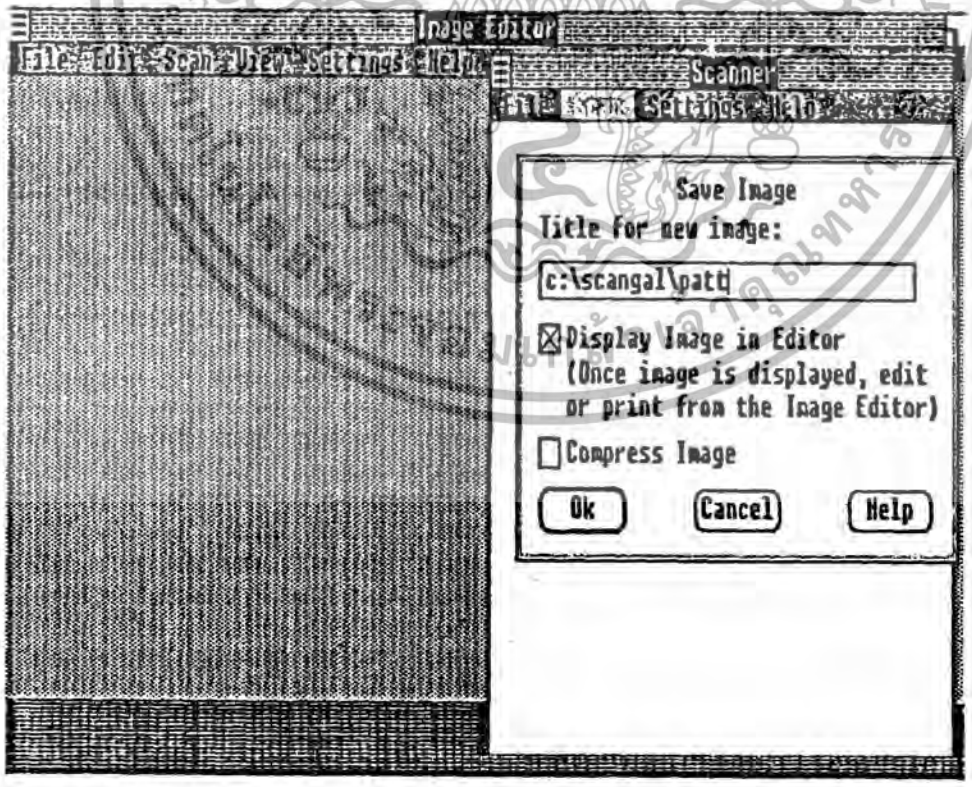
รูปที่ ข-7 แสดงการทำ Final Scan

แล้วเครื่องจะถามเงื่อนไขในการ Final Scan ในรูปที่ ข-8 แล้วจึงจัดเก็บข้อมูลลง file ดังในรูปที่ ข-9 โดยการเปลี่ยนชื่อข้อมูลแล้วจัดเก็บ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



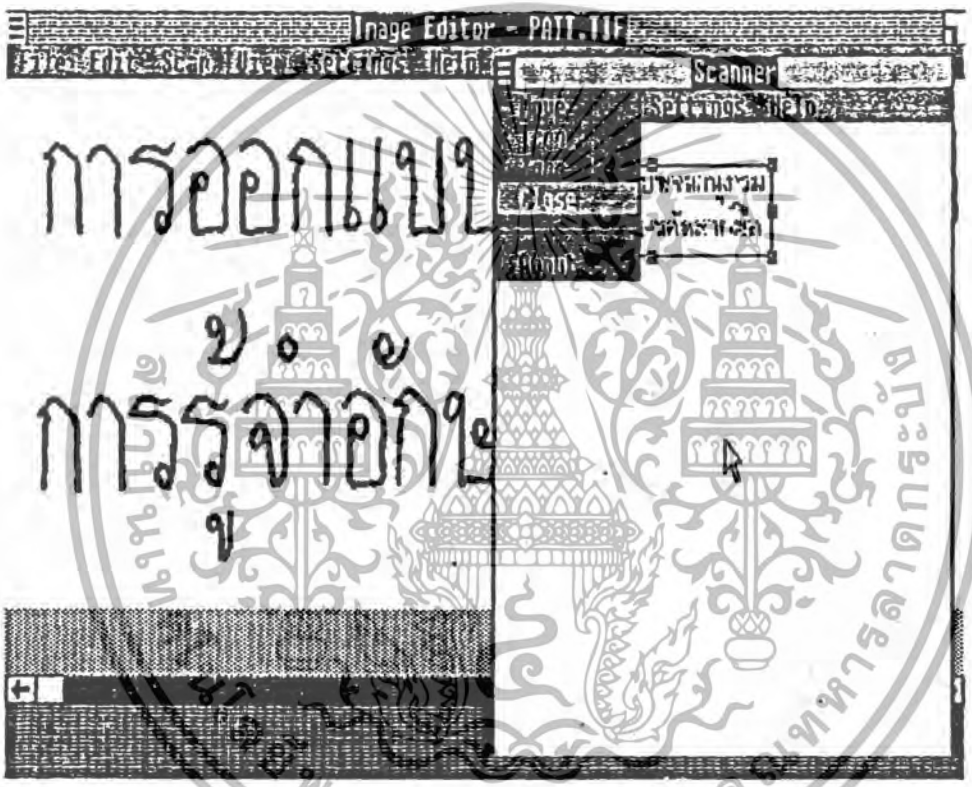
รูปที่ ข-8 แสดงการกำหนดเงื่อนไขของ Final Scan



รูปที่ ข-9 แสดงการเก็บข้อมูลลง file

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6. การออกจากส่วนของการ scan window โดยใช้ Alt-F แล้วเลื่อนไปทางซ้ายมือ กด C จะได้ดังรูปที่ ข-10 เมื่อกด <ENTER> ก็จะแสดงเฉพาะในส่วนของ edit window ในรูปที่ ข-11



รูปที่ ข-10 แสดงการออกจาก scanner window

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ ข-11 แสดง edit window

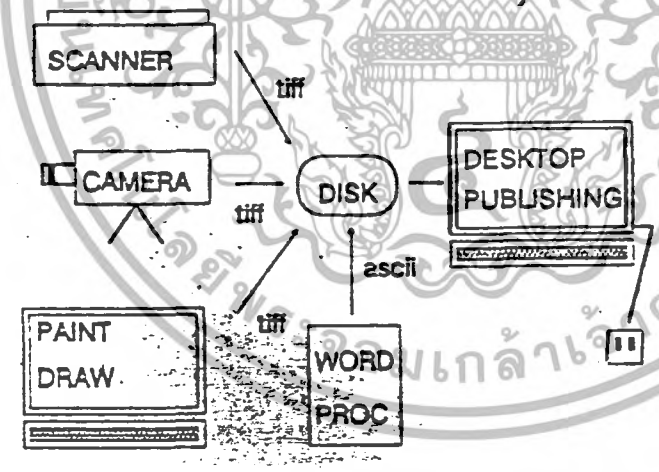
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ค

รูปแบบของ TIFF File

(TIFF - TAG IMAGE FORMAT FILE)

ในการเก็บข้อมูลของภาพด้วยเครื่องตรวจกวาดภาพ (Image Scanner) รูปแบบของไฟล์ข้อมูลภาพจะอยู่ในรูปแบบของ TIFF File ซึ่งเป็นรูปแบบใหม่ของการเก็บข้อมูลภาพ โดยมีลักษณะของรูปแบบไฟล์ ในรูปของการผสมผสาน 2 อย่างคือ มีส่วนรูปแบบของการบอกเงื่อนไขของข้อมูลภาพ ซึ่งบอกรายละเอียดของภาพนั้นๆว่ามีลักษณะเป็นอย่างไร และส่วนของขนาดของเงื่อนไขที่ได้ระบุไว้ในส่วนแรก ซึ่งแตกต่างจากลักษณะข้อมูลภาพทั่วไป โดยจะระบุตำแหน่งของขนาดข้อมูลที่แสดงเงื่อนไขของภาพในตำแหน่งที่แน่นอนในไฟล์ข้อมูลภาพ รูปแบบของข้อมูลแบบ TIFF นั้นจะสอดคล้องกับอุปกรณ์ประเภท เครื่องตรวจกวาดภาพ (Scanner) กล้องถ่ายภาพนิ่ง (Camera) และ ในส่วนโปรแกรมตระกูล Paint (Paint Program) ดังแสดงในรูป ค-1



รูปที่ ค-1 แสดงการประยุกต์ใช้งาน TIFF File

โครงสร้างทั่วไปของ TIFF File

เพื่อการนำข้อมูลภาพในรูปแบบของ TIFF File ไปใช้งาน เราสามารถแยกแยะโครงสร้างของข้อมูลภาพแบบ TIFF ได้ออกเป็น 4 ส่วนด้วยกันคือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

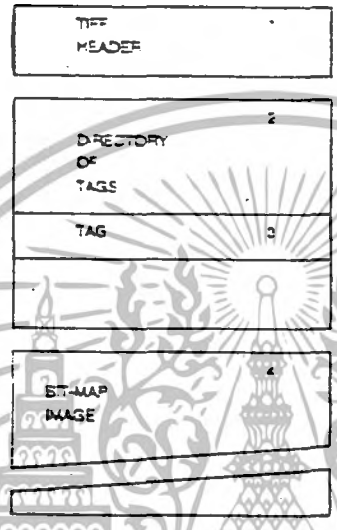
1. ส่วนหัวเรื่องของข้อมูลภาพ (Details of the Header Section)

2. ส่วนเงื่อนไขข้อมูลภาพ (Details of the Directory)

3. ส่วนแสดงขนาดของเงื่อนไขข้อมูลภาพ (Organization of Eack Tag Entry)

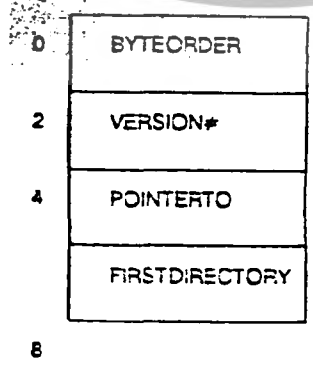
4. ส่วนข้อมูลภาพ (Bit-Map Image)

ซึ่งสามารถเขียนแสดงเป็นโครงสร้างดังรูป ค-2



รูปที่ ค-2 แสดงโครงสร้างของ TIFF File

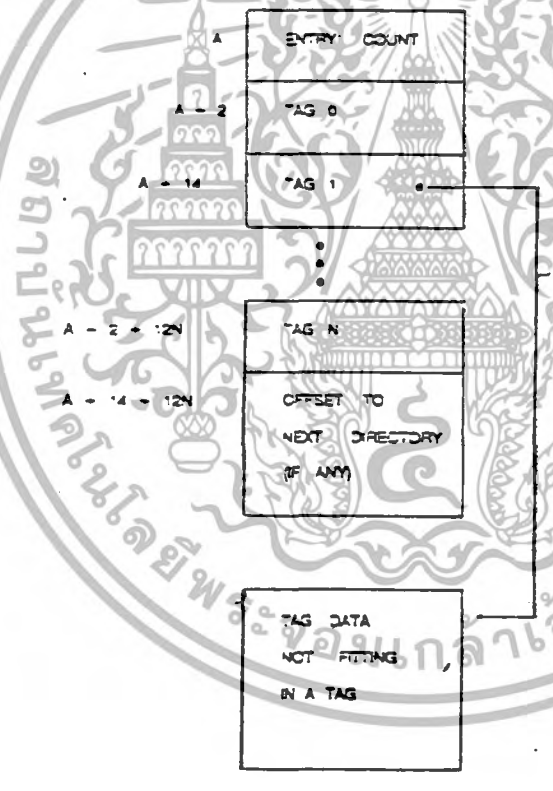
1. ส่วนหัวเรื่องของข้อมูลภาพ เป็นส่วนแรกที่ปรากฏอยู่ใน TIFF File และเป็นโครงสร้างที่สำคัญ ซึ่งจะบอกรายละเอียดของข้อมูลเบื้องต้นในลักษณะของลำดับสองไบท์ โดยแสดงรายละเอียดของรูปแบบการเก็บข้อมูล (BYTE ORDER) ตัวเลขแสดงรุ่นของข้อมูล TIFF Version (VERSION#) และข้อมูลแสดงตำแหน่งแรกของการแสดงรายละเอียดของข้อมูลเข้า (POINTERTO) ดังแสดงในรูปที่ ค-3



รูปที่ ค-3 แสดงส่วนหัวเรื่องข้อมูลภาพ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. ส่วนเงื่อนไขข้อมูลภาพ เป็นส่วนของข้อมูลที่แสดงหัวข้อรายละเอียดของข้อมูล สามารถแบ่งออกได้เป็น 3 ส่วนด้วยกันคือ ส่วนที่หนึ่งบอกด้วยสองไบท์แรกของส่วน Directory บอกรายละเอียดเกี่ยวกับจำนวนของ Tag (รายละเอียดหลักของข้อมูลภาพ) ส่วนที่สองเป็นส่วนที่แสดงรายละเอียดของข้อมูลในแต่ละ Tag ซึ่งประกอบด้วยข้อมูลแสดงรายละเอียดของแต่ละ Tag จำนวน 12 byte เป็นส่วนของการบอกรายละเอียดข้อมูลของภาพ เช่น ความกว้างของภาพ ความยาวของภาพ เทคนิคการเก็บภาพ ลักษณะการเก็บข้อมูลภาพ เป็นต้น ตามจำนวนของ Tag ซึ่งในแต่ละ Tag จะบอกรายละเอียดของภาพหนึ่งอย่างตามรูปแบบการเก็บของข้อมูลภาพ และส่วนที่สามเป็นส่วนท้ายที่แสดงค่า offset ซึ่งจะชี้ต่อไปยังส่วนของรายละเอียดอื่นๆ ซึ่งถ้าไม่มีส่วนของรายละเอียดถัดไป ก็จะแสดงค่าเป็นศูนย์ ดังแสดงในรูปที่ ค-4

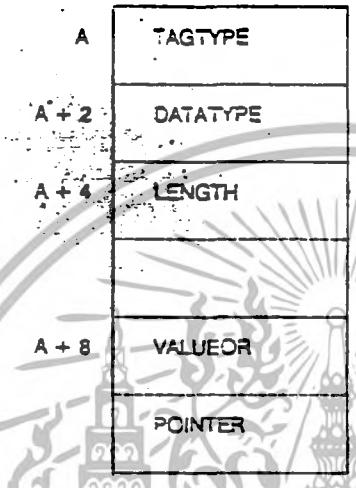


รูปที่ ค-4 แสดงส่วนเงื่อนไขข้อมูลภาพ

3. ส่วนแสดงขนาดของเงื่อนไขข้อมูลภาพ องค์ประกอบโครงสร้างของแฟ้มข้อมูลในส่วนของแต่ละ Tag จะประกอบด้วยข้อมูลยาว 12 byte สองไบท์แรกแสดง Tag ตำแหน่งอ้างอิงของข้อมูลหลัก (TAGTYPE) ส่วนสองไบท์ถัดไป(ไบท์ที่ 3 และ 4) แสดงชนิดของข่าวสาร (DATATYPE) เช่นถ้าสองไบท์นั้นมีค่าเป็น 282 (11AH) จะบอกรายละเอียดของความละเอียด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการสแกนตามแกน X ของข้อมูลภาพ ส่วนอีกสี่ไบท์ถัดไป(ไบท์ที่ 5,6,7 และ 8) จะแสดงค่าความยาวของข้อมูล (LENGTH) ส่วนอีกสี่ไบท์ถัดไป(ไบท์ที่ 9,10,11 และ 12) จะแสดงค่าของข้อมูล (VALUE/OFFSET) เช่น ถ้าในส่วนของ DATATYPE มีค่าเป็น 100H และส่วนของ VALUE มีค่าเป็น 0AH ก็จะได้ขนาดความกว้างของข้อมูลภาพมีขนาด 10 จุดภาพ เป็นต้น ดังแสดงในรูปที่ ค-5



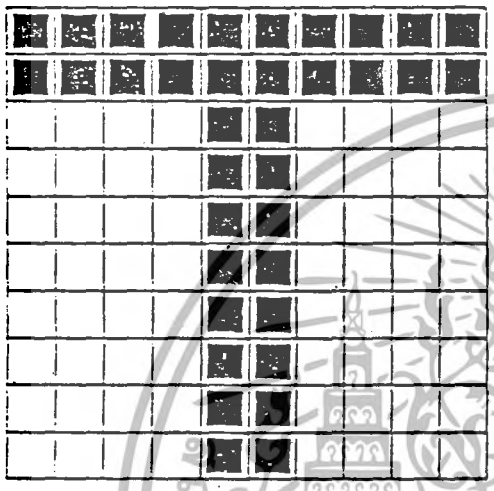
รูปที่ ค-5 แสดงส่วนแสดงขนาดของเงื่อนไขข้อมูลภาพ

4. ส่วนข้อมูลภาพ เป็นส่วนเฉพาะข้อมูลที่มีการจัดเก็บตามลักษณะที่ได้กำหนดไว้ในรูปของข้อมูลแบบ binary ตามในส่วนของ Directory Section ที่ได้กล่าวแล้ว ซึ่งจากโครงสร้างที่ได้กล่าวแล้วนั้น จะทำให้เข้าใจโครงสร้างลักษณะของข้อมูลต่างๆ การจัดเก็บตลอดจนถึงการอ่านข้อมูลออกมาใช้งานอย่างถูกต้อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่างการแจกแจงข้อมูลภาพแบบ TIFF File

An 8086-format 10 pixel x 10 pixel Black and White TIFF Image



300 DPI Image
BINARY (Line art) image
No data compression
8086 byte order

	Tag #	Tag Name	Value
__x__	0FFH	Subfile Type	Value:__1___
__x__	100H	Image Width	Value:__10__
__x__	101H	Image Length	Value:__10__
__x__	102H	Bits per sample	Value:__1___
__x__	103H	Data compression	Value:__1___
__x__	106H	Photometric Interp.	Value:__1___
_____	107H	Thresholding	Value:_____
_____	108H	Cell Width	Value:_____
_____	109H	Cell Length	Value:_____
__x__	10AH	Fill Order	Value:__1___

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

_____	10DH	Document name	Value:_____
_____	10EH	Image description	Ascii:_____
_____	10FH	Scanner make	Ascii:_____
_____	110H	Scanner model	Ascii:_____
__x__	111H	Strip Offsets	Value/Offset:Offset
__x__	112H	Orientation	Value:__1___
__x__	115H	Samples per pixel	Value:__1___
_____✓	116H	RowsPerStrip	Value:_____
_____✓	117H	StripByteCounts	Value/Offset:_____
__x__	118H	Min Sample Value	Value:__0___
__x__	119H	Max Sample Value	Value:__1___
__x__	11AH	X resolution	Value:_300__
__x__	11BH	Y resolution	Value:_300__
__x__	11CH	Planar configuration	Value:___1__
_____	11DH	Page Name	Ascii:_____
_____	11EH	X position	Value:_____
_____	11FH	Y position	Value:_____
_____	120H	FreeOffsets	Value/Offset:_____
_____	121H	FreeByteCounts	Value/Offset:_____
_____	122H	GrayResponseUnit	Value:_____
_____	123H	GrayResponseCurve	Value/Offset:_____
_____	124H	Group3Options	Value:_____
_____	125H	Group4Options	Value:_____
_____	128H	ResolutionUnit	Value:_____
_____	129H	PageNumber	Value:_____
_____	12CH	ColorResponseUnit	Value:_____
_____	12DH	ColorResponseCurves	Value/Offset:_____

File worksheet

I. Header

Offset	Value	
000	49	Note byte order is Intel format.
001	49	
002	2A	Version #42, Note byte reversal
003	00	
004	08	Pointer to first directory
005	00	
006	00	
007	00	

II. Directory

008	0F	15 tags in this directory
009	00	

	Offset	Tag	Type	Length	Value/Offset
Tag 1	00A	FF 00	03 00	01 00 00 00	01 00 00 00 (Subfile type)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Tag 2	016	00 01	03 00	01 00 00 00	0A 00 00 00	(Image width 10 pixels)
Tag 3	022	01 01	03 00	01 00 00 00	0A 00 00 00	(Image length 10 pixels)
Tag 4	02E	02 01	03 00	01 00 00 00	01 00 00 00	(1 Bit per sample)
Tag 5	03A	03 01	03 00	01 00 00 00	01 00 00 00	(Packed data compression)
Tag 6	046	06 01	03 00	01 00 00 00	00 00 00 00	(Black is 1)
Tag 7	052	0A 01	03 00	01 00 00 00	01 00 00 00	(Filled MSB to LSB)
Tag 8	05E	11 01	04 00	01 00 00 00	D2 00 00 00	(Offset to raster data)
Tag 9	06A	12 01	03 00	01 00 00 00	01 00 00 00	(Normal orientation)
Tag 10	076	15 01	03 00	01 00 00 00	01 00 00 00	(1 sample/pixel)
Tag 11	082	18 01	03 00	01 00 00 00	00 00 00 00	(Min value is 0)
Tag 12	08E	19 01	03 00	01 00 00 00	01 00 00 00	(Max value is 1)
Tag 13	09A	1A 01	05 00	01 00 00 00	C2 00 00 00	(Offset to X resolution)
Tag 14	0A6	1B 01	05 00	01 00 00 00	CA 00 00 00	(Offset to Y resolution)
Tag 15	0B2	1C 01	03 00	01 00 00 00	01 00 00 00	(Planar config is normal)

End of directory

BE 00 00 00 00 (no more directories)

III. Extended tag information (as needed)

C2	2C 01 00 00	01 00 00 00 (300 dpi X resolution)
CA	2C 01 00 00	01 00 00 00 (300 dpi Y resolution)

IV. Raster data (as needed)

D2	FFC0	(10 black pixels)
D4	FFC0	(10 black pixels)
D6	0C00	(2 black pixels)
D8	0C00	(2 black pixels)
DA	0C00	(2 black pixels)
DC	0C00	(2 black pixels)
DE	0C00	(2 black pixels)
E0	0C00	(2 black pixels)
E2	0C00	(2 black pixels)
E4	0C00	(2 black pixels)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

V. Strip offset pointers (as needed)

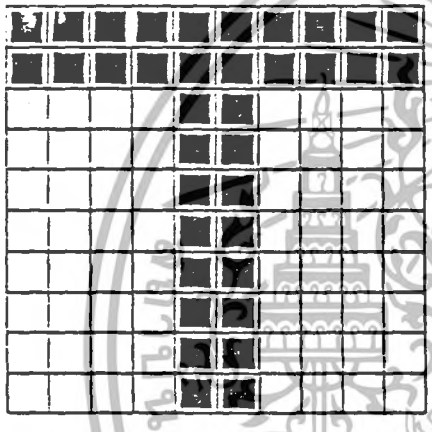
None needed.

VI. Strip byte counts (as needed)

None needed.

File is E6 bytes long.

A 68000- format 10 pixel x 10 pixel Black and White TIFF Image



300 DPI Image
BINARY (Line art) image
No data compression
68000 byte order

	Hex	Name	Value
__x__	0FFH	Subfile Type	Value:__1___
__x__	100H	Image Width	Value:__10__
__x__	101H	Image Length	Value:__10__
__x__	102H	Bits per sample	Value:__1___
__x__	103H	Data compression	Value:__1___
__x__	106H	Photometric Interp.	Value:__1___
_____	107H	Thresholding	Value:_____
_____	108H	Cell Width	Value:_____
_____	109H	Cell Length	Value:_____
__x__	10AH	Fill Order	Value:__1___
_____	10DH	Document name	Value:_____
_____	10EH	Image description	Ascii:_____

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

_____	10FH	Scanner make	Ascii:_____
_____	110H	Scanner model	Ascii:_____
__x__	111H	Strip Offsets	Value/Offset:Offset
__x__	112H	Orientation	Value:__1___
__x__	115H	Samples per pixel	Value:__1___
_____	116H	RowsPerStrip	Value:_____
_____	117H	StripByteCounts	Value/Offset:_____
__x__	118H	Min Sample Value	Value:__0___
__x__	119H	Max Sample Value	Value:__1___
__x__	11AH	X resolution	Value:_300__
__x__	11BH	Y resolution	Value:_300__
__x__	11CH	Planar configuration	Value:___1__
_____	11DH	Page Name	Ascii:_____
_____	11EH	X position	Value:_____
_____	11FH	Y position	Value:_____
_____	120H	FreeOffsets	Value/Offset:_____
_____	121H	FreeByteCounts	Value/Offset:_____
_____	122H	GrayResponseUnit	Value:_____
_____	123H	GrayResponseCurve	Value/Offset:_____
_____	124H	Group3Options	Value:_____
_____	125H	Group4Options	Value:_____
_____	128H	ResolutionUnit	Value:_____
_____	129H	PageNumber	Value:_____
_____	12CH	ColorResponseUnit	Value:_____
_____	12DH	ColorResponseCurves	Value/Offset:_____

File worksheet

I. Header

Offset	Value	
000	4D	Note byte order is Motorola format.
001	4D	
002	00	Version #42, Note byte order
003	2A	
004	00	Pointer to first directory
005	08	
006	00	
007	00	

II. Directory

008	00	15 tags in this directory
009	0F	

	Offset	Tag	Type	Length	Value/Offset	
Tag 1	00A	00 FF	00 03	00 00 00 01	00 01 00 00	(Subfile type)
Tag 2	016	01 00	00 03	00 00 00 01	00 0A 00 00	(Image width 10 pixels)
Tag 3	022	01 01	00 03	00 00 00 01	00 0A 00 00	(Image length 10 pixels)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Tag 4	02E	01 02	00 03	00 00 00 01	00 01 00 00	(1 Bit per sample)
Tag 5	03A	01 03	00 03	00 01 00 00	00 01 00 00	(Packed data compression)
Tag 6	046	01 06	00 03	00 00 00 01	00 00 00 00	(Black is 1)
Tag 7	052	01 0A	00 03	00 00 00 01	00 01 00 00	(Filled MSB to LSB)
Tag 8	05E	01 11	00 04	00 00 00 01	00 00 00 D2	(Offset to raster data)
Tag 9	06A	01 12	00 03	00 00 00 01	00 01 00 00	(Normal orientation)
Tag 10	076	01 15	00 03	00 00 00 01	00 01 00 00	(1 sample/pixel)
Tag 11	082	01 18	00 03	00 00 00 01	00 00 00 00	(Min value is 0)
Tag 12	08E	01 19	00 03	00 00 00 01	00 01 00 00	(Max value is 1)
Tag 13	09A	01 1A	00 05	00 00 00 01	00 00 00 C2	(Offset to X resolution)
Tag 14	0A6	01 1B	00 05	00 00 00 01	00 00 00 CA	(Offset to Y resolution)
Tag 15	0B2	01 1C	00 03	00 00 00 01	00 01 00 00	(Planar config is normal)

End of directory

BE 00 00 00 00 (no more directories)

III. Extended tag information (as needed)

C2	00 00 01 2C	00 00 00 01 (300 dpi X resolution)
CA	00 00 01 2C	00 00 00 01 (300 dpi Y resolution)

IV. Raster data (as needed)

D2	FFC0	(10 black pixels)
D4	FFC0	(10 black pixels)
D6	0C00	(2 black pixels)
D8	0C00	(2 black pixels)
DA	0C00	(2 black pixels)
DC	0C00	(2 black pixels)
DE	0C00	(2 black pixels)
E0	0C00	(2 black pixels)
E2	0C00	(2 black pixels)
E4	0C00	(2 black pixels)

V. Strip offset pointers (as needed)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

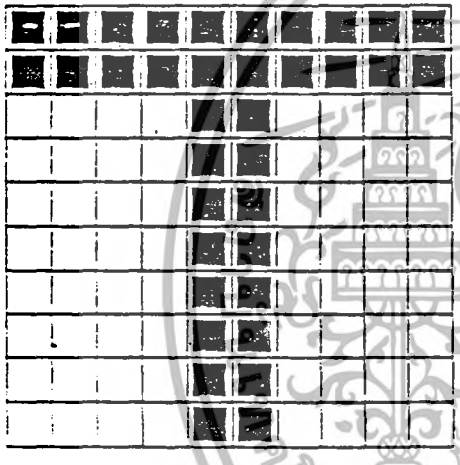
None needed.

VI. Strip byte counts (as needed)

None needed.

File is E6 bytes long.

. An 8086-format 10 pixel x 10 pixel CCITT/3 Compressed B&W TIFF Image



300 DPI Image
BINARY (Line art) image
CCITT/3 data compression
8086 byte order

	Hex	Name	Value
__x__	0FFH	Subfile Type	Value:__1___
__x__	100H	Image Width	Value:__10__
__x__	101H	Image Length	Value:__10__
__x__	102H	Bits per sample	Value:__1___
__x__	103H	Data compression	Value:__2___
__x__	106H	Photometric Interp.	Value:__1___
_____	107H	Thresholding	Value:_____
_____	109H	Cell Width	Value:_____

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

_____	109H	Cell Length	Value:_____
__x__	10AH	Fill Order	Value:__1___
_____	10DH	Document name	Value:_____
_____	10EH	Image description	Ascii:_____
_____	10FH	Scanner make	Ascii:_____
_____	110H	Scanner model	Ascii:_____
__x__	111H	Strip Offsets	Value/Offset:Offset
__x__	112H	Orientation	Value:__1___
__x__	115H	Samples per pixel	Value:__1___
__x__	116H	RowsPerStrip	Value:__1___
__x__	117H	StripByteCounts	Value/Offset:Offset
__x__	118H	Min Sample Value	Value:__0___
__x__	119H	Max Sample Value	Value:__1___
__x__	11AH	X resolution	Value:_300__
__x__	11BH	Y resolution	Value:_300__
__x__	11CH	Planar configuration	Value:___1__
_____	11DH	Page Name	Ascii:_____
_____	11EH	X position	Value:_____
_____	11FH	Y position	Value:_____
_____	120H	FreeOffsets	Value/Offset:_____
_____	121H	FreeByteCounts	Value/Offset:_____
_____	122H	GrayResponseUnit	Value:_____
_____	123H	GrayResponseCurve	Value/Offset:_____
_____	124H	Group3Options	Value:_____
_____	125H	Group4Options	Value:_____
_____	128H	ResolutionUnit	Value:_____
_____	129H	PageNumber	Value:_____
_____	12CH	ColorResponseUnit	Value:_____
_____	12DH	ColorResponseCurves	Value/Offset:_____

File worksheet

I. Header

Offset	Value	
000	49	Note byte order is Intel format.
001	49	
002	2A	Version #42, Note byte reversal
003	00	
004	08	Pointer to first directory
005	00	
006	00	
007	00	

II. Directory

008	11	17 tags in this directory
009	00	

	Offset	Tag	Type	Length	Value/Offset
Tag 1	00A	FF 00	03 00	01 00 00 00	01 00 00 00 (Subfile type)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Tag 2	016	00 01	03 00	01 00 00 00	0A 00 00 00	(Image width 10 pixels)
Tag 3	022	01 01	03 00	01 00 00 00	0A 00 00 00	(Image length 10 pixels)
Tag 4	02E	02 01	03 00	01 00 00 00	01 00 00 00	(1 Bit per sample)
Tag 5	03A	03 01	03 00	01 00 00 00	02 00 00 00	(CCITT/3 compression)
Tag 6	046	06 01	03 00	01 00 00 00	00 00 00 00	(Black is 1; however doesn't apply with CCITT/3)
Tag 7	052	0A 01	03 00	01 00 00 00	01 00 00 00	(Filled MSB to LSB)
Tag 8	05E	11 01	04 00	0A 00 00 00	FE 00 00 00	(Offset to pointers)
Tag 9	06A	12 01	03 00	01 00 00 00	01 00 00 00	(Normal orientation)
Tag 10	076	15 01	03 00	01 00 00 00	01 00 00 00	(1 sample/pixel)
Tag 11	082	16 01	04 00	01 00 00 00	01 00 00 00	(1 row/strip)
Tag 12	08E	17 01	04 00	0A 00 00 00	26 01 00 00	(Pointer to byte counts)
Tag 13	09A	18 01	03 00	01 00 00 00	00 00 00 00	(Min value is 0)
Tag 14	0A6	19 01	03 00	01 00 00 00	01 00 00 00	(Max value is 1)
Tag 15	0B2	1A 01	05 00	01 00 00 00	DA 00 00 00	(Offset to X resolution)
Tag 16	0BE	1B 01	05 00	01 00 00 00	E2 00 00 00	(Offset to Y resolution)
Tag 17	0CA	1C 01	03 00	01 00 00 00	01 00 00 00	(Planar config is normal)

End of directory

D6 00 00 00 00 (no more directories)

III. Extended tag information (as needed)

DA	2C 01 00 00	01 00 00 00 (300 dpi X resolution)
E2	2C 01 00 00	01 00 00 00 (300 dpi Y resolution)

IV. Compressed raster data (as needed)

EA	3508	(10 black pixels)
EC	3508	(10 black pixels)
EE	BEC0	(2 black pixels)
F0	BEC0	(2 black pixels)
F2	BEC0	(2 black pixels)
F4	BEC0	(2 black pixels)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

F6	BEC0	(2 black pixels)
F8	BEC0	(2 black pixels)
FA	BEC0	(2 black pixels)
FC	BEC0	(2 black pixels)

V. Strip offset pointers

FE	EA 00 00 00
102	EC 00 00 00
106	EE 00 00 00
10A	F0 00 00 00
10E	F2 00 00 00
112	F4 00 00 00
116	F6 00 00 00
11A	F8 00 00 00
11E	FA 00 00 00
122	FC 00 00 00

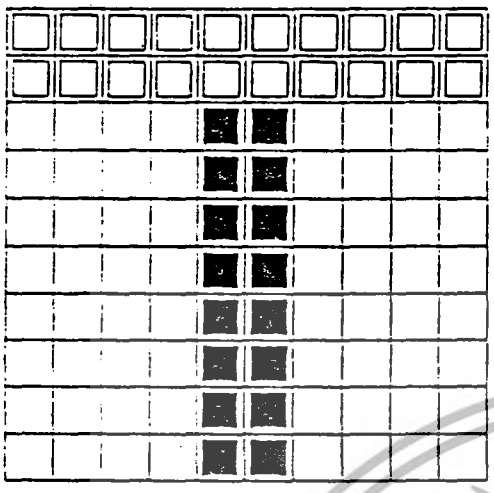
VI. Strip byte counts (as needed)

126	02 00 00 00
12A	02 00 00 00
12E	02 00 00 00
132	02 00 00 00
136	02 00 00 00
13A	02 00 00 00
13E	02 00 00 00
142	02 00 00 00
146	02 00 00 00
14A	02 00 00 00

File is 14E bytes long.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

An 8086-format 10 pixel x 10 pixel 4 Bit Gray Scale TIFF Image



300 DPI Image
4 bit gray scale image
No data compression
8086 byte order

	Hex	Name	Value
__x__	0FFH	Subfile Type	Value:__1___
__x__	100H	Image Width	Value:__10__
__x__	101H	Image Length	Value:__10__
__x__	102H	Bits per sample	Value:__4___
__x__	103H	Data compression	Value:__1___
__x__	106H	Photometric Interp.	Value:__1___
_____	107H	Thresholding	Value:_____
_____	108H	Cell Width	Value:_____
_____	109H	Cell Length	Value:_____
__x__	10AH	Fill Order	Value:__1___
_____	10DH	Document name	Value:_____
_____	10EH	Image description	Ascii:_____
_____	10FH	Scanner make	Ascii:_____
_____	110H	Scanner model	Ascii:_____
__x__	111H	Strip Offsets	Value/Offset:Offset
__x__	112H	Orientation	Value:__1___
__x__	115H	Samples per pixel	Value:__1___
_____	116H	RowsPerStrip	Value:_____
_____	117H	StripByteCounts	Value/Offset:_____
__x__	118H	Min Sample Value	Value:__0___
__x__	119H	Max Sample Value	Value:__1___
__x__	11AH	X resolution	Value:_300__
__x__	11BH	Y resolution	Value:_300__
__x__	11CH	Planar configuration	Value:___1__
_____	11DH	Page Name	Ascii:_____
_____	11EH	X position	Value:_____

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

____	11FH	Y position	Value:____
____	120H	FreeOffsets	Value/Offset:____
____	121H	FreeByteCounts	Value/Offset:____
__x__	122H	GrayResponseUnit	Value:__2___
__x__	123H	GrayResponseCurve	Value/Offset:Offset
____	124H	Group3Options	Value:____
____	125H	Group4Options	Value:____
____	128H	ResolutionUnit	Value:____
____	129H	PageNumber	Value:____
____	12CH	ColorResponseUnit	Value:____
____	12DH	ColorResponseCurves	Value/Offset:____

File worksheet

I. Header

Offset	Value	
000	49	Note byte order is Intel format.
001	49	
002	2A	Version #42, Note byte reversal
003	00	
004	08	Pointer to first directory
005	00	
006	00	
007	00	

II. Directory

008	11	17 tags in this directory
009	00	

	Offset	Tag	Type	Length	Value/Offset	
Tag 1	00A	FF 00	03 00	01 00 00 00	01 00 00 00	(Subfile type)
Tag 2	016	00 01	03 00	01 00 00 00	0A 00 00 00	(Image width 10 pixels)
Tag 3	022	01 01	03 00	01 00 00 00	0A 00 00 00	(Image length 10 pixels)
Tag 4	02E	02 01	03 00	01 00 00 00	04 00 00 00	(4 Bits per sample)
Tag 5	03A	03 01	03 00	01 00 00 00	01 00 00 00	(Packed data compression)
Tag 6	046	06 01	03 00	01 00 00 00	00 00 00 00	(Black is 1)
Tag 7	052	0A 01	03 00	01 00 00 00	01 00 00 00	(Filled MSB to LSB)
Tag 8	05E	11 01	04 00	01 00 00 00	0E 01 00 00	(Offset to raster data)
Tag 9	06A	12 01	03 00	01 00 00 00	01 00 00 00	(Normal orientation)
Tag 10	076	15 01	03 00	01 00 00 00	01 00 00 00	(1 sample/pixel)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Tag 11	082	18 01	03 00	01 00 00 00	00 00 00 00 (Min value is 0)
Tag 12	08E	19 01	03 00	01 00 00 00	0F 00 00 00 (Max value is 15)
Tag 13	09A	1A 01	05 00	01 00 00 00	DA 00 00 00 (Offset to X resolution)
Tag 14	0A6	1B 01	05 00	01 00 00 00	E2 00 00 00 (Offset to Y resolution)
Tag 15	0B2	1C 01	03 00	01 00 00 00	01 00 00 00 (Planar config is normal)
Tag 16	0BE	90 02	03 00	01 00 00 00	02 00 00 00 (Hundreths)
Tag 17	0CA	91 02	03 00	10 00 00 00	EA 00 00 00 (Pt: gray scale information)

End of directory

D6 00 00 00 00 (no more directories)

III. Extended tag information (as needed)

DA	2C 01 00 00	01 00 00 00 (300 dpi X resolution)
E2	2C 01 00 00	01 00 00 00 (300 dpi Y resolution)

EA	00 00 09 00	(gray response curve)
F2	10 00 13 00	
F6	17 00 1B 00	
FA	20 00 24 00	
FE	2B 00 32 00	
102	39 00 44 00	
106	4B 00 64 00	
10A	82 00 A0 00	

IV. Raster data (as needed)

10E	66 66 66 66 66 (10 medium gray pixels)
113	66 66 66 66 66 (10 medium gray pixels)
118	00 00 FF 00 00 (2 black pixels)
11D	00 00 FF 00 00 (2 black pixels)
122	00 00 FF 00 00 (2 black pixels)
127	00 00 FF 00 00 (2 black pixels)
12C	00 00 FF 00 00 (2 black pixels)
131	00 00 FF 00 00 (2 black pixels)
136	00 00 FF 00 00 (2 black pixels)
13B	00 00 FF 00 00 (2 black pixels)

V. Strip offset pointers (as needed)

None needed.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ง

โปรแกรมการรู้จำอักษรคัดลายมือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
/*************************************************
*  Automatic Design Dictionary For Thai Character *
*          ....  Pattern Recognition ....         *
*       Compile by :   TURBO C 2.0                *
*       Model      :   Compact                    *
*       Process    :   Segmentation               *
*                  :   Feature Extraction         *
*                  :   Unification                *
*                  :   Concentration              *
*                  :   Dictionary                 *
*       Data input :   TIFF format file           *
*       Programmer :   Mr.LEK                     *
*       Enhancer   :   COMP.Junior KMITL.         *
*       Check  by  :   T W E C (r)                *
*       Purpose    :   Find freq. of each code    *
*************************************************/

nclude <stdio.h>
nclude <alloc.h>
nclude <string.h>
nclude <conio.h>
nclude <stdlib.h>
nclude <graphics.h>

 SCREEN.C is subroutine to help displaying on text screen */
 This subroutine is compact to TWSCREEN.C but run on Thai language*/

nclude <dos.h>

d plot(int x,int y,char attr,char character);
d gotoxy(int x,int y);
d window(int x1,int y1,int x2,int y2);
d clearword(int a,int b,int c,int d);
d printstring(int x,int y,int length,char attr,char tprint[80]);


d plot(int x,int y,char attr,char character)  /* put character and attribute on
creen */

nt addr;
har  far *screenmem = MK_FP(0xb000,0);

F (x>79) x=79;
F (y>24) x =24;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
ddr = y*160+x*2;
(screenmem + addr) = character;
(screenmem + addr + 1) = attr;


d gotoxy(int x,int y)  /* locate cursor in column x and row y */

nion REGS i;

 i.h.dh = y;
 i.h.dl = x;
 i.h.ah = 2;
 i.h.bh = 0;
 int86(16,&i,&i);


.d window(int x1,int y1,int x2,int y2)  /*  draw window */

.nt e;
'define   a1 149
'define   a2 155
'define   a3 151
'define   a4 157
'define   a5 159
'define   a6 158

)lot(x1,y1,0xf,a1);
)lot(x1,y2,0xf,a2);
)lot(x2,y1,0xf,a3);
)lot(x2,y2,0xf,a4);
'or(e=x1+1;e<x2;e++)
  {
  plot(e,y1,0xf,a6);
  plot(e,y2,0xf,a6);
  }
'or (e = y1+1;e<y2;e++)
  {
  plot(x1,e,0xf,a5);
  plot(x2,e,0xf,a5);
  }


.d clearword(int a,int b,int c,int d)    /* clear  block screen */
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
t e,f;

r(f = b;f<=d;f++)
for(e = a;e<=c;e++)
  plot(e,f,0xf,' ');


 printstring(int x,int y,int length,char attr,char tprint[80])

t a;
r (a=0;a < length;a++)
plot(a+x,y,attr,tprint[a]);


ine  false     0
ine  true      1
ine  boolean int

gned char *buffer;
gned char *block;
     *beef;
     fname[30],dicname[30],letterpath[30],diccall[60];
     *fp,*fdic,*dicfile;
     *beat;
     hsize,vsize,listcount,fnum,fsize;
     ff = 12;  /* FormFeed */
gned long memreq;
ct hist_t {
            unsigned int code;
            int freq;
          } hist[8192];
     check;
 comlist[5][27] ={"  1. สร้างพจนานุกรม          ",
                  "  2. เพิ่มเติมพจนานุกรม      ",
                  "  3. ทดสอบคำอังกฤษ           ",
                  "  4. กำหนดPATHไฟล์อังกฤษ     ",
                  "  5. ออกจากโปรแกรม           "};

---------------------------------------------------------- */
scanver(buff,x)                  /* Return 0 if no black dot */
gned char far *buff;
     x;

gned char mask;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
t       i,j,h2,exist;

exist = 0;
h2 = (hsize + 7) /8;
mask = 0x80 >> (x % 8);
j = x >> 3;
for (i = 0; (i<vsize) && !exist; i++)
 if (!(buff[i*h2+j] & mask))   /* There is black dot */
 exist = 1;
 return (exist);


t scanhor(buff,y,x1,x2)          /* Return 0 if no black dot */
signed char far *buff;
t       y,x1,x2;

signed char mask,data;
t       i,j,h2,exist,nloop;

exist = 0;
nloop = x2 - x1 + 1;
h2 = (hsize + 7)/8;
mask = 0x80 >> (x1 % 8);
data = buff[y*h2+(x1>>3)];
for (i=0; (i<nloop) && !exist;i++)
{
  if (!(data & mask))          /* There is black dot */
  exist = 1;
  x1++;
  if (!(mask >>= 1))
  {
    data = buff[y*h2+(x1>>3)];
    mask = 0x80;
  }
}
return (exist);


 ------------------------------------------------------------ */
id segment(buff,x1,x2,y1,y2,block)
signed char far *buff;
signed char far *block;
t       x1,x2,y1,y2;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
gned char mask,data;
gned char far *b2;
     xsize;
     i,j,x,y,h2;
     message[40];


* segmenting */

2 = (hsize+7) >> 3;
size = x2-x1+1;
or (y=y1,j=0; y<=y2; y++,j++)

 mask = 0x80 >> (x1 % 8);
 i = 0;
 data = buff[y*h2 + ((x1+i) >> 3)];
 while (i<xsize)
 {
      block[i+j*xsize] = data & mask ? 0:1 ;
      i++;
      if (!(mask >>= 1))
      {
         mask = 0x80;
         data = buff[y*h2 + ((x1+i) >> 3)];
      }
 }


---------------------------------------------------- */
 displaypix(block,xlen,ylen)
gned char *block;
 xlen,ylen;


 x,y,dot,ln;
igned char rest;
ar far *scrmem = MK_FP(0xb800,0x590);

t = 0;
 (y=0;y<=14;y++)
or (x=0;x<=7;x++)
  for(ln = 0;ln<5;ln++)
    *(scrmem+0x2000*ln+x+y*90) = rest;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
or (      y  = 0;y  <=(ylen/4);y++)
 for (    x  = 0;x  <=(xlen/8);x++)
   for (  ln = 0;ln < 5;      ln++)
     {
       rest = 0;
       for (dot = 0;dot< 8;dot ++)
         rest = (rest << 1)+ *(block+((y*4)+ln)*xlen+(y*4+ln)+(x*8)+dot);
       *(scrmem+0x2000*ln+x+y*90) = rest;
     }
```

```
---------------------------------------------------- */
r p_code(ee,nn,ww,ss)
 unsigned char  ee,nn,ww,ss;

 char  c;
 if(((1-nn*ss)*ee*ww)==1)     c='h';
 if(((1-ee*ww)*nn*ss)==1)     c='v';
 if(((1-ee*ww)*(1-nn*ss))==1) c='s';
 if((ee*nn*ww*ss)==1)         c='i';
 return(c);

 q_code(eq,nq,wq,sq)
  boolean  eq,nq,wq,sq;

     char  cc;
                              cc='0';   if(sq)              cc='1';
     if(wq)          cc='2';  if(wq&&sq)          cc='3';
     if(nq)          cc='4';  if(nq&&sq)          cc='5';
     if(nq&&wq)      cc='6';  if(nq&&wq&&sq)      cc='7';
     if(eq)          cc='8';  if(eq&&sq)          cc='9';
     if(eq&&wq)      cc='A';  if(eq&&wq&&sq)      cc='B';
     if(eq&&nq)      cc='C';  if(eq&&nq&&sq)      cc='D';
     if(eq&&nq&&wq) cc='E';   if(eq&&nq&&wq&&sq) cc='F';
     return(cc);


 extract(bloc,x1,x2,y1,y2,bee)
  unsigned char far *bloc;
  unsigned char far *bee;
  int      x1,x2,y1,y2;

 t      x,y,i,j,k,l,xsize;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
boolean  qs,qe,qn,qw;
char     p_code(), q_code();

xsize=x2-x1+1;
x=x2-x1;
y=y2-y1;
for(i=0;i<=y;i++)

  for(j=0;j<=x;j++)
  if(*(bloc+xsize*i+j)==1)
  {
       if(i==0) *(bee+xsize*i+j)=p_code(*(bloc+xsize*i+j+1),0,
                      *(bloc+xsize*i+j-1),*(bloc+xsize*(i+1)+j));
       else
       if(i==y) *(bee+xsize*i+j)=p_code(*(bloc+xsize*i+j+1),
                      *(bloc+xsize*(i-1)+j),*(bloc+xsize*i+j-1),0);
       else
       if(j==0) *(bee+xsize*i+j)=p_code(*(bloc+xsize*i+j+1),
                      *(bloc+xsize*(i-1)+j),0,*(bloc+xsize*(i+1)+j));
       else
       if(j==x) *(bee+xsize*i+j)=p_code(0,*(bloc+xsize*(i-1)+j),
                      *(bloc+xsize*i+j-1),*(bloc+xsize*(i+1)+j));
       else
                *(bee+xsize*i+j)=p_code(*(bloc+xsize*i+j+1),
                *(bloc+xsize*(i-1)+j),*(bloc+xsize*i+j-1),
                *(bloc+xsize*(i+1)+j));
  }
  else
  {
      k=i;l=j;
        while((k++<=y)&&(*(bloc+(xsize*k)+l)==0));
        if (k<=y) qs=true; else  qs=false;
      k=i;l=j;
        while((k-->=0)&&(*(bloc+(xsize*k)+l)==0));
        if (k>=0) qn=true; else  qn=false;
      k=i;l=j;
        while((l++<=x)&&(*(bloc+(xsize*k)+l)==0));
        if (l<=x) qe=true; else qe=false;
      k=i;l=j;
        while((l-->=0)&&(*(bloc+(xsize*k)+l)==0));
        if (l>=0) qw=true; else qw=false;
      *(bee+xsize*i+j) = q_code(qe,qn,qw,qs);
  }
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
; ------------------------------------------------------------ */
ıar unif(i,j,wid,bee)
    unsigned char far *bee;
    int       i,j,wid;

 har c;
witch(*(bee+wid*i+j))

 case'F': if((*(bee+i*wid+j+1)=='E')||(*(bee+(i-1)*wid+j)=='E')||
                    (*(bee+i*wid+j-1)=='E')||(*(bee+(i+1)*wid+j)=='E'))
                    c='E';
     else if((*(bee+i*wid+j+1)=='D')||(*(bee+(i-1)*wid+j)=='D')||
                    (*(bee+i*wid+j-1)=='D')||(*(bee+(i+1)*wid+j)=='D'))
                    c='D';
        else if((*(bee+i*wid+j+1)=='C')||(*(bee+(i-1)*wid+j)=='C')||
                    (*(bee+i*wid+j-1)=='C')||(*(bee+(i+1)*wid+j)=='C'))
                    c='C';
        else if((*(bee+i*wid+j+1)=='B')||(*(bee+(i-1)*wid+j)=='B')||
                    (*(bee+i*wid+j-1)=='B')||(*(bee+(i+1)*wid+j)=='B'))
                    c='B';
     else if((*(bee+i*wid+j+1)=='A')||(*(bee+(i-1)*wid+j)=='A')||
                    (*(bee+i*wid+j-1)=='A')||(*(bee+(i+1)*wid+j)=='A'))
                    c='A';
     else if((*(bee+i*wid+j+1)=='9')||(*(bee+(i-1)*wid+j)=='9')||
                    (*(bee+i*wid+j-1)=='9')||(*(bee+(i+1)*wid+j)=='9'))
                    c='9';
     else if((*(bee+i*wid+j+1)=='7')||(*(bee+(i-1)*wid+j)=='7')||
                    (*(bee+i*wid+j-1)=='7')||(*(bee+(i+1)*wid+j)=='7'))
                    c='7';
     else if((*(bee+i*wid+j+1)=='5')||(*(bee+(i-1)*wid+j)=='5')||
                    (*(bee+i*wid+j-1)=='5')||(*(bee+(i+1)*wid+j)=='5'))
                    c='5';
        else if((*(bee+i*wid+j+1)=='3')||(*(bee+(i-1)*wid+j)=='3')||
                    (*(bee+i*wid+j-1)=='3')||(*(bee+(i+1)*wid+j)=='3'))
                    c='3';
        else    c='F';
     break;

 case'E': if((*(bee+i*wid+j+1)=='C')||(*(bee+(i-1)*wid+j)=='C')||
                    (*(bee+i*wid+j-1)=='C')||(*(bee+(i+1)*wid+j)=='C'))
                    c='C';
        else if((*(bee+i*wid+j+1)=='B')||(*(bee+(i-1)*wid+j)=='B')||
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
                (*(bee+i*wid+j-1)=='B')||(*(bee+(i+1)*wid+j)=='B'))
                c='B';
      else if((*(bee+i*wid+j+1)=='A')||(*(bee+(i-1)*wid+j)=='A')||
                (*(bee+i*wid+j-1)=='A')||(*(bee+(i+1)*wid+j)=='A'))
                c='A';
      else if((*(bee+i*wid+j+1)=='6')||(*(bee+(i-1)*wid+j)=='6')||
                (*(bee+i*wid+j-1)=='6')||(*(bee+(i+1)*wid+j)=='6'))
                c='6';
   else    c='E';
   break;

case'D': if((*(bee+i*wid+j+1)=='C')||(*(bee+(i-1)*wid+j)=='C')||
                (*(bee+i*wid+j-1)=='C')||(*(bee+(i+1)*wid+j)=='C'))
                c='C';
   else if((*(bee+i*wid+j+1)=='9')||(*(bee+(i-1)*wid+j)=='9')||
                (*(bee+i*wid+j-1)=='9')||(*(bee+(i+1)*wid+j)=='9'))
                c='9';
      else if((*(bee+i*wid+j+1)=='7')||(*(bee+(i-1)*wid+j)=='7')||
                (*(bee+i*wid+j-1)=='7')||(*(bee+(i+1)*wid+j)=='7'))
                c='7';
      else if((*(bee+i*wid+j+1)=='5')||(*(bee+(i-1)*wid+j)=='5')||
                (*(bee+i*wid+j-1)=='5')||(*(bee+(i+1)*wid+j)=='5'))
                c='5';
   else    c='D';
   break;

case'B': if((*(bee+i*wid+j+1)=='A')||(*(bee+(i-1)*wid+j)=='A')||
                (*(bee+i*wid+j-1)=='A')||(*(bee+(i+1)*wid+j)=='A'))
                c='A';
      else if((*(bee+i*wid+j+1)=='9')||(*(bee+(i-1)*wid+j)=='9')||
                (*(bee+i*wid+j-1)=='9')||(*(bee+(i+1)*wid+j)=='9'))
                c='9';
      else if((*(bee+i*wid+j+1)=='3')||(*(bee+(i-1)*wid+j)=='3')||
                (*(bee+i*wid+j-1)=='3')||(*(bee+(i+1)*wid+j)=='3'))
                c='3';
            break;

case'7': if((*(bee+i*wid+j+1)=='6')||(*(bee+(i-1)*wid+j)=='6')||
                (*(bee+i*wid+j-1)=='6')||(*(bee+(i+1)*wid+j)=='6'))
                c='6';
      else if((*(bee+i*wid+j+1)=='5')||(*(bee+(i-1)*wid+j)=='5')||
                (*(bee+i*wid+j-1)=='5')||(*(bee+(i+1)*wid+j)=='5'))
                c='5';
      else if((*(bee+i*wid+j+1)=='3')||(*(bee+(i-1)*wid+j)=='3')||
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
                (*(bee+i*wid+j-1)=='3')||(*(bee+(i+1)*wid+j)=='3'))
                c='3';
    else    c='7';
    break;

default : c=(*(bee+i*wid+j));
}
return(c);


id unificate(bee,wid,lon)
  unsigned char far *bee;
  int      wid,lon;

  int   i,j;
  char  unif();

  for(i=0;i<=lon;i++)
    for(j=0;j<=wid;j++)
        *(bee+i*wid+j) = unif(i,j,wid,bee);
  for(i=0;i<=lon;i++)
    for(j=wid;j>=0;j--)
        *(bee+i*wid+j) = unif(i,j,wid,bee);
  for(i=lon;i>=0;i--)
    for(j=wid;j>=0;j--)
        *(bee+i*wid+j) = unif(i,j,wid,bee);
  for(i=lon;i>=0;i--)
    for(j=0;j<=wid;j++)
        *(bee+i*wid+j) = unif(i,j,wid,bee);
  for(i=0;i<=wid;i++)
    for(j=0;j<=lon;j++)
        *(bee+i*wid+j) = unif(i,j,wid,bee);
  for(i=0;i<=wid;i++)
    for(j=lon;j>=0;j--)
        *(bee+i*wid+j) = unif(i,j,wid,bee);
  for(i=wid;i>=0;i--)
    for(j=0;j<=lon;j++)
        *(bee+i*wid+j) = unif(i,j,wid,bee);
  for(i=wid;i>=0;i--)
    for(j=lon;j>=0;j--)
        *(bee+i*wid+j) = unif(i,j,wid,bee);
  for(i=0;i<lon;i++)
  { fprintf(stdprn,"\n");
      for(j=0;j<wid;j++)
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    { fprintf(stdprn,"%c",*(bee+i*wid+j));  }
}


---------------------------------------------- */
createdic(re,rn,rw,rs,rc)
char re,rn,rw,rs,rc;

t    conw;
nw=0x0000;

(rc>'G')
itch (rc)

ase 'h':conw=conw+0x0001; break;   case 'v':conw=conw+0x0002; break;
ase 's':conw=conw+0x0004; break;   case 'i':conw=conw+0x0008;

se
itch(rc)

ase '3':conw=conw+0x0003 ; break;  case '5':conw=conw+0x0005 ; break;
ase '6':conw=conw+0x0006 ; break;  case '7':conw=conw+0x0007 ; break;
ase '9':conw=conw+0x0009 ; break;  case 'A':conw=conw+0x000a ; break;
ase 'B':conw=conw+0x000b ; break;  case 'C':conw=conw+0x000c ; break;
ase 'D':conw=conw+0x000d ; break;  case 'E':conw=conw+0x000e ; break;
ase 'F':conw=conw+0x000f ;

itch(re)

ase '3':conw=conw+0x8000; break;   case '5':conw=conw+0xa000; break;
ase '6':conw=conw+0x2000; break;   case '7':conw=conw+0xa000; break;
ase '9':conw=conw+0xc000; break;   case 'A':conw=conw+0x4000; break;
ase 'B':conw=conw+0xc000; break;   case 'C':conw=conw+0x6000; break;
ase 'D':conw=conw+0xe000; break;   case 'E':conw=conw+0x6000; break;
ase 'F':conw=conw+0xe000;

itch(rn)

ase '3':conw=conw+0x0400; break;   case '5':conw=conw+0x0800; break;
ase '6':conw=conw+0x0c00; break;   case '7':conw=conw+0x0c00; break;
ase '9':conw=conw+0x1000; break;   case 'A':conw=conw+0x1400; break;
ase 'B':conw=conw+0x1400; break;   case 'C':conw=conw+0x1800; break;
ase 'D':conw=conw+0x1800; break;   case 'E':conw=conw+0x1c00; break;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
 case 'F':conw=conw+0x1c00;
}
switch(rw)
{
 case '3':conw=conw+0x0180; break;   case '5':conw=conw+0x0280; break;
 case '6':conw=conw+0x0300; break;   case '7':conw=conw+0x0380; break;
 case '9':conw=conw+0x0080; break;   case 'A':conw=conw+0x0100; break;
 case 'B':conw=conw+0x0180; break;   case 'C':conw=conw+0x0200; break;
 case 'D':conw=conw+0x0280; break;   case 'E':conw=conw+0x0300; break;
 case 'F':conw=conw+0x0380;
}
switch(rs)
{
 case '3':conw=conw+0x0060; break;   case '5':conw=conw+0x0020; break;
 case '6':conw=conw+0x0040; break;   case '7':conw=conw+0x0060; break;
 case '9':conw=conw+0x0030; break;   case 'A':conw=conw+0x0050; break;
 case 'B':conw=conw+0x0070; break;   case 'C':conw=conw+0x0010; break;
 case 'D':conw=conw+0x0030; break;   case 'E':conw=conw+0x0050; break;
 case 'F':conw=conw+0x0070;
}
return(conw);


id con_word(bee,wi,ln,dic)
ar   *bee;
t    *dic;
t    wi,ln;

char ec,nc,wc,sc;
int  i,j,k,l;
int  createdic();

for(i=0;i<ln;i++)
 for(j=0;j<wi;j++)
  { ec=nc=wc=sc='n';
       if (*(bee+wi*i+j)>'G')     /* black stoke (>G) */
    {
         k=i;l=j;
              while ((l++<wi)&&(*(bee+wi*k+l)>'G'));
              if (l<wi) ec=*(bee+wi*k+l); else ec='n';
         k=i;l=j;
              while ((k-->=0)&&(*(bee+wi*k+l)>'G'));
              if (k>=0) nc=*(bee+wi*k+l); else nc='n';
         k=i;l=j;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
            while ((l-->=0)&&(*(bee+wi*k+l)>'G'));
            if (l>=0) wc=*(bee+wi*k+l); else wc='n';
        k=i;l=j;
            while ((k++<ln)&&(*(bee+wi*k+l)>'G'));
            if (k<ln) sc=*(bee+wi*k+l); else sc='n';
    }
      else                              /* white stoke (<G) */
    {
        k=i;l=j+1;
            while (l++<wi-1)
            { if((*(bee+wi*k+l)>'G')&&(*(bee+wi*k+l+1)<'G')) break; }
            if (l<wi-1)  ec=*(bee+wi*k+l+1);  else ec='n';
        k=i-1;l=j;
            while (k-->0)
            { if((*(bee+wi*k+l)>'G')&&(*(bee+wi*(k-1)+l)<'G')) break; }
            if (k>0)  nc=*(bee+wi*(k-1)+l);  else nc='n';
        k=i;l=j-1;
            while (l-->0)
            { if((*(bee+wi*k+l)>'G')&&(*(bee+wi*k+l-1)<'G')) break; }
            if (l>0)  wc=*(bee+wi*k+l-1);  else wc='n';
        k=i+1;l=j;
            while (k++<ln-1)
            { if((*(bee+wi*k+l)>'G')&&(*(bee+wi*(k+1)+l)<'G')) break; }
            if (k<ln-1)  sc=*(bee+wi*(k+1)+l);  else sc='n';
    }
  *(dic+(wi*i+j))=createdic(ec,nc,wc,sc,*(bee+wi*i+j));
  }
}


* ------------------------------------------------------ */
oid sortdic(dic,xsize,ysize)
    unsigned  *dic;
    int   xsize,ysize;

nt    i,j,ncode,sum,cmp();
loat  perc,xsum;
truct hist_t temp;

sum=xsize*ysize;
qsort(dic,sum,2,cmp);

for(i=0;i<8192;i++)
  { hist[i].code = 0;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
 .hist[i].freq = 0;
}

or(i=0,ncode=0;i<sum;i++)
if(*(dic+i)==*(dic+i+1))   hist[ncode].freq++;
else
   { hist[ncode].code = *(dic+i);
     hist[ncode++].freq = hist[ncode].freq+1; }

or(i=0;i<ncode;i++)
for(j=ncode-1; j>i; j--)
  if((hist[j].freq-hist[i].freq) > 0)
    {
    temp.freq    = hist[i].freq; temp.code    = hist[i].code;
    hist[i].freq = hist[j].freq; hist[i].code = hist[j].code;
    hist[j].freq = temp.freq;    hist[j].code = temp.code;
    }
for(i=0;i<5;i++)
fprintf(stdout," %5x-%3d ",hist[i].code,hist[i].freq);
 fprintf(stdout," %5x ",hist[i].code);                    */


================================================================ */
cmp(e1,e2)
gned *e1,*e2;
eturn(*e1-*e2);  }

================================================================ */
doprocess(void)

     x,y,j,exist,first,last;
     h1,h2,v1,v2,hor,ver;
     input[255];


      if((fp = fopen(fname,"rb")) == NULL)
      { gotoxy(33,19);
        printf("ไม่พบแฟ้มข้อมูล %s",fname);
        return(1);
      }
    clearword(33,19,77,22);
       gotoxy(34,20);printf("กำลังวิเคราะห์แฟ้มข้อมูล %s...",fname);
    fseek(fp,0x1e,SEEK_SET);        /* Get x width of TIFF image */
    hsize = getc(fp);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
hsize += getc(fp)*256;
fseek(fp,0x2a,SEEK_SET);        /* Get y length of TIFF image */
vsize = getc(fp);
vsize += getc(fp)*256;

memreq = (unsigned long)((hsize+7)/8*vsize);

if((buffer = (unsigned char *) farmalloc(memreq)) == NULL)
  { gotoxy(33,19);
    fprintf(stdout,"ไม่สามารถทำงานได้ หน่วยความจำไม่พอ"); return(1); }

fseek(fp,0x156,SEEK_SET);
if(fread(buffer,memreq,1,fp) != 1)
  { gotoxy(33,19);
    fprintf(stderr,"ไม่พบไฟล์"); return(1); }
   fclose(fp);

x = 0;                  /* Extract each character */
   while (x<hsize)
        {
          while (!scanver(buffer,x) && (x<hsize)) x++;
     if (x >= hsize) break;
     h1 = x;
     while (scanver(buffer,x) && (x<hsize))  x++;
     h2 = x-1;
     y = 0;
          while (!scanhor(buffer,y,h1,h2) && (y<vsize)) y++;
     v1 = y;
     while (scanhor(buffer,y,h1,h2) && (y<vsize))  y++;
     v2 = y-1;

     hor = h2-h1+1;   ver = v2-v1+1;

     if((block = (unsigned char *) farmalloc(hor*ver)) == NULL)
          { fprintf(stderr,"Cannot allocate block memory"); exit(0); }
     if((beef = (char *) farmalloc(hor*ver)) == NULL)
          { fprintf(stderr,"Cannot allocate beef memory");  exit(0); }
     if((beat = (int *) farmalloc(2*hor*ver)) == NULL)
          { fprintf(stderr,"Cannot allocate beat memory");  exit(0); }

     gotoxy(33,22);printf(">>>>>");
     segment(buffer,h1,h2,v1,v2,block);
     displaypix(block,h2-h1,v2-v1);
     gotoxy(33,22);printf(">>>>>>>>>>>>>>>>>>");
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        extract(block,h1,h2,v1,v2,beef);
        gotoxy(33,22);printf(">>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>");
        farfree(block);
        unificate(beef,hor,ver);
        gotoxy(33,22);printf(">>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>");
        con_word(beef,hor,ver,beat);
        sortdic(beat,hor,ver);
        gotoxy(33,22);printf("****************************************");
        farfree(beat);
        farfree(beef);
        clearword(32,18,77,22);
       }
       farfree(buffer);


============================================================ */
. doprocess1(void)

.     x,y,i,j,k,exist,first;
.     h1,h2,v1,v2,hor,ver,aa1;
r     input[100];
.     count,last,check,fs1,fsh;
igned int  diccheck[101][6];
r     test,REST[3];
igned int resth,rest1;

     if((fp = fopen(fname,"rb")) == NULL)
    {
            clearword(33,18,77,22);
            gotoxy(33,21);printf("ไม่สามารถเปิดแฟ้มภาพกราฟฟิกได้");
            gotoxy(33,22);printf("จบการทำงานโดยผิดพลาด");
            getch();        return;
    }
     clearword(33,18,77,22);
     gotoxy(34,20);printf("กำลังวิเคราะห์แฟ้มภาพ %s...",fname);
     fseek(fp,0x1e,SEEK_SET);        /* Get x width of TIFF image */
     hsize = getc(fp);
     hsize += getc(fp)*256;
     fseek(fp,0x2a,SEEK_SET);       /* Get y length of TIFF image */
     vsize = getc(fp);
     vsize += getc(fp)*256;

     memreq = (unsigned long)((hsize+7)/8*vsize);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
if((buffer = (unsigned char *) farmalloc(memreq)) == NULL)
{    clearword(33,19,77,22);
        gotoxy(33,19);
        fprintf(stdout,"ไม่สามารถทำงานได้ หน่วยความจำ BUFFER ไม่พอ");
        getch();  return;
}
fseek(fp,0x156,SEEK_SET);
if(fread(buffer,memreq,1,fp) != 1)
{    clearword(33,19,77,22);
        gotoxy(33,19);
        fprintf(stderr,"ไม่พบข้อมูลที่ต้องการจากไฟล์");
        getch();  return;
}
fclose(fp);

x = j = 0 ;                /* Extract each character */
while (x<hsize)
{
  if ((dicfile = fopen(dicname,"rb")) == NULL)
  { clearword(33,19,77,22);
    gotoxy(33,19);printf("ไม่สามารถเปิดแฟ้มพจนานุกรมได้");
    getch();  return;
  }

  while (!scanver(buffer,x) && (x<hsize)) x++;
  if (x >= hsize) break;
  h1 = x;
  while (scanver(buffer,x) && (x<hsize))  x++;
  h2 = x-1;

  y = 0;
  while (y<vsize)
  {
        while (!scanhor(buffer,y,h1,h2) && (y<vsize)) y++;
        v1 = y;
        while (scanhor(buffer,y,h1,h2) && (y<vsize))  y++;
        v2 = y-1;
        y = y+1;
        hor = h2-h1+1;   ver = v2-v1+1;

        if(hor*ver==0) break;
        if((block = (unsigned char *) farmalloc(hor*ver)) == NULL)
        {    clearword(33,19,77,22);
             gotoxy(33,19);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        fprintf(stderr,"ไม่สามารถจองหน่วยความจำในส่วนของ block ได้");
        getch();  return;
}
if((beef = (char *) farmalloc(hor*ver)) == NULL)
{    clearword(33,19,77,22);
        gotoxy(33,19);
        fprintf(stderr,"ไม่สามารถจองหน่วยความจำในส่วนของ beef ได้");
        getch();  return;
}
if((beat = (int *) farmalloc(2*hor*ver)) == NULL)
{       clearword(33,19,77,22);
        gotoxy(33,19);
        fprintf(stderr,"ไม่สามารถจองหน่วยความจำในส่วนของ beat ได้");
        getch();   return;
}
clearword(33,19,77,22);
gotoxy(33,19);printf("scanning.....");
gotoxy(33,22);printf(">>>>>");
segment(buffer,h1,h2,v1,v2,block);
displaypix(block,h2-h1,v2-v1);
gotoxy(33,22);printf(">>>>>>>>>>>>>>>>>>");
extract(block,h1,h2,v1,v2,beef);
gotoxy(33,22);printf(">>>>>>>>>>>>>>>>>>>>>>>>>>>>>");
farfree(block);
uniticate(beef,hor,ver);
gotoxy(33,22);printf(">>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>");
con_word(beef,hor,ver,beat);
farfree(beef);
sortdic(beat,hor,ver);
gotoxy(33,22);printf("*****************************************");
farfree(beat);

last= 1;
fs1 = getc(dicfile);
fsh = getc(dicfile);
fsize =fsh*256+fs1;

for(last=1;last<=fsize;last++)
{
        i=0;
        resth = getc(dicfile);
        restl = getc(dicfile);
        diccheck[last][0] = resth * 256 + restl;
        resth = getc(dicfile);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
            restl = getc(dicfile);
            diccheck[last][1] = resth * 256 + restl;
            resth = getc(dicfile);
            restl = getc(dicfile);
            diccheck[last][2] = resth * 256 + restl;
            resth = getc(dicfile);
            restl = getc(dicfile);
            diccheck[last][3] = resth * 256 + restl;
            resth = getc(dicfile);
            restl = getc(dicfile);
            diccheck[last][4] = resth * 256 + restl;
            resth = getc(dicfile);
            restl = getc(dicfile);
            diccheck[last][5] = resth * 256 + restl;

            if(hist[0].code!=diccheck[last][0]) continue;
            if(hist[1].code!=diccheck[last][1]) continue;
    if(hist[2].code!=diccheck[last][2]) continue;
            if(hist[3].code!=diccheck[last][3]) continue;
            if(hist[4].code!=diccheck[last][4]) continue;

            j = j+1;
            clearword(33,19,77,22);
            gotoxy(33+j,18);printf("%c",(char)diccheck[last][5]);
            gotoxy(33,20);printf("คำว่าภาษาคือ ('%c') code:%d",
            (char)diccheck[last][5],diccheck[last][5]);
            gotoxy(33,22);printf("โปรดรอสักครู่ ขณะเพื่อนทำงานต่อ...");
            /*getch();*/
            fclose(dicfile);
            i=1;
            break;
     }
     if(i==1) continue;
     clearword(33,19,77,22);
     gotoxy(33,19);
     printf("เสียใจด้วยนะ! เราไม่ทราบว่าเป็นอักษรอะไร");
     gotoxy(33,20);
     printf("คุณต้องการให้เราจดจำคำว่าภาษาหรือไม่?(Y/N)");

     for(test=0x0;((test!='y')&&(test!='Y')&&(test!='n')&&
     (test!='N'));  test=getch());
     printf("%c",test);

     if ((test == 'N') || (test =='n'))
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
        {       fclose(dicfile);
                break;
        }

        gotoxy(33,22);printf("ใปรดใส ศ.ัวอ.ภษร :");
        gets(REST);
        fnum=((unsigned char)*REST);
        j = j+1;
        gotoxy(33+j,18);printf("%c",fnum);

        dicfile = fopen(dicname,"ab");
        fputc(hist[0].code>>8,dicfile);fputc(hist[0].code%256,dicfile);
        fputc(hist[1].code>>8,dicfile);fputc(hist[1].code%256,dicfile);
        fputc(hist[2].code>>8,dicfile);fputc(hist[2].code%256,dicfile);
        fputc(hist[3].code>>8,dicfile);fputc(hist[3].code%256,dicfile);
        fputc(hist[4].code>>8,dicfile);fputc(hist[4].code%256,dicfile);
        fputc(fnum >> 8,dicfile);       fputc(fnum % 256,dicfile);
        fclose(dicfile);

        dicfile = fopen(dicname,"r+b");
        fseek(dicfile,0x00,SEEK_SET);
        fsl = getc(dicfile);
        fsh = getc(dicfile);

        fseek(dicfile,0x00,SEEK_SET);
        fsize =fsh*256+fsl+1;
        fputc(fsize % 256,dicfile);fputc(fsize >> 8,dicfile);
        fclose(dicfile);
        }
    }
    farfree(buffer);


:==========================================================================*/
i displaycommand(void)


it a;

intstring(11,12,13,0x78,"  ใปรดเลือก  ");
or (a = 0;a<5;a++) printstring(3,a+14,27,0x07,comlist[a]);
intstring(3,14+listcount,27,0x70,comlist[listcount]);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
d display(void)

lrscr();
indow(1,1,79,24);
indow(2,2,78,9);
indow(68,3,77,8);
indow(2,10,30,23);
indow(31,10,78,16);
indow(31,17,78,23);

rintstring(25,4,31,0x0f,"โปรแกรมออกแบบพจนานุกรมอังกฤษ");
rintstring(31,5,20,0x0f,"สำหรับผู้ใช้คอมพิวเตอร์");
rintstring(18,6,46,0x07,"สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง");
rintstring(37,7,9,0x07,"รุ่นที่ 1");

isplaycommand();

rintstring(33,12,17,0x0f,"แฟ้มพจนานุกรม :");
rintstring(33,13,22,0x0f,"Pathของไฟล์ศัพท์อังกฤษ: ");
rintstring(33,14,17,0x0f,"ศัพท์อังกฤษที่สอบ :");
lot(60,14,0x0f,':');



                                                                    */

 Createdic(int check)

ar input[30],incheck;
t  count,first,last,test;
t fsl,fsh;
ize = 0;

earword(33,19,77,22);
toxy(33,19);printf("โปรดใส่ชื่อแฟ้มพจนานุกรม :");gets(input);
 (input !=NULL)

  sprintf(dicname,"%s",input);
  printstring(50,12,20,0x0f,"                    ");
  printstring(50,12,strlen(dicname),0x0f,dicname);

  if (check == 0)
    {
    dicfile = fopen(dicname,"wb");
     fputc(0x00,dictile);fputc(0x00,dicfile);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
    }
  else
    dicfile = fopen(dicname,"ab");
  incheck = 'n';
  while ((incheck == 'n') || (incheck == 'N'))
    {
       clearword(33,20,77,22);
       gotoxy(33,20);printf("ป้อนค่าเริ่มต้นของรหัสที่ต้องการเริ่มต้น :");gets(input);
       first = atoi(input);
       gotoxy(33,21);printf("ป้อนค่าสิ้นสุดของรหัสที่ต้องการสิ้นสุด :");gets(input);
       last  = atoi(input);
       gotoxy(33,22);printf("ท่านต้องการเปลี่ยนแปลง %d ถึง %d (Y/N):",first,last);
       incheck = getch();printf("%c",check);
    }
  clearword(33,20,77,22);
  for (count = first;count <=last;count++)
    {
      fnum = count;
      sprintf(fname,"%s%d.TIF",letterpath,fnum);
      gotoxy(50,14);
      fprintf(stdout," %3d      : ('%c')",fnum,(char)fnum);
      test = doprocess();
      if (test != 1)
      {
      fputc(hist[0].code >> 8,dicfile);fputc(hist[0].code % 256,dicfile);
      fputc(hist[1].code >> 8,dicfile);fputc(hist[1].code % 256,dicfile);
      fputc(hist[2].code >> 8,dicfile);fputc(hist[2].code % 256,dicfile);
      fputc(hist[3].code >> 8,dicfile);fputc(hist[3].code % 256,dicfile);
      fputc(hist[4].code >> 8,dicfile);fputc(hist[4].code % 256,dicfile);
      fputc(count >> 8,dicfile);       fputc(count % 256,dicfile);
      fsize++;
      }
    }
fclose(dicfile);

dicfile = fopen(dicname,"r+b");
fseek(dicfile,0x00,SEEK_SET);
fsl = getc(dicfile);
fsh = getc(dicfile);
fseek(dicfile,0x00,SEEK_SET);
fsize +=fsh*256+fsl;
fputc(fsize % 256,dicfile);fputc(fsize >> 8,dicfile);
fclose(dicfile);
}
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
*-------------------------------------------------------------------*/

oid Searchdic(void)

 char input[30];

 clearword(33,19,77,22);
 gotoxy(33,19);printf("โปรดใส่ชื่อแฟ้มพจนานุกรม :");gets(input);
 if (input !=NULL)
  {
    sprintf(dicname,"%s",input);
    printstring(50,12,20,0x0f,"                    ");
    printstring(50,12,strlen(dicname),0x0f,dicname);
    gotoxy(33,20);printf("โปรดใส่ชื่อแฟ้มข้อมูลที่ต้องการทดสอบ :");
    gotoxy(33,21);gets(fname);
    printstring(50,14,20,0x0f,"                    ");
    gotoxy(50,14);
    fprintf(stdout,"%s",fname);
    sprintf(diccall,"%s",fname);
    sprintf(fname,"%s%s",letterpath,diccall);
       check = doprocess1();
       fclose(dicfile);
 }


 ==================================================================*/
 in()

 sprintf(letterpath,"");
 listcount = 0;
 display();
 while(listcount < 10)
 {
 displaycommand();
 check = getch();
 if (check == 0x0)
   {
     check = getch();
     if (check == 72) listcount--;
     if (check == 80) listcount++;
     if (listcount < 0) listcount = 4;
     if (listcount > 4) listcount = 0;
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
  }
if (check == 0x0d)
  {
   printstring(3,14+listcount,27,0x0f,comlist[listcount]);
   switch(listcount)
     {
       case 0 : Createdic(0);break;
       case 1 : Createdic(1);break;
       case 2 : Searchdic();break;
       case 3 : {
                  clearword(33,19,77,22);
                  gotoxy(33,20);printf("ป้อนชื่อ, PATH ของพจนานุกรม :");
                  gets(diccall);
                  if (strlen(diccall) != 0)
                  {
                    sprintf(letterpath,"%s",diccall);
                    gotoxy(53,13);printf("                    ");
                    gotoxy(53,13);printf("%s",letterpath);
                  }
                }
                break;
       case 4 : {
                  clearword(33,19,77,22);
                  gotoxy(33,20);printf("ต้องการออกจากโปรแกรม?(y/n)");
                  check = getch();
                  if ((check == 'y') || (check == 'Y'))
                    {
                      clrscr();
                      exit(0);
                    }
                }
     }
  }
clearword(33,19,77,22);
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้