



บทที่ ๓

ระบบการจัดตัวอักษรด้วยคอมพิวเตอร์

ปัจจุบันระบบการสื่อสารในประเทศไทยได้ใช้ภาษาไทยปนภาษาอังกฤษ ทั้งนี้เนื่องจากประเทศไทยได้รับเอาวิชาการและเทคโนโลยีมาจากประเทศทางตะวันตก ด้วยเหตุนี้การประมวลผลด้วยคำเพื่อทำการจัดพิมพ์จึงจำเป็นต้องสามารถประมวลผลได้ทั้งภาษาไทยและภาษาอังกฤษ ปัญหาที่สำคัญประการหนึ่งที่เกิดขึ้นในการประมวลผลด้วยอักษรไทยก็คือ ระบบคอมพิวเตอร์ที่ใช้กันอยู่ในปัจจุบันได้ออกแบบมาจากประเทศตะวันตก ซึ่งใช้ภาษาที่มีโครงสร้างตัวอักษรอยู่ในบรรทัดเดียวกันหรือ ๑ ระดับ เป็นข้อมูลเพื่อป้อนเข้าสู่ระบบคอมพิวเตอร์ (Input string) โดยมีลักษณะเป็นตัวอักษรที่เรียงต่อเนื่องกันและใช้แสดงผลลัพท์ ทำให้ระบบการประมวลผลด้วยคำที่ได้พัฒนาขึ้นในตะวันตกไม่สามารถประยุกต์ได้โดยตรงกับการประมวลผลด้วยอักษรไทย เนื่องด้วยโครงสร้างตัวอักษรในภาษาไทยแบ่งออกเป็น ๔ ระดับ โดยที่ตำแหน่งหนึ่ง ๆ ของภาษาไทยสามารถมีอักษรได้มากกว่า ๑ ตัว ดังนั้นเพื่อที่จะให้ภาษาไทยสามารถใช้กับระบบคอมพิวเตอร์ได้ อักษรของภาษาไทยจึงต้องเก็บอยู่ในลักษณะที่อยู่ในบรรทัดเดียวกัน เช่นเดียวกับภาษาอังกฤษ เช่น คำว่า "สื่อสาร" จะต้องถูกเก็บไว้ในลักษณะ "ส ๑ ๐ อ ส ๑ ร" ฉะนั้นจึงต้องจัดอักษรในภาษาไทยจาก ๔ ระดับ มาเป็น ๑ ระดับ เพื่อเป็นข้อมูลสำหรับป้อนเข้าสู่ระบบคอมพิวเตอร์และในการพิมพ์ภาษาไทยจึงต้องมีการจัดให้อักษรในภาษาไทยกลับมาเป็น ๔ ระดับอย่างเดิม เพื่อพิมพ์ออกมาอยู่ในรูปที่อ่านแล้วเข้าใจได้โดยง่าย

โครงสร้างตัวอักษรในภาษาไทย

ลักษณะโครงสร้างของภาษาไทย ประกอบด้วย

๑. พยัญชนะ ก-ฮ ทั้งหมด ๔๔ ตัว และ ฤ ฦ
๒. สระ ะ ั ำ ิ ึ ื ุ ู เ แ โ ใ ใ และ ๆ ำ
๓. วรรณยุกต์ ไม้ ่ ้ ๊ ๋ และ ์ (หยาดน้ำค้าง)

คำในภาษาไทยคำหนึ่ง ๆ อาจจะประกอบด้วย พยางค์ สระ วรรณยุกต์ต่าง ๆ ซึ่งอยู่ในบรรทัดเดียวกันหรืออยู่คนละบรรทัด โดยที่สระอาจจะมีทั้งอยู่ในเหนือบรรทัด หรือใต้บรรทัดของพยางค์ และยังมีวรรณยุกต์อยู่อีกหนึ่งบรรทัด ดังนั้นจึงพอจะสรุปได้ว่าลักษณะโครงสร้างของตัวอักษรในภาษาไทยแบ่งออกเป็น ๔ ระดับ ดังนี้

ระดับที่ ๑ (TYPE1) เป็นระดับของตัวอักษรประเภทวรรณยุกต์และตัวการ์นต์ มีทั้งหมด ๔ ตัว ได้แก่ ' ° ° ° °

ระดับที่ ๒ (TYPE2) เป็นระดับของตัวอักษรประเภทสระบางประเภท รวมทั้งไม้ไต่คู้ และหยาดน้ำค้างมีทั้งหมด ๗ ตัว ได้แก่ ° ° ° ° ° ° ° (หยาดน้ำค้าง)

ระดับที่ ๓ (TYPE3) เป็นระดับที่สำคัญของภาษาไทย มีอักขระมากที่สุด มีทั้งหมด ๔๖ ตัว เป็นพยางค์ ๔๖ ตัว (ก-ช และ ฤ ฦ) และสระอีก ๑๐ ตัว (รวมทั้ง ๆ ๆ) ได้แก่ ำ แ อ โ อ ะ ำ ๆ และ ๆ

ระดับที่ ๔ (TYPE4) เป็นระดับของตัวอักษรประเภทสระบางประเภทมีทั้งหมด ๒ ตัว ได้แก่ ุ ู

รวมตัวอักษรในภาษาไทยทั้งหมดมี ๗๐ ตัว

การจัดเรียงลำดับตัวอักษรในภาษาไทย เพื่อป้อนเข้าสู่ระบบคอมพิวเตอร์

ดังที่กล่าวมาแล้วว่า ระบบคอมพิวเตอร์ได้ออกแบบมาเพื่อใช้กับภาษาที่มีโครงสร้างของตัวอักษรอยู่ในบรรทัดเดียวกัน หรือ ๑ ระดับ ดังนั้นข้อมูลที่เป็นอักขระในภาษาไทยจึงจำเป็นต้องจัดลำดับหรือตำแหน่งของอักขระก่อน เพื่อให้อยู่ในรูปที่ระบบคอมพิวเตอร์สามารถอ่านได้ เช่นคำว่า "สื่อสาร" จะต้องจัดตัวอักษรให้อยู่ในบรรทัดเดียวกันและเรียงลำดับหรือตำแหน่งของตัวอักษรดังนี้ "ส อสื่อสาร" หรือ "ส อสื่อสาร" เป็นต้น

จากกฎเกณฑ์นี้พอจะสรุปได้ว่า ถ้าตำแหน่งใดมีตัวอักขระมากกว่าหนึ่งตัว ตัวอักษรประเภทพยางค์จะนำหน้าตัวอักษรอื่น ๆ แล้วจึงตามด้วยอักขระประเภทอื่น ตัวอย่างเช่นคำว่า "อู" อาจเป็น "ค . °" หรือ "ค . °" ก็ได้ หรือคำว่า "ค" อาจเป็น "ค . °" หรือ "ค . °" ก็ได้ การสลับตำแหน่งของตัวอักษรในระดับ ๑, ๒ และ ๔ ไม่มีผลต่อการเรียงพิมพ์

การเรียงพิมพ์ภาษาไทย

เนื่องด้วยข้อมูลที่เป็นอักขระไทยที่จะเข้าระบบคอมพิวเตอร์ ต้องจัดตัวอักขระให้อยู่ในบรรทัดเดียวกันและจัดลำดับหรือตำแหน่งตัวอักขระก่อน เช่นคำว่า "สื่อสาร" อาจถูกจัดเป็น "ส" "อ" "สาร" หรือ "ส" "อ" "สาร" ดังที่กล่าวมาแล้ว ในการพิมพ์ภาษาไทยจึงต้องมีการจัดเรียงพิมพ์ เพื่อจัดระดับและตำแหน่งของตัวอักขระให้กลับไปอยู่ในลักษณะ ๔ ระดับ ตามเดิม ดังนั้นการเรียงพิมพ์ภาษาไทยจึงต้องประกอบด้วยการจัดประเภทของอักขระว่าอยู่ในระดับใด ตำแหน่งการพิมพ์จะเคลื่อนไปทางขวาหนึ่งตำแหน่ง เมื่ออักขระที่พบเป็นประเภทร้อยอยู่ในระดับที่ ๓ เท่านั้น เนื่องจากการสลับตำแหน่งของตัวอักขระในระดับ ๑, ๒ และ ๔ ไม่มีผลต่อการจัดเรียงพิมพ์ (๕)

สำหรับรหัสที่ใช้แทนตัวอักขระภาษาไทย ภาษาอังกฤษ ตัวเลขและอักขระพิเศษ ในระบบรหัส EBCDIC (Extended Binary Coded Decimal Interchange Code) มีรหัสดังรูป ๓.๑

