บทที่ 2

ผลงานที่ผ่านมาและทฤษฎี

ได้มีการศึกษาค้นคว้าเทคนิคทางคณิตศาสตร์ออกมามากมาย [12]-[15] เพื่อใช้ในการแก้ปัญหาการรู้จำแบบรูปต่าง ๆ ซึ่งพอจะสามารถแบ่งแยกออกได้เป็น 2 แนวทาง คือ

1. Statistical approach (หรือ decision - theoretical approach) แนวทางนี้จะอาศัยหลักเกณฑ์ทางสถิติในการจำแนกแบบรูป โดยมีเซ็ตของการวัดทางคุณลักษณะซึ่งเรียกว่า feature ที่ดึงออกมาจากข้อมูลภาพ และถูกใช้ในการกำหนดแต่ละ feature vector ไปยัง class ใด class หนึ่งที่จัดไว้ ซึ่งคณิตศาสตร์ที่นำมาใช้ได้แก่ สมการการตัดสินใจแบบเชิงเส้น (linear decision function), การวัดการกระจายโดยการใช้ Karhunen - Loeve expansion, และการคำนวณคำลักษณะต่อเนื่องของจุดภาพ (connectivity) เป็นต้น

2. Syntactic approach (หรือ Structural approach) แนวทางนี้จะนำเอาข้อมูลเกี่ยวกับโครงร่างของรูปมาใช้ประโยชน์ การอธิบายรูปจะทำโดยอาศัยรูปย่อย (subpattern) ที่ง่ายแก่การจำซึ่งจะเรียกว่า primitive แต่ละรูปจะเกิดจากการประกอบกันของ primitive เหล่านี้ และกฎเกณฑ์ในการประกอบกันของ primitive เป็นรูปจะกำหนดโดยไวยากรณ์ (grammar) ตารางแสดงการเปรียบเทียบระหว่าง Statistical approach กับ Syntactic approach จะเป็นดังตารางที่ 2.1

สำหรับในประเทศไทยก็มีการทำวิจัยในเรื่องการรู้จำตัวอักษรไทยโดยนักวิจัยหลายท่านแต่ก็ไม่ได้แพร่หลายนัก ซึ่งแต่ละท่านก็ได้นำเสนอแนวทางและกรรมวิธีในการรู้จำแบบตัวอักษรที่แตกต่างกันไป อาทิเช่น

สุรสิทธิ์ ราตรี (2530) [7] ได้นำเสนอวิธีการรู้จำตัวอักษรไทยแบบตัวพิมพ์โดยวิธีวิเคราะห์โครงสร้างแบบมีไวยากรณ์ ทำโดยการแบ่งพื้นที่ของแต่ละตัวอักษรออกเป็นสี่ส่วน แล้วนำคณิตศาสตร์ทางสถิติมาตรวจนับเส้นโค้ง เส้นตรง และวงกลมในแต่ละส่วนพื้นที่ แล้วนำคณิตศาสตร์ที่คำนวณได้มาประกอบกันเพื่อแสดงความหมายว่าเป็นเส้นโค้ง เส้นตรง วงกลม หรือลักษณะพื้นฐานที่เกิดขึ้นในแต่ละส่วนของพื้นที่ แล้วนำลักษณะพื้นฐานเหล่านี้ใน

ตารางที่ 2.1 การเปรียบเทียบระหว่าง Statistical approach กับ Syntactic approach

		Statistical Approach	Syntactic Approach
1.	Pattern Generation (Storing) Basis	Probabilistic Models	Formal Grammars
2.	Pattern Classification (Recognition/Description) Basis	Estimation/Decision Theory	Parsing
3.	Feature Organization	Feature Vector	Primitives and Observed Relations
4.	Typical Learning (Training) Approaches		
	Supervised:	Density/distribution estimation (usually parametric)	Forming Grammars (heuristic or grammatical inference)
	Unsupervised:	Clustering	Clustering
5.	Limitations	Difficulty in expressing structural information	Difficulty in learning structural rules

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

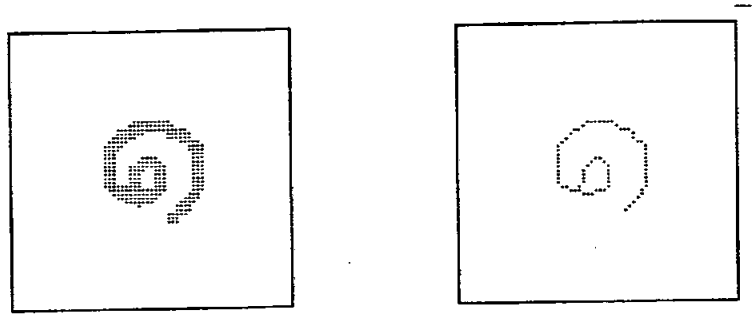
ทุกส่วนของพื้นที่มาประกอบกันเป็นคุณสมบัติของตัวอักษรต่าง ๆ โดยกำหนดขนาดภาพข้อมูลที่ใช้เป็น 20 x 45 จุดภาพ ซึ่งจากการทดสอบให้ผลการรู้จำประมาณ 98 % วิธีการที่นำเสนอนี้จะมีข้อจำกัดที่ว่าถ้ามีการเปลี่ยนแปลงขนาดของข้อมูลภาพและ font ที่นำมาใช้ จะมีผลให้ต้องทำการเปลี่ยนแปลงค่าที่ใช้เป็นเงื่อนไขในการกำหนดรหัสสัญลักษณ์ในแต่ละส่วนของพื้นที่

วัชระ ฉัตรวิริยะ (2530) [8] นำเสนอวิธีการจดจำตัวอักษรไทยแบบตัวพิมพ์ โดยใช้วิธีการหาคุณสมบัติของตัวอักษร โดยเลือกใช้คุณสมบัติของตำแหน่งของจุดศูนย์ถ่วงของจุดที่ประกอบเป็นตัวอักษรที่เทียบกับกรอบอ้างอิงที่กำหนด ตำแหน่งนี้จะเปลี่ยนไปตามรูปร่างของตัวอักษรหรือลักษณะการกระจาย และ จำนวนจุดของเนื้อตัวอักษร โดยกำหนดขนาดข้อมูลภาพที่ใช้เป็น 96 x 96 จุดภาพ ซึ่งข้อจำกัดของวิธีการหาตำแหน่งในการนำไปใช้แยกแยะตัวอักษร ก็คือ สามารถใช้ได้กับตัวอักษรที่เป็นตัวพิมพ์ และ ตารางที่เก็บค่าของพารามิเตอร์จะใช้เฉพาะตัวอักษรเฉพาะเท่านั้น ถ้าเป็นตัวอักษรแบบอื่นก็ต้องสร้างตารางใหม่ด้วย

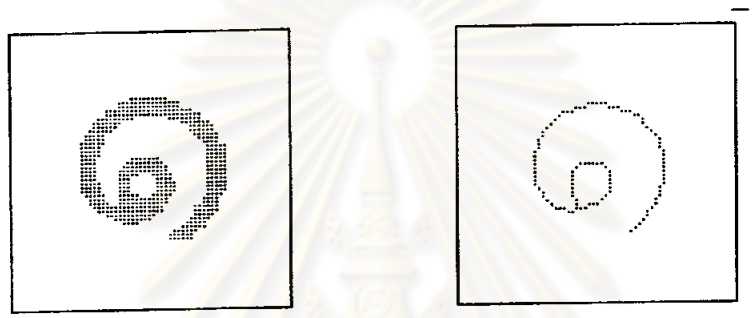
สุรพันธุ์ เอื้อไพบูลย์ (2531) [9] นำเสนอการรู้จำตัวอักษรไทยแบบตัวเขียน โดยการนำเอาลักษณะเด่นของตัวอักษรไทย คือ หัวของตัวอักษร มาใช้ในการจำแนกกลุ่มของตัวอักษร (rough classification) และใช้คุณสมบัติทางโทโปโลยี่ (topology) ของตัวอักษร ในการแยกตัวอักษรออกจากกลุ่ม (fine classification) ตลอดจนนำเอาเทคนิคในการแยกตัวอักษรหลาย ๆ วิธีมาใช้แยกตัวอักษรที่มีความคล้ายคลึงกันออกจากกันด้วย ซึ่งปรากฏว่าจากการทดสอบโดยกำหนดขนาดข้อมูลภาพเป็น 64 x 64 จุดภาพ สามารถให้ผลการรู้จำประมาณ 98.82 % โดยข้อจำกัดของวิธีการนี้คือตัวอักษรไทยที่ใช้จะต้องเป็นตัวอักษรที่เขียนเลียนแบบตัวพิมพ์ และ ต้องมีหัวที่ชัดเจนด้วย

ชัยณรงค์ คล้ายมณี และ ทรงชัย วีระทวีมาศ (2533) [10] นำเสนอการรู้จำตัวเลขไทยแบบตัวเขียน โดยใช้คุณสมบัติทางโทโปโลยี่ ซึ่งได้แก่ จุดตัด, จุดปลาย, จุดแยก, จุดต่อ, และจุดอิสระ วิธีการนี้สามารถให้ผลการรู้จำประมาณ 80 %

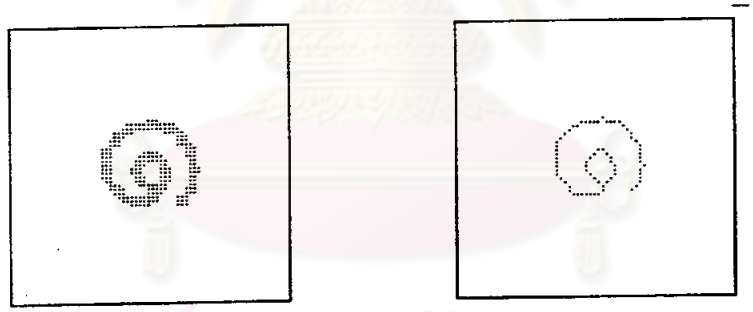
จากตัวอย่างงานวิจัยที่ยกขึ้นมา จะเห็นว่าวิธีการที่ใช้ส่วนใหญ่จะเป็นไปในแนวทาง Statistical ซึ่งมีข้อจำกัดอยู่ที่ต้องกำหนดขนาดของข้อมูลภาพที่แน่นอนและจะใช้งานได้เฉพาะรูปแบบตัวอักษรที่กำหนดเท่านั้น ดังเช่นในงานวิจัย 3 ชิ้นแรก ส่วนในงานวิจัยสุดท้ายนั้นผลการรู้จำยังไม่สูงพอที่จะนำไปพัฒนาออกใช้งานจริงได้



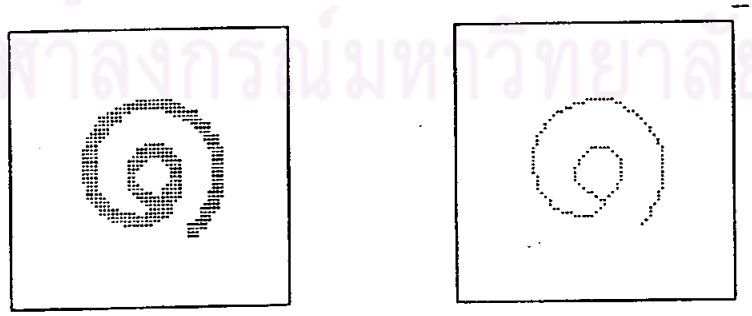
(ก) CordiaUPC 18



(ข) CordiaUPC 26



(ค) DilleniaUPC 18



(ง) DilleniaUPC 26

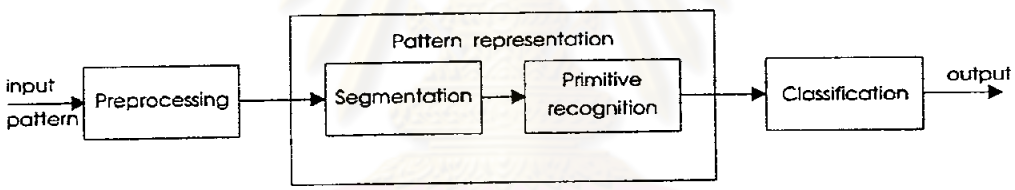
รูปที่ 2.1 ตัวอย่างตัวเลข ๑ ที่ font และขนาดต่างกัน กับโครงร่างของตัวเลขนั้น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สำหรับวิธีการที่นำเสนอในงานวิจัยนี้ คือ วิธีซินแทกติก จะมีจุดเด่นอยู่ที่นำเอาโครงร่างของภาพตัวอักษรมาใช้ประโยชน์ในการรู้จำตัวอักษรนั้น ซึ่งโดยปกติแล้วตัวอักษรแต่ละตัวในแบบ font ต่างกัน หรือ ขนาดต่างกัน จะยังคงมีส่วนที่เป็นโครงร่างที่คล้ายคลึงกัน ดังแสดงในรูปที่ 2.1 จึงน่าที่แนวทางที่นำเสนอนี้จะสามารถให้ผลในการรู้จำภาพตัวเลขไทยในแบบ font ต่างกัน ที่ขนาดต่าง ๆ ได้ดี

2.1 โมเดลของระบบรู้จำด้วยวิธีซินแทกติก

บล็อกไดอะแกรมของระบบรู้จำด้วยวิธีซินแทกติกจะเป็นดังรูปที่ 2.2 ซึ่งจะประกอบไปด้วยส่วนต่าง ๆ คือ Preprocessing, Pattern representation, และClassification โดยในส่วน Pattern representation จะประกอบด้วย 2 ส่วนย่อย คือ Segmentation (หรือ Decomposition) กับ Primitive recognition



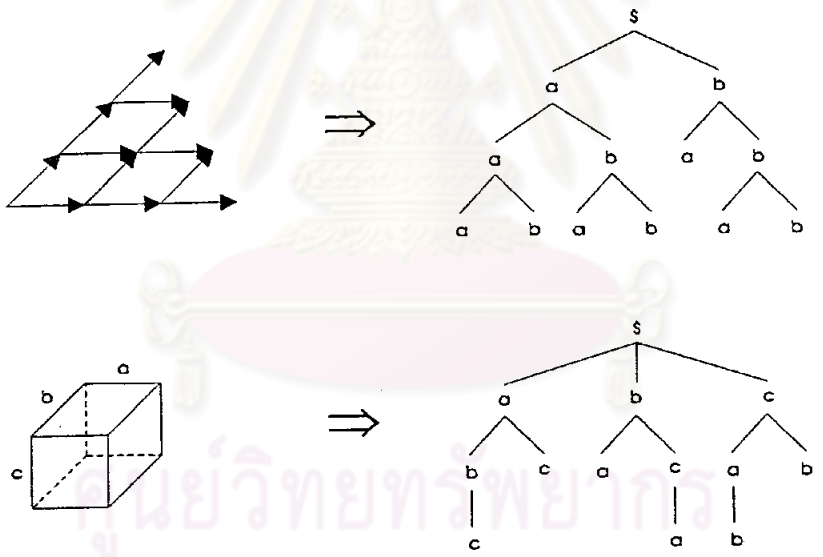
รูปที่ 2.2 บล็อกไดอะแกรมของระบบรู้จำด้วยวิธีซินแทกติก

หลังจากที่ได้รับอินพุตของรูปเข้ามาก็จะผ่านไปที่ส่วนแรกคือ Preprocessing ซึ่งจะมีหน้าที่ในการปรับแต่งรูปให้มีคุณภาพที่ดีขึ้นและมีความเหมาะสมที่จะนำไปใช้ในส่วนถัดไป กระบวนการในส่วนนี้ได้แก่ การกำจัดสัญญาณรบกวน, การปรับรูปให้มีขนาดมาตรฐาน, และการทำให้บาง เป็นต้น ในส่วนที่สอง Pattern representation จะเป็นส่วนที่ทำการแทนรูปด้วยโครงสร้างของภาษาซึ่งกำหนดโดยชนิดของไวยากรณ์ที่เลือกใช้ ในส่วนนี้จะประกอบไปด้วย 2 กระบวนการคือ Segmentation และ Primitive recognition ส่วน Segmentation จะเป็นส่วนที่ทำการแบ่งแยกโครงร่างของรูปออกเป็นส่วนย่อย ๆ โดยจัดเก็บตามรูปแบบของภาษาที่ใช้ จากนั้นส่วน Primitive recognition จะทำการแทนแต่ละส่วนย่อยด้วย primitive ที่ได้กำหนดไว้ ส่วนสุดท้าย Classification จะเป็นส่วนที่ทำการจำแนกรูปแต่ละรูปซึ่งในการ

ศึกษานี้จะใช้เทคนิคในการหา Distance ระหว่าง อินพุต กับ ต้นแบบ โดยรายละเอียดของวิธีการจะกล่าวในบทต่อไป

2.2 ไวยากรณ์แบบต้นไม้ (Tree Grammar) [14]

ได้มีการศึกษาวิธีการแทนโครงสร้างของภาษาด้วยหลักทางคณิตศาสตร์เกิดเป็นไวยากรณ์ (grammar) แบบต่าง ๆ ออกมามากมาย อาทิเช่น context-free grammar, context-sensitive grammar, finite-state grammar, web grammar, และ tree grammar เป็นต้น โดยไวยากรณ์ที่ยกตัวอย่างมา 3 อย่างแรกนั้นการอ้างความสัมพันธ์ระหว่าง primitive แต่ละตัวที่เชื่อมต่อกันจะเป็นไปในลักษณะซ้ายและขวาเท่านั้น ซึ่งในการใช้งานกับแบบรูปที่มีโครงร่างซับซ้อน คือ มิได้เป็นเพียงลายเส้นเดี่ยว ๆ แต่มีการแตกแยกออกเป็นหลาย ๆ เส้น จะทำการอธิบายโครงร่างนั้นทำได้ยาก แต่สำหรับไวยากรณ์ 2 อย่างหลังจะสามารถใช้ในการ



(ก) รูป (ข) tree

รูปที่ 2.3 ตัวอย่างของรูปและ tree ของรูปที่สอดคล้องกัน

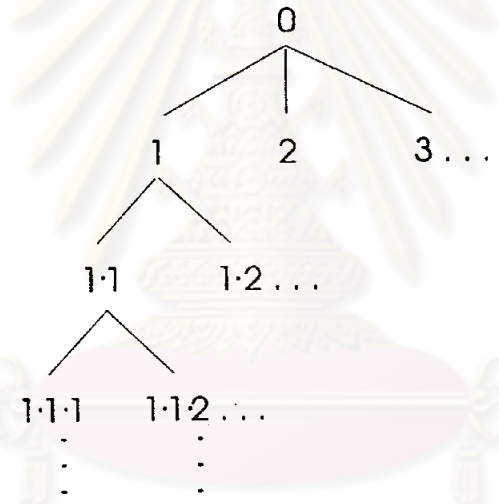
อธิบายโครงร่างที่มีความซับซ้อนได้ดีกว่า ในการศึกษานี้จะเลือกใช้ tree grammar ซึ่งเราจะสามารถใช้ tree grammar ในการสร้างรูปได้โดยง่าย เพียงแต่รูปนั้นจะต้องสามารถอธิบายได้ด้วยโครงสร้างแบบ tree ดังรูปที่ 2.3 จะแสดงรูป และ โครงสร้างแบบ tree ของรูปนั้น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

กำหนดให้ N^+ เป็นเซ็ตของจำนวนเต็มบวก และให้ U เป็น Universal tree domain มีลักษณะดังรูปที่ 2.4 (โดยมีจุดเริ่มต้นที่ 0 และกำหนด "·" เป็นเครื่องหมายการกระทำการแบบไบนารี่) ความลึกของ $a \in U$ จะเขียนด้วย d(a) และกำหนดให้เป็นดังนี้ คือ $d(0) = 0$, $d(a \cdot i) = d(a) + 1$ โดยที่ $i \in N^+$ $a \leq b$ ถ้าและเพียงแต่ถ้ามี $x \in U$ ที่ $a \cdot x = b$ a และ b จะไม่สามารถเปรียบเทียบกันได้ถ้าและเพียงแต่ถ้า a ไม่น้อยกว่าหรือเท่ากับ b และ b ไม่น้อยกว่าหรือเท่ากับ a D จะเป็น tree domain ถ้าและเพียงแต่ถ้า D เป็นเซ็ตย่อยที่จำกัดของ U ที่เป็นไปตาม

(ก) ถ้า $b \in D$ และ $a < b$ แล้วจะถือได้ว่า $a \in D$

(ข) ถ้า $a \cdot j \in D$ และ $i < j$ ใน N^+ แล้วจะถือได้ว่า $a \cdot i \in D$



รูปที่ 2.4 Universal tree domain

Ranked alphabet จะเป็นคู่ลำดับ $< \Sigma , r >$ โดยที่ Σ เป็นเซ็ตที่จำกัดของสัญลักษณ์ และ

$$r : \Sigma \rightarrow N = N^+ \cup \{ 0 \}$$

สำหรับ $a \in \Sigma$, r(a) จะถูกเรียกว่า rank ของ a และให้ $\Sigma_n = r^{-1}(n)$

Tree บน Σ (บน $< \Sigma, r >$) จะเป็นฟังก์ชัน

$$\alpha : D \rightarrow \Sigma$$

โดยที่ D เป็น tree domain และ

$$r[\alpha(a)] = \max \{ i \mid a \cdot i \in D \}$$

นั้นคือ rank ของที่ตำแหน่ง a จะต้องเท่ากับจำนวนของสาขาใน tree domain ที่ a และ domain ของ tree จะเขียนด้วย $D(\alpha)$ หรือ D_α กำหนดให้ T_Σ เป็นเซ็ตของ tree บน Σ ทั้งหมด ความลึกของ α จะกำหนดเป็น $d(\alpha) = \max \{ d(a) \mid a \in D(\alpha) \}$ ถ้าให้ α เป็น tree และ a เป็นสมาชิกของ $D(\alpha)$ แล้ว α/a จะหมายถึง subtree ของ α ที่ a ซึ่งจะกำหนดเป็น

$$\alpha/a = \{ (b,x) \mid (a \cdot b,x) \in \alpha \}$$

นิยาม

Regular tree grammar บน $< V_T , r >$ จะเป็นดังนี้

$$G_t = (V , r' , P , S)$$

ซึ่งจะสอดคล้องตามเงื่อนไขต่อไปนี้

1. $< V , r' >$ เป็น ranked alphabet ที่จำกัด ที่มี $V_T \subset V$ และ $r' / V_T = r$

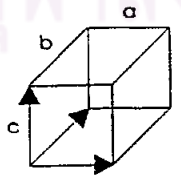
   $V - V_T = V_N$ โดย V_N เป็นเซ็ตของ nonterminals.

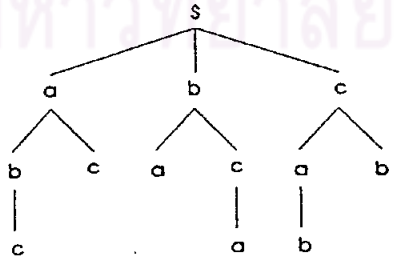
2. P เป็นเซ็ตที่จำกัดของ productions ที่มีรูปแบบเป็น $\phi \rightarrow \psi$ โดยที่ ϕ และ ψ เป็น tree บน $< V , r' >$

3. S เป็น subset ที่จำกัดของ T_v โดยที่ T_v เป็นเซ็ตของ trees บนอักขระ V

ตัวอย่างที่ 2.1 วัตถุรูปสี่เหลี่ยมลูกบาศก์ดังรูป 2.5(ก) สามารถอธิบายได้โดยโครงสร้างแบบ tree ในรูป 2.5(ข)

primitives :





(ก) (ข)

รูปที่ 2.5 การแทน tree ของวัตถุรูปลูกบาศก์

ตัวอย่างที่ 2.2 Tree grammar

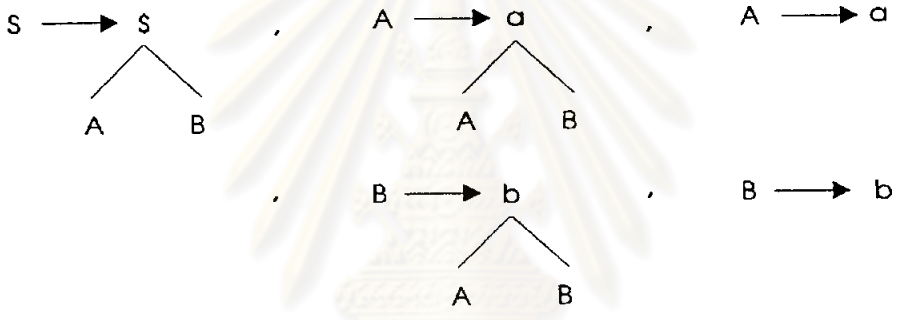
$$G_t = (V , r , P , S)$$

โดยที่

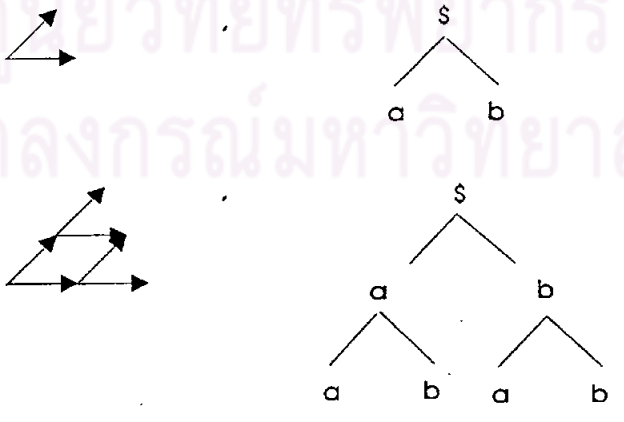
$$V = \{ S , a , b , \$, A , B \} , \quad V_T = \{ a , b , s \}$$

$$r(a) = r(b) = \{ 2,0 \} , \quad r(\$) = 2$$

และ P :



จะทำให้เกิดรูปดังเช่น



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

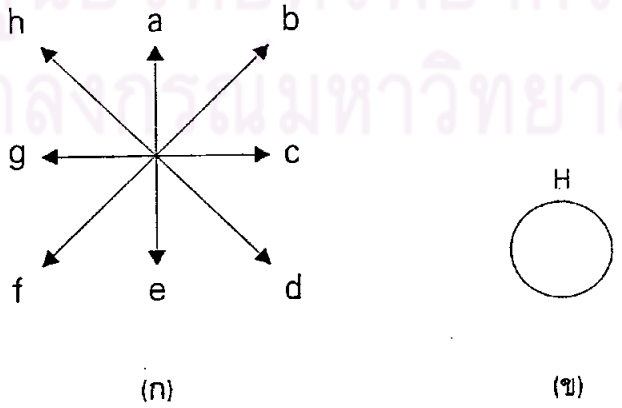
2.3 คุณสมบัติของตัวเลขไทย และ primitive

เมื่อพิจารณาดูลักษณะของตัวเลขไทยแบบตัวพิมพ์ทั้ง 10 ตัว (จากเลข ๐ ถึง ๙) ดังรูป 2.6 จะเห็นว่าโครงสร้างของตัวเลขไทยจะประกอบด้วยส่วนของเส้นโค้ง (curve) และส่วนที่เป็นลูป (loop) เป็นหลัก ซึ่งในการพิจารณาหา primitive ที่จะใช้ในส่วนที่เป็นเส้นโค้ง ถ้าเราเลือกใช้ primitive ที่มีลักษณะเป็นเส้นโค้งในแบบต่าง ๆ การแม็ช (matching) ของรูปย่อยที่เป็นเส้นโค้งกับ primitive ที่เลือกไว้จะต้องมีกระบวนการที่ค่อนข้างยุ่งยากและเสียเวลา

๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙

รูปที่ 2.6 ตัวเลขไทยตั้งแต่ ๐ ถึง ๙

คือ จะต้องทำการตรวจสอบหารัศมีความโค้งของเส้นโค้ง และ ลักษณะของการโค้งของเส้นโค้งนั้น [16] เพื่อใช้ในการเปรียบเทียบกับ primitive ที่เป็นต้นแบบ ซึ่งโดยการที่เราทำการแบ่งส่วนของเส้นโค้งออกเป็นส่วนของเส้นตรงย่อยหรือเวคเตอร์ (vector) ในทิศทางต่าง ๆ ที่มาประกอบกัน [11] จะทำให้เราสามารถได้ primitive ที่ง่ายแก่การแม็ช คือ ทำการตรวจสอบเฉพาะทิศทางของแต่ละเวคเตอร์เท่านั้น ซึ่งชุดของ primitive ที่สอดคล้องกับส่วนย่อยดังกล่าว ก็คือ Freeman Vector ดังแสดงในรูปที่ 2.7(ก) สำหรับในส่วนที่เป็นลูปเราจะเลือกใช้ primitive ดังรูป 2.7(ข) ซึ่งในแต่ละ primitive จะทำการเขียนแทนด้วยตัวอักษรตามที่ได้เขียนกำกับไว้ในรูป



(ก) (ข)

รูปที่ 2.7 primitives (ก) Freeman vector (ข) loop

ซึ่งจากชุดของ primitive ที่กำหนดดังรูปที่ 2.7 นี้เราจะนำไปใช้งานในกระบวนการรู้จำตัวเลขไทยที่จะกล่าวถึงรายละเอียดในบทถัดไป โดยกระบวนการที่ทำการแทน primitive เหล่านี้จะปรากฏอยู่ในหัวข้อที่ 3.6

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย