

ประกาศกระทรวงอุตสาหกรรม

ฉบับที่ ๒๕๘๕ (พ.ศ. ๒๕๕๒)

ออกตามความในพระราชบัญญัติมาตรฐานผลิตภัณฑ์อุตสาหกรรม

พ.ศ. ๒๕๑๑

เรื่อง กำหนดมาตรฐานผลิตภัณฑ์อุตสาหกรรม

อักษรวิธียาไทยสำหรับคอมพิวเตอร์

อาศัยอำนาจตามความในมาตรา ๑๕ แห่งพระราชบัญญัติมาตรฐานผลิตภัณฑ์
อุตสาหกรรม พ.ศ. ๒๕๑๑ รัฐมนตรีว่าการกระทรวงอุตสาหกรรมออกประกาศ
กำหนดมาตรฐานผลิตภัณฑ์อุตสาหกรรม อักษรวิธียาไทยสำหรับคอมพิวเตอร์
มาตรฐานเลขที่ มอก. ๑๕๖๖-๒๕๕๑ ไว้ ดังมีรายการละเอียดต่อท้ายประกาศนี้

ประกาศ ณ วันที่ ๑๗ มิถุนายน พ.ศ. ๒๕๕๒

สุวัจน์ ลิปตพัลลภ

รัฐมนตรีว่าการกระทรวงอุตสาหกรรม

มาตรฐานผลิตภัณฑ์อุตสาหกรรม อักษรวิธีกาษาไทยสำหรับคอมพิวเตอร์

1. ขอบข่าย

- 1.1 มาตรฐานผลิตภัณฑ์อุตสาหกรรมนี้ครอบคลุมถึง การจำแนกประเภทของสัญลักษณ์ต่าง ๆ ในภาษาไทยเมื่อพิจารณาตามความต้องการของการประมวลผลด้วยคอมพิวเตอร์และการกำหนดพฤติกรรมของโปรแกรมระบบและโปรแกรมประยุกต์ เมื่อทำงานเกี่ยวข้องกับชุดอักขระภาษาไทย นอกจากนี้ได้กำหนดพฤติกรรมร่วมเมื่อมีการใช้งานเป็นภาษาอังกฤษปนกับภาษาไทย พฤติกรรมดังกล่าวครอบคลุมถึงวิธีการป้อนข้อมูล (input method) และวิธีการแสดงผล (output method) ของโปรแกรมประยุกต์ต่าง ๆ ผ่านระบบภาษาไทยใด ๆ ที่มีคุณสมบัติตามมาตรฐานนี้ กล่าวคือ โปรแกรมประยุกต์ต่าง ๆ ภายใต้มาตรฐานนี้เมื่อได้รับสายอักขระ (string) ชุดหนึ่งจะต้องมีพฤติกรรมอย่างเดียวกันไม่ว่าจะทำงานภายใต้ระบบปฏิบัติการหรือเครื่องคอมพิวเตอร์เครื่องใด ๆ ในกรณีพิเศษซึ่งกล่าวเฉพาะการจัดทำแบบใดแบบหนึ่ง จะมีหมายเหตุเขียนไว้อย่างชัดเจน

2. บทนิยาม

ความหมายของคำที่ใช้ในมาตรฐานผลิตภัณฑ์อุตสาหกรรมนี้ มีดังต่อไปนี้

- 2.1 อักขระ (character) หมายถึง ตัวอักษร ตัวเลข เครื่องหมายพิเศษ และเครื่องหมายอื่นใด รวมทั้งอักขระควบคุมที่สามารถป้อนบันทึกลงสื่อและแสดงผลทางเครื่องคอมพิวเตอร์ได้
- 2.2 รหัส (code) หมายถึง ตัวเลขที่ใช้แทนความหมายต่าง ๆ ที่กำหนดขึ้น
- 2.3 รหัสอักขระ (character code) หมายถึง เลขจำนวนเต็มซึ่งใช้แทนอักขระต่าง ๆ ในชุดอักขระสำหรับรหัสชุดพื้นฐานตามมาตรฐานนี้ ใช้รหัสอักขระแบบ 8 บิต (เลขรหัสมีค่าตั้งแต่ 0 ถึง 255)
- 2.4 ชุดอักขระ (character set) หมายถึง กลุ่มอักขระเข้ารหัสสำหรับใช้ในระบคอมพิวเตอร์ ในมาตรฐานนี้ได้กำหนดชุดอักขระขึ้นมา 2 ชุด ได้แก่ ชุดอักขระพื้นฐาน (basic character set) และชุดอักขระ Code Page 437 แต่ละชุดมีอักขระ 256 แบบ ดังรายละเอียดในข้อ 4.
- 2.5 ตารางรหัส (code table) หมายถึง ตารางแสดงตำแหน่งของอักขระต่าง ๆ ในกรณีของรหัสชนิด 8 บิต ใช้ขนาด 16 x 16 ช่อง ซึ่งมีเส้นทางแนวตั้งระบุค่าของ 4 บิตท้าย และแนวอนระบุค่าของ 4 บิตต้นของรหัสภายในตารางเป็นการระบุความหมายของรหัสตรงตำแหน่งช่องนั้น ๆ หากค่าของรหัสเป็นตัวควบคุมในตารางรหัส จะแสดงค้าย่อของตัวควบคุม หากค่าของรหัสเป็นตัวอักษรที่เห็นเป็นรูปได้ (graphic character) จะแสดงเป็นรูปของอักขระนั้น ๆ
- 2.6 อักขระไทย (Thai character) หมายถึง อักขระจำนวน 83 ตัวตาม มอก.620 ซึ่งจำแนกเบื้องต้นตามมาตรฐานนั้นออกเป็นตัวอักษรไทย ตัวเลขไทย และเครื่องหมายพิเศษ และจำแนกโดยละเอียดตามมาตรฐานนี้ออกเป็นรายละเอียดต่าง ๆ ตามข้อ 4.
- 2.7 รหัสอักขระพื้นฐานตามมาตรฐาน (common character code) หมายถึง รหัสอักขระชุดพื้นฐานตามมาตรฐานนี้ ซึ่งมีชื่อย่อว่า TACTIS (Thai API Consortium/Thai Industrial Standard) ประกอบด้วยรหัสมาตรฐาน ISO

- 646 ร่วมกับ มอก.620 (ส่วนที่เป็นประเภท ASCII) ดังรายละเอียดในข้อ 4.
- 2.8 **อักษรวิธีภาษาไทย** (Thai input and output methods) คำว่า "อักษรวิธี" ตามพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ.2525ให้ความหมายไว้ว่า "วิธีเขียนและอ่านหนังสือให้อีกต้อง" แต่ในมาตรฐานนี้จะใช้ความหมายเฉพาะ ส่วนวิธีเขียนหนังสือไทยที่ถูกต้อง ซึ่งในด้านคอมพิวเตอร์นั้นจะหมายถึงวิธีเขียน (แสดงผล) ให้อีกต้อง ตรงกับภาษาอังกฤษว่า Thai output method โดยความถูกต้องในที่นี้หมายถึงพฤติกรรมใด ๆ ที่สอดคล้องกับมาตรฐานนี้
 - 2.9 **ข้อกำหนดเพื่อการเขียนโปรแกรมประยุกต์** (Application Programming Interface หรือ API) หมายถึง ข้อตกลงหรือการประกาศเป็นภาษาโปรแกรมภาษาใด ๆ (ในที่นี้ใช้ภาษา C) เพื่อให้ผู้พัฒนาโปรแกรมประยุกต์สามารถใช้งานระบบปฏิบัติการที่มีข้อกำหนดเดียวกันได้อย่างถูกต้อง ทั้งนี้ไม่ขึ้นอยู่กับตัวเครื่อง
 - 2.10 **การจำแนกประเภทของอักขระ** (character classification) หมายถึง การแบ่งกลุ่มสัญลักษณ์ต่าง ๆ ในรหัส TACTIS ให้เป็นหมวดหมู่ที่เป็นประโยชน์ต่อการนำไปประมวลผลในเชิงการกำหนดอักษรวีธีของแต่ละกลุ่มอย่างชัดเจน และการกำหนดวิธีการตรวจสอบความถูกต้องของลำดับตัวอักษรในสายอักขระและในการป้อนข้อมูล
 - 2.11 **ระดับการแสดงผลอักขระ** (character display level) หมายถึง ตำแหน่งการวางอักษระลงบนสื่อแสดงผล โดยเทียบกับแนวเส้นบรรทัดที่กำลังแสดง ระดับต่าง ๆ ในมาตรฐานนี้แบ่งเป็น 5 ระดับ คือ **ระดับไม่แสดงผล** (NONDISP) **ระดับบนสุด** (TOP) **ระดับเหนือฐาน** (ABOVE) **ระดับฐาน** (BASE) และ**ระดับต่ำกว่าฐาน** (BELOW)
 - 2.12 **ตัวหยุด หรือ ตัวหยุดอยู่กับที่** (dead character) หมายถึง สัญลักษณ์ประเภทสระบน สระล่าง วรรณยุกต์ และเครื่องหมายกำกับเสียงที่อยู่นหรือได้พยัญชนะต่าง ๆ สัญลักษณ์เหล่านี้โดยปกติแล้วจะต้องได้รับการป้อนตามหลังอักขระที่สามารถประสมกับตัวมันได้ และเมื่อแสดงผล ตัวชี้ตำแหน่ง (cursor) จะไม่เดินต่อไปทางขวามือ ค่าซึ่งตรงกันข้ามกับ ตัวหยุด คือ **ตัวเดินหน้า** (forward character)
 - 2.13 **เซลล์แสดงผลภาษาไทย** (Thai display cell) หมายถึง หน่วยแสดงผล 1 หน่วยที่อาจจะประกอบด้วยสัญลักษณ์ภาษาไทยจำนวนตั้งแต่ 1 ถึง 3 สัญลักษณ์ ประสมอยู่ในตำแหน่งแสดงผลเดียวกันแต่ต่างระดับตามอักษรวีธีที่กำหนดในข้อ 4. คุณสมบัติของเซลล์หรือช่องแสดงผลภาษาไทยจะถูกบรรยายโดยข้อมูลซึ่งมีโครงสร้างเป็น THAI_CELL ตามคำจำกัดความโดยละเอียดในข้ออื่น
 - 2.14 **การประสมอักขระ** (character composition) หมายถึง การนำอักขระที่สามารถประสมกันได้มารวมกันในเซลล์เดียวกัน เช่น พยัญชนะสามารถประสมกับสระบนได้โดยปรากฏอยู่ในเซลล์เดียวกัน เช่น
มี เกิดจาก ม มา ประสมกับ สระอี
ตุ้ เกิดจาก ต เต่า ประสมกับสระอุ และไม้โท
 - 2.15 **สัญลักษณ์ที่มีหลายอักขระ** (multi-character symbols) หมายถึง สัญลักษณ์ทางภาษาไทยซึ่งมีการใช้งานตามหลักภาษา โดยการนำอักขระมากกว่า 1 ตัวมารวมกันแสดงผล ตัวอย่างสัญลักษณ์ที่มีหลายอักขระ ได้แก่
ฯลฯ เรียกว่า ไปยาลใหญ่
... เรียกว่า ไช้ปลา
ฯะ เรียกว่า อังคัณวิสรรชนีย์

๗๘๘ เรียกว่า อังคณวิสรรชนีย์และโคมูตร
 :- เรียกว่า วิกิษภาค

- 2.16 **สายอักขระ** (character string) หมายถึง แถวลำดับ(array)ของอักขระในหน่วยความจำของเครื่องคอมพิวเตอร์หรือเทียบเท่า โดยมีการจัดอักขระที่เรียงกันในแต่ละ 1 ตัว อักขระเข้ารหัสตาม มอก. 620 อย่างเคร่งครัด กล่าวคือ สระบน สระล่างหรือพยัญชนะ ต่างก็ถูกเก็บในหน่วยความจำแบบแยกกันตัวละ 1 ไบต์ สายอักขระชุดหนึ่งสิ้นสุดเมื่อพบอักขระที่มีรหัส 0 (NULL character) ตรงกับการเรียกตามความนิยมว่า ASCIIZ
- 2.17 **ความยาวสายอักขระ**(string length;strlen) หมายถึง จำนวนตัวอักขระในสายอักขระไม่นับ NULL character
- 2.18 **ความกว้างสายอักขระ** (string width) หมายถึง จำนวนช่องหรือเซลล์ที่ระบบแสดงผลตามอักขรวิธีภาษาไทยสำหรับคอมพิวเตอร์ ต้องใช้แสดงสายอักขระสายนั้นเช่นสายอักขระ มหาวิทยาลัยอรรถศาสตร์ มีความยาวเป็น 21 ตัว และมีความกว้างเป็น 18 เซลล์
- 2.19 **น้ำหนักเซลล์แสดงผล** (cell weight) หมายถึง จำนวนสัญลักษณ์ที่ปรากฏในเซลล์แสดงผล 1 เซลล์มีค่าตั้งแต่ 1 ถึง 3 เช่น
- เซลล์ สู้ มีน้ำหนักเป็น 3
 เซลล์ ปี มีน้ำหนักเป็น 2

3. ัญนิยม

- 3.1 **สัญลักษณ์เกี่ยวกับรูปแบบของเอกสาร**
 มาตรฐานนี้ ได้อาศัยข้อตกลงในการพิมพ์เอกสารดังนี้
- 3.1.1 **ตัวเอียง** ใช้ในการเรียกคำศัพท์เทคนิค ซึ่งมีคำอธิบายเกี่ยวกับศัพท์นั้น ๆ เป็นครั้งแรกในย่อหน้าที่ปรากฏคำนั้นในสภาพตัวเอียง นอกจากนี้ยังใช้ในการบรรยายชื่อฟังก์ชันและประเภทของข้อมูล (data type) ที่เป็นภาษา C ด้วย (ใช้ในกรณีที่เรียกลอย ๆ ไม่ใช่แสดง program listing แบบเต็ม)
- 3.1.2 **ตัวหนา** ใช้เรียกตัวแปรบอกสภาพแวดล้อมของระบบ (ตัวแปรแวดล้อม(environment variable)) โดยจะเรียกใช้ในสภาพอักษรตัวใหญ่ทั้งคำ เช่น PATH
- 3.1.3 **ตัวภาษาอังกฤษซึ่งมีความกว้างคงที่** (โดยใช้ชุดแบบอักษร Courier) ใช้แสดงตัวอย่างการโต้ตอบทางจอภาพหรือ คำคงตัว ตัวแปร หรือชื่อฟังก์ชันในโปรแกรมที่เขียนเป็นภาษา C (หรือภาษาอื่นใดที่จะมีผู้พัฒนาขึ้นมาเพิ่มเติม)
- 3.1.4 **นิพจน์ทางคอมพิวเตอร์ต่าง ๆ** ในมาตรฐานนี้หากไม่มีการระบุชื่อภาษา จะบรรยายด้วยภาษา C เป็นหลัก หากมีการพัฒนาเป็นภาษาอื่นใดขึ้นมาเพิ่มเติม ให้เพียงแต่ยึดถือคำศัพท์เรียกชื่อในมาตรฐานนี้เป็นแนวทางเท่านั้น
- 3.1.5 **รูป** ในเอกสารนี้หมายถึง แผนผังหรือตารางซึ่งมีรายละเอียดมากกว่าการจัดพิมพ์ด้วยตัวพิมพ์แต่เพียงอย่างเดียว เช่น อาจจะมีการแรเงา ลากเส้นเฉียง หรือลูกศร หรือมีภาพทรงต่าง ๆ มาประกอบ
- 3.1.6 **ตาราง** หมายถึง การจัดข้อความหรือตัวเลขเป็นกลุ่มทางแนวตั้ง/แนวนอนโดยมีเส้นแบ่งกลุ่มในแนวตั้งหรือแนวนอนแต่เพียงอย่างเดียว

- 3.1.7 ตัวเลขหรือจำนวนต่าง ๆ ในมาตรฐานนี้แสดงเป็นเลขธรรมชาติฐานสิบ (decimal radix) ยกเว้นจะมีระบุเป็นพิเศษว่าเป็นฐานอื่นใด ในกรณีที่ยืนยันเป็นจำนวนฐานสิบหกโดยไม่มีการระบุไว้ จะนำหน้าจำนวนต่าง ๆ ด้วยสัญลักษณ์ 0x อย่างชัดเจน ตามวิธีการของภาษา C
- 3.2 การใช้ศัพท์ที่อ้างอิงถึงความสอดคล้องกับมาตรฐาน
 - 3.2.1 ต้อง (shall) หมายถึง มีความจำเป็นต้องปฏิบัติตามมาตรฐานนี้ จึงจะจัดได้ว่าระบบที่จัดทำขึ้นมีความสอดคล้อง (conform) กับมาตรฐานอย่างแท้จริง มิฉะนั้นแล้วคุณสมบัติของระบบที่ขัดแย้งกับเงื่อนไขของมาตรฐานนี้จะถือได้ว่าไม่สอดคล้อง ข้อกำหนดเชิงบังคับที่ใช้คำว่าต้องถือว่าเป็นข้อกำหนดขั้นต่ำที่ผู้สร้างระบบทุกรายต้องสร้างขึ้นมาให้ตรงกัน มิฉะนั้นจะเกิดปัญหาการทำงานร่วมกันระหว่างระบบต่าง ๆ อันเป็นต้นปัญหาที่มาตรฐานนี้พยายามขจัด
 - 3.2.2 ควร (should) หมายถึง คำแนะนำเพื่อนำไปปฏิบัติให้ตรงกัน แม้ว่าเป็นสิ่งที่พึงปรารถนาคล้ายคำว่าต้อง แต่อาจจะมีความยืดหยุ่นบ้างประการ ทำให้จำเป็นต้องผ่อนผันให้ลดหย่อนความจำเป็นนั้นได้ เช่น ในกรณีของสัญลักษณ์ สรรพ ประสมกับนิคทิด (เช่น วิสุ ในภาษาบาลี) อาจจะแสดงผลไม่ได้กับอุปกรณ์ไทย การ์ดจำนวนมากที่มีใช้งานอยู่ทั่วไป มาตรฐานนี้จะได้ลดหย่อนเพื่อเปิดโอกาสให้ระบบที่ยืนอยู่กึ่งฮาร์ดแวร์เก่ามีโอกาสดำเนินการที่สอดคล้องกับมาตรฐานนี้ได้ แต่สำหรับระบบใหม่ ๆ ซึ่งมีข้อกำหนดน้อยลง ผู้จัดทำระบบควรจะต้องดำเนินการที่มีความสามารถครอบคลุมถึงหัวข้อคำแนะนำเหล่านี้โดยสมบูรณ์ได้
 - 3.2.3 อาจจะ (may) หมายถึง ทางเลือกที่มาตรฐานนี้เสนอแนะ โดยผู้จัดทำระบบอาจจะคัดเลือกหรือทำหรือไม่ทำก็ได้ ในกรณีที่มาตรฐานนี้เสนอมากกว่าหนึ่งทางเลือก ผู้จัดทำระบบอาจจะเลือกปฏิบัติเพียงทางหนึ่งทางใดก็ได้ คำว่า อาจจะ มีความหมายเชิงข้อกำหนดสำหรับมาตรฐานน้อยมากและเป็นประเด็นที่จะไม่กระทบกระเทือนต่อความสอดคล้องของระบบ ต่อมาตรฐานแต่อย่างใด
 - 3.2.4 ไม่กำหนด (undefined) ได้แก่ กรณีที่ ค่า (value) หรือ พฤติกรรม (behavior) ใด ๆ ซึ่งมาตรฐานนี้ไม่ได้กล่าวถึงในเชิงบังคับว่าจะมีผลเสียต่อการทำงานร่วมกันอย่างไร ผู้จัดทำระบบอาจจะนำสิ่งที่ไม่กำหนดไปใช้งานด้านการขยายระบบด้านหนึ่งด้านใดได้ตราบใดที่ยังคงไม่กระทบกระเทือนวิธีการทำงานร่วมกับระบบย่อยของผู้อื่นที่สอดคล้องกับมาตรฐานนี้ การกำหนดค่าหรือพฤติกรรมใด ๆ ลงในช่องว่างที่มาตรฐานนี้ไม่ได้กล่าวถึง ควรจะเสนอต่อคณะกรรมการวิชาการคณะที่ 536/2 หรือ Thai API Consortium เพื่อบรรจุเข้าพิจารณาหาความเหมาะสมในการนำมารวมเป็นมาตรฐานฉบับที่ปรับปรุงสำหรับอนาคตต่อไปได้
 - 3.2.5 ห้ามนำไปใช้งาน หมายถึง ค่าหรือพฤติกรรมใด ๆ ที่มาตรฐานนี้จะเจาะจงให้เว้นว่างไปก่อนจนกว่าจะมีการกำหนดความหมายอย่างเป็นทางการในอนาคต หากผู้จัดทำระบบใด ๆ นำไปใช้งานจะถือว่าเป็นการขยายระบบจำเพาะกับเทคโนโลยีของผู้ใช้ และระบบนั้น ๆ อาจจะผ่านหรือไม่ผ่านการทดสอบความสอดคล้อง (conformance verification) กับมาตรฐานนี้ได้
- 3.2.6 กำหนดขึ้นโดยผู้จัดทำระบบ หมายถึง ค่าหรือพฤติกรรมใด ๆ ที่กำหนดขึ้นโดยผู้จัดทำระบบ โดยที่การ

จัดทำระบบนั้นได้กำหนดความต้องการเกี่ยวกับการเขียนโปรแกรมและโครงสร้างข้อมูลโดยมีเอกสารประกอบอย่างครบถ้วนและเป็นสิ่งที่ไม่ได้กำหนดเอาไว้ในมาตรฐานนี้

- 3.2.7 เอกสารยืนยันความสอดคล้อง (conformance document) ได้แก่ เอกสารซึ่งผู้จัดทำระบบทำขึ้นเพื่ออธิบายรายละเอียดเกี่ยวกับระบบนั้น ๆ และเป็นการแจ้งให้ผู้ใช้ระบบทราบว่าระบบนั้น ๆ มีความสอดคล้องกับมาตรฐานนี้อย่างไรโดยจัดเค้าโครงลำดับของการนำเสนอที่สอดคล้องกับมาตรฐานนี้เรียงตามหัวข้อของมาตรฐาน เอกสารนี้ต้องไม่กล่าวถึงความสามารถหรือลักษณะเพิ่มเติมที่มากไปกว่ามาตรฐานระบบที่จะอ้างว่าสอดคล้องกับมาตรฐานทุกระบบต้องมีเอกสารยืนยันความสอดคล้องประกอบเสมอ

3.3 สัญนิยมเกี่ยวกับการเรียกชื่อต่าง ๆ (API naming convention)

- 3.3.1 ตัวแปรแวดล้อม (environment variable) ใช้สัญลักษณ์ภาษาอังกฤษล้วน และเป็นตัวพิมพ์ใหญ่ทั้งหมด นำด้วย TAC_ ตัวอย่างเช่น การกำหนดลักษณะของเครื่องพิมพ์หมายเลข 1 ใช้ตัวแปรชื่อ TAC_LPT1 ค่าของตัวแปรแวดล้อมต้องกำหนดให้กับระบบปฏิบัติการผ่านทางคำสั่งของ OS นั้น ๆ เช่น ในระบบ MS-DOS™ อาจสั่งว่า

```
set TAC_LPT1=EPM113
```

เพื่อแจ้งให้ทราบว่า เครื่องพิมพ์หมายเลข 1 เป็น Epson MX-80 (EPM1) รหัสแบบ 13 เป็นต้น

- 3.3.2 ค่าคงตัว (constant) ใช้สัญลักษณ์ภาษาอังกฤษล้วน เป็นตัวพิมพ์ใหญ่ทั้งหมด นำด้วย TAC_ เช่นเดียวกับตัวแปรแวดล้อม แต่อนุโลมให้ละคำว่า TAC_ ได้ในทางปฏิบัติที่ผู้ใช้แน่ใจว่าไม่ก่อให้เกิดความกำกวม ตัวอย่างเช่น การกำหนดค่ารหัสมาตรฐาน ในแฟ้ม TIS.H ตามภาคผนวก ก. ได้กำหนดค่าคงตัวซึ่งหมายถึง ค ควาย ว่า

```
#define TAC_KHOKHWAI 164
```

หรือ #define KHOKHWAI 164

- 3.3.3 ประเภทของข้อมูล (data type) ใช้สัญลักษณ์ภาษาอังกฤษล้วน และเป็นตัวใหญ่ทั้งหมด ไม่มีการนำด้วยเครื่องหมายใด ๆ เช่น CRT_INFO, THAI_CELL เป็นต้น โปรดสังเกตว่าการแยกคำย่อยใช้สัญลักษณ์ underscore (_) เป็นตัวแยก นอกจากนี้ประเภทของข้อมูลระดับต่ำ (primitive data type) และโครงสร้างข้อมูลระดับต่ำ หากเขียนเป็นภาษา C ให้ยึดถือตามข้อกำหนดต่อไปนี้ด้วย

```
#define CHAR char
#define SHORT short
#define LONG long
#define INT int
typedef unsigned char UCHAR;
typedef unsigned short USHORT;
typedef unsigned long ULONG;
typedef unsigned int UINT;
typedef unsigned short BCool;
typedef unsigned char BYTE;
typedef unsigned int WORD;
```

typedef unsigned long *DWORD*;

- 3.3.4 ฟังก์ชันมาตรฐานเกี่ยวกับอักขรวิธีและสายอักขระไทย (Thai Output and String API) ให้นำชื่อฟังก์ชันด้วย TAC และตามด้วยชื่อของฟังก์ชันที่ตั้งขึ้น ตามกติกาต่าง ๆ ดังนี้

string function ใช้สูตรการตั้งชื่อว่า s<act> หรือ sn<act> เช่น TACswid

conversion function ใช้สูตรการตั้งชื่อว่า <i>to<o> เช่น TACTtod

character test function ใช้สูตรการตั้งชื่อว่า is<what> เช่น TACiscons

โดย <act> หมายถึง action หรือสิ่งที่จะคำนวณหรือทำ เช่น swid เป็นการหาค่า string width ตามวิธีของภาษาไทย

<i> และ <o> หมายถึง input data และ output data ตามลำดับ

<what> เป็นคำถามว่าอักขระเป็นประเภทนี้หรือไม่ เช่น iscons เป็นการตรวจสอบว่า

อักขระที่เป็น parameter เป็นพยัญชนะหรือไม่

โดยปกติการเรียกชื่อฟังก์ชันต่าง ๆ ให้เรียกด้วยชื่อเต็มเสมอ แต่ในกรณีที่มีความประสงค์จะเรียกเป็นชื่อย่อ ให้จัดทำเป็น header file เพิ่มเติม เพื่อใช้ในการขยายชื่อย่อให้กลายเป็นชื่อเต็ม การย่อให้ละ TAC ที่นำหน้าออกไป

- 3.3.5 ฟังก์ชันประเภทจอภาพ+แป้นพิมพ์ และเครื่องพิมพ์ (Display/Keyboard API and Printer API) ฟังก์ชันเหล่านี้ให้นำหน้าด้วย TAC จากนั้นตามด้วย v (จอภาพ) หรือ k (แป้นพิมพ์) หรือ p (เครื่องพิมพ์) และสัญลักษณ์ซึ่งมีความหมายตามสูตรต่อไปนี้

ก. <noun> หรือ <data structure> เช่น

TACvbox วัตถุกรอบบนจอแสดงผล

TACvxyas เขียน string ณ ตำแหน่ง x, y บนจอแสดงผลโดยใช้สี a ที่กำหนดให้

TACkstate ถานสถานะของแป้นพิมพ์

TACpstring พิมพ์ string ออกทางเครื่องพิมพ์

ข. <verb> ใช้สิ่งเปลี่ยนสถานะของสิ่งต่าง ๆ หรือกระทำการบางอย่าง เช่น

TACkread อ่านข้อความเข้ามาจากแป้นพิมพ์

TACcompose สั่งให้เริ่มการพิมพ์ภาษาไทยแบบจัดระดับ

ค. <verb> <noun> ใช้สิ่งกระทำการอะไรบางอย่าง เช่น

TACmoverect ย้ายช่องสี่เหลี่ยมบนจอภาพ

TACkgetshift สอบถามค่า shift ต่าง ๆ ในระบบแป้นพิมพ์

TACpgetinfo สอบถามคุณสมบัติของเครื่องพิมพ์

4. คุณลักษณะที่ต้องการ

4.1 ชุดอักขระ

ในระบบคอมพิวเตอร์ภาษาไทยตามมาตรฐานนี้ประกอบด้วยสัญลักษณ์ต่าง ๆ ซึ่งเข้ารหัสเป็นตัวเลข (coded character) ตามมาตรฐานอุตสาหกรรม โดยแบ่งเป็นชุดอักขระพื้นฐาน 1 ชุด และชุดอักขระเสริมอย่างน้อย 1 ชุด ตามคำจำกัดความต่อไปนี้

4.1.1 ชุดอักขระพื้นฐาน (basic character set) ได้แก่ ชุดอักขระที่เข้ารหัสเลขฐานสองขนาด 8 บิต ไม่ติดเครื่องหมาย มีค่า 0 ถึง 255 (unsigned char หรือ UCHAR) ตามแผนผังในรูปที่ 1 ซึ่งให้ถือว่าเป็นรหัส TACTIS สำหรับใช้ในมาตรฐานนี้

ตารางรหัสนี้เป็นตารางรหัสที่เกิดจากการรวมมาตรฐาน ISO/IEC 646 เข้ากับมาตรฐาน มอก.620 ทั้งนี้ ได้คัดเลือกเฉพาะส่วนที่เข้ากับ ISO/IEC 646 ได้ (ไม่ใช่ส่วนที่เป็น EBCDIC) และได้จัดทำรหัสนี้ขึ้นตามขอบเขตที่ ISO/IEC 2022 ระบุไว้ ดังนั้นจึงเป็นที่คาดหวังว่ารหัสตามรูปที่ 1 สามารถนำไปใช้ร่วมกับ การแลกเปลี่ยนข้อมูลระหว่างประเทศได้อย่างถูกต้อง

รหัส TACTIS เป็นรหัสที่อุตสาหกรรมคอมพิวเตอร์ในประเทศไทยมีความคุ้นเคยกันเป็นอย่างดีเพราะเป็นวิธีการนำ ISO 646 มารวมกับ มอก.620 อย่างถูกต้อง ทั้งนี้ในการเก็บข่าวสารลงในสื่อหรือส่งผ่านระบบ เครือข่าย จะต้องปฏิบัติภายในกรอบของรหัสอย่างเคร่งครัด โดยไม่ใช้รหัสที่ ห้ามใช้ ในมาตรฐานนี้ หากมีการกล่าวถึง อักขระ โดยไม่มีคำอธิบายอื่นใด ให้หมายถึง ชุดอักขระพื้นฐาน TACTIS ตามข้อกำหนดในรูปที่ 1

ชุดอักขระพื้นฐานในระบบภาษาไทยตามมาตรฐาน จะมีรหัสเพียงแบบเดียว แต่เปิดโอกาสให้จัดทำเป็นหลายชุดแบบอักษร (multi-fonts) ได้ โดยอาศัยกลไกของระบบแสดงผลตามมาตรฐานนี้ การแสดงผลโดยมีรูปแบบอักษรที่ต่างกันแต่เป็นอักขระชุดเดียวกันให้ใช้อักษรวิธีเดียวกัน

4.1.2 ชุดอักขระเสริม (auxiliary character set) ได้แก่ ชุดอักขระที่ผู้สร้างระบบสามารถจัดทำอย่างน้อย 1 ชุด และอาจจะเลือกทำหลาย ๆ ชุดก็ได้ โดยมีข้อกำหนดว่าอักขระวิธีของชุดอักขระเสริมให้เป็นไปตามระบบภาษาอังกฤษ กล่าวคือ รหัสในช่วง 32-126 และ 161-254 เป็นรหัสสำหรับอักขระที่กินเนื้อที่แสดงผล 1 เซลล์เสมอ

ชุดอักขระเสริมที่จะต้องมี ได้แก่ ชุดอักขระที่มีภาพการแสดงผลตรงกับ Code Page 437 ของ MS-DOS™ ทั้งนี้อนุโลมให้ระบบแสดงผลที่มีข้อจำกัดทางเทคนิค แสดงชุดตัวอักษรเพียงบางส่วนของ Code Page 437 ได้ตามรูปที่ 2 (ก) อักขระใด ๆ ที่ไม่สามารถแสดงได้ให้ใช้สัญลักษณ์จุด (.) แทน นอกจากนี้ หากระบบยังมีข้อจำกัดอื่นใด ให้อนุโลมรูปแบบแสดงภาพต่าง ๆ ในรูปที่ 2 (ข) เป็นแบบง่ายตามรูปที่ 2 (ค) ด้วย

ชุดอักขระเสริมชุดอื่น ๆ ที่ผู้จัดทำระบบประสงค์จะจัดทำขึ้น ให้ใช้อักขระวิธีแบบภาษาอังกฤษเหมือนกันหมดทุกชุด

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	NUL	DLE	SP	0	@	P		p	Ç	É	á	■	L	⌌	α	≡
1	SOH	DC1	!	1	A	Q	a	q	Û	æ	í	■	⌌	⌌	β	±
2	STX	DC2	"	2	B	R	b	r	é	Æ	ó	■	⌌	⌌	Γ	≥
3	ETX	DC3	#	3	C	S	c	s	â	ô	ú		⌌	⌌	π	≤
4	EOT	DC4	\$	4	D	T	d	t	ã	ö	ñ	⌌	⌌	⌌	Σ	(
5	ENQ	NAK	%	5	E	U	e	u	ä	ò	Ñ	⌌	⌌	⌌	σ)
6	ACK	SYN	&	6	F	V	f	v	å	ó	ª	⌌	⌌	⌌	μ	÷
7	BEL	ETB	'	7	G	W	g	w	ç	ù	º	⌌	⌌	⌌	τ	≈
8	BS	CAN	(8	H	X	h	x	ê	ÿ	¿	⌌	⌌	⌌	Φ	°
9	HT	EM)	9	I	Y	i	y	ë	ö	9	⌌	⌌	⌌	θ	•
A	LF	SUB	*	:	J	Z	j	z	è	ü	À	⌌	⌌	⌌	Ω	-
B	VT	ESC	+	;	K	[k	;	í	ç	½	⌌	⌌	■	δ	√
C	FF	FS	,	<	L	\	l	l	î	£	¼	⌌	⌌	■	ω	ⁿ
D	CR	GS	-	=	M]	m]	ï	¥	ï	⌌	=	■	∅	²
E	SO	RS	.	>	N	^	n	-	Ä	E	«	⌌	⌌	■	ε	■
F	SI	US	/	?	O	_	o	DEL	Å	F	»	⌌	⌌	■	∩	

รูปที่ 2 (ก) Code Page 437 US

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	NUL	DLE	SP	0	@	P		p					L	⌋		
1	SOH	DC1	!	1	A	Q	a	q					⌋	⌋		
2	STX	DC2	"	2	B	R	b	r					⌋	⌋		
3	ETX	DC3	#	3	C	S	c	s						⌋	⌋	
4	EOT	DC4	\$	4	D	T	d	t					⌋	⌋	⌋	
5	ENQ	NAK	%	5	E	U	e	u					⌋	⌋	⌋	
6	ACK	SYN	&	6	F	V	f	v					⌋	⌋	⌋	
7	BEL	ETB	'	7	G	W	g	w					⌋	⌋	⌋	
8	BS	CAN	(8	H	X	h	x					⌋	⌋	⌋	
9	HT	EM)	9	I	Y	i	y					⌋	⌋	⌋	
A	LF	SUB	*	:	J	Z	j	z					⌋	⌋	⌋	
B	VT	ESC	+	:	K	[k	[⌋	⌋		
C	FF	FS	,	<	L	\	l	l					⌋	⌋		
D	CR	GS	-	*	M]	m]					⌋	=		
E	SO	RS	.	>	N	^	n	-					⌋	⌋		
F	SI	US	/	?	O	_	o	DEL					⌋	⌋		

รูปที่ 2 (ข) ชุดอักขระเสริมขั้นต่ำสุดตามมาตรฐาน

	A	B	C	D	E	F
			L	⌋		
			⌋	⌋		
			⌋	⌋		
			⌋	⌋		
				⌋	⌋	
			⌋	⌋	⌋	
			⌋	⌋	⌋	
			⌋	⌋	⌋	
			⌋	⌋	⌋	
			⌋	⌋	⌋	
			⌋	⌋	⌋	
			⌋	⌋	⌋	
			⌋	⌋	⌋	
			⌋	⌋	⌋	
			⌋	⌋	⌋	

รูปที่ 2 (ค) ตัวอย่างการสร้างชุดอักขระเสริมในระบบที่มีข้อจำกัดด้านจำนวนตัวอักษร

รูปที่ 2 ชุดอักขระเสริม
(ข้อ 4.1.2)

4.2 ประเภทของอักขระไทย/อังกฤษ

การจำแนกประเภทอักขระพื้นฐานในมาตรฐานนี้เป็นการจำแนกประเภทเพื่อความสะดวกในการประมวลผลด้วยคอมพิวเตอร์ด้านการแสดงผล ซึ่งนำไปเปรียบเทียบกับกรจำแนกทางภาษาเพื่อสรุปว่าเป็นสิ่งที่ถูก หรือ ผิดไม่ได้ อย่างไรก็ตามข้อกำหนดต่าง ๆ ในย่อหน้านี้จัดทำขึ้นเพื่อให้คอมพิวเตอร์สามารถแสดงผลเป็นภาษาไทยได้อย่างถูกต้องสมบูรณ์ตามหลักภาษาไทยที่ปรากฏในตำราภาษาไทยต่าง ๆ เช่น หนังสือ *หลักภาษาไทย* (พระยาอุปกิตศิลปสาร) หนังสือ *มูลบทบรรพกิจ* (พระยาศรีสุนทรโวหาร) และ *ศัพท์บัญญัติวิชาการพิมพ์* (ราชบัณฑิตยสถาน) ประโยชน์ของการจำแนกประเภทอักขระไทยตามมาตรฐานนี้อีกด้านหนึ่งก็คือ การบรรยายหรือกำหนดอักขระวิธของไทยโดยละเอียดซึ่งเป็นข้อกำหนดเชิงกลุ่มตัวอักษร มีดัง ในข้อ 4.3 และ ข้อ 4.4

จากหลักการข้างต้น มาตรฐานนี้ได้จำแนกประเภทอักขระไทย/อังกฤษในชุดอักขระพื้นฐานไว้ดังนี้

- 4.2.1 *อักขระควบคุม* (control character หรือ CTRL) ได้แก่อักขระที่ไม่ปรากฏการแสดงผลบนอุปกรณ์แสดงผล (non-display character) แต่ใช้ในการควบคุมการทำงานของอุปกรณ์แสดงผลหรืออุปกรณ์สื่อสาร อักขระเหล่านี้มีรหัสอยู่ในช่วง 0-31, 127, 128-159 และ 255 รวมทั้งสิ้น 66 ตัว
- 4.2.2 *พยัญชนะไทย* (consonant หรือ CONS) มี 44 ตัว ได้แก่ ก ข ช ค ต ข ฉ ง จ ฉ ฎ ฏ ฐ ฑ ฒ ณ ด ต ถ ท ธ น บ ป ผ ฝ พ ฟ ภ ม ย ร ล ว ศ ษ ส ห อ ฮ
- 4.2.3 *สระไทย* (vowel หรือ -V) ประกอบด้วยกลุ่มสัญลักษณ์ต่อไปนี้
- 4.2.3.1 *สระนำ* คือสระที่นำหน้าพยัญชนะต้น (leading vowel หรือ LV) มี 5 ตัว ได้แก่ แ ไ อ ไ
- 4.2.3.2 *สระตาม* คือสระที่อยู่หลังพยัญชนะต้น (following vowel หรือ FV) มี 6 ตัว ได้แก่ *สระตามทั่วไป* 4 ตัว คือ ะ ำ า ฤ *สระตามพิเศษ* 2 ตัว คือ ฤ ฎ
- 4.2.3.3 *สระล่าง* คือสระที่อยู่ใต้พยัญชนะต้น (below vowel หรือ BV) มี 2 ตัว ได้แก่ ุ ู
- 4.2.3.4 *สระบน* คือสระที่อยู่เหนือพยัญชนะต้น (above vowel หรือ AV) มี 5 ตัว ได้แก่ ิ ี ึ ื ุ
- 4.2.4 *เครื่องหมายวรรณยุกต์* (tonemark หรือ TONE) มี 4 รูป ได้แก่ ˊ ˋ ˊ ˋ
- 4.2.5 *เครื่องหมายกำกับเสียง* (diacritic หรือ -D) ประกอบด้วยกลุ่มสัญลักษณ์ดังนี้
- 4.2.5.1 เครื่องหมายกำกับเสียงเหนือพยัญชนะต้นหรือตัวสะกด (above diacritic หรือ AD) มี 4 ตัว ได้แก่ ˊ (นิคหิต) ˋ (ทัณฑฆาต) ˊ (ไม้ไต่คู้) และ ˋ (ยามักการ)
- 4.2.5.2 เครื่องหมายกำกับเสียงใต้ตัวสะกดหรือตัวควบ (below diacritic หรือ BD) มี 1 ตัว คือ ˋ (พินทุ)
- 4.2.6 *สัญลักษณ์อื่น ๆ* (non-composable หรือ NON) ได้แก่ สัญลักษณ์ต่าง ๆ ที่นอกเหนือไปจาก 5 ประเภทที่กล่าวมาแล้วข้างต้นนี้และมิคุณสมบัติเป็นสัญลักษณ์ที่ใช้กับอักขระวิธแบบอังกฤษทั่วไป คือ ตัวละ 1 ช่อง ไม่มี การนำสัญลักษณ์อื่นใดมาประสม (compose) เข้าด้วยกัน อักขระประเภทนี้มี 119 ตัว โดยประกอบด้วยกลุ่มย่อย ๆ ดังนี้
- 4.2.6.1 *อักขระแสดงผลสากล* (graphic character) ตาม ISO 646 จำนวน 94 ตัว ซึ่งประกอบด้วยตัวอักษรภาษาอังกฤษ 52 ตัว (ได้แก่ A B C D E F G H I J K L M N O P Q R S T U V W X Y Z a b c d e f g h i j k l m n o p q r s t u v w x y z) เลขสากล 10 ตัว (ได้แก่ 0 1 2 3 4 5 6 7 8 9) และอักขระพิเศษ 32 ตัว ได้แก่ ! " # \$ % & ' () * + , - . / : ; < = > ? @ [\] ^ _ ` |) ~
- 4.2.6.2 *อักขระเว้นวรรคสากล* (space) 1 ตัว คือ รหัส 32 (0x20)

มอก. 1566-2541

- 4.2.6.3 อักขระห้ามแยก (no-break space) มี 1 ตัว คือ รหัส 160 (0xA0)
- 4.2.6.4 ตัวเลขไทย ตาม มอก.620 มี 10 ตัว ได้แก่ ๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙
- 4.2.6.5 เครื่องหมายพิเศษ ตาม มอก.620 มี 6 ตัว ได้แก่ ๗ (ไปยาลน้อย) ๘ (บาท) ๙ (ไม้ยมก) © (ฟองมัน) ๙ (อังก๋ันคู่) และ ¸ (โคมุต)
- 4.2.6.6 สัญลักษณ์แยกคำในสายอักขระไทย (word break) มี 1 ตัว (รหัส 220 หรือ 0xDC) สำหรับใช้แทรก
ระหว่างคำไทยในฐานะข้อมูลประเภทข้อความ อักขระนี้โดยปกติเป็นตัวไม่แสดงผลในการประยุกต์
ใช้งานทั่วไป ในบางกรณี เช่น การตรวจสอบความถูกต้องของเอกสาร ผู้จัดทำระบบอาจกำหนดให้
แสดงผลเป็นรูปร่างใด ๆ ที่แสดงการแยกของคำได้ตามความเหมาะสมตามสภาพการใช้งาน
- 4.2.6.7 รหัสห้ามนำไปใช้งาน มี 6 รหัส ได้แก่ 219, 221, 222, 252, 253 และ 254 (0xDB, 0xDD,
0xDE,0xFC, 0xFD, 0xFE)

เพื่อช่วยให้สามารถบรรยายข้อกำหนดทางด้านอักขรวิธีภาษาไทยที่สมบูรณ์ การจำแนกประเภทของอักขระในกลุ่ม
FV,BV,AV และ AD จำเป็นต้องแบ่งย่อยลงไปอีกโดยการใช้หมายเลขกำกับกลุ่มย่อย (เช่น FV ประกอบด้วย
FV1,FV2 และ FV3) ซึ่งจะนำมาใช้ประโยชน์ดังรายละเอียดในตารางที่ 1

4.3 คุณสมบัติของอักขระในชุดอักขระพื้นฐาน

อักขระไทย/อังกฤษเมื่อนำมาใช้งานแสดงผลทางคอมพิวเตอร์ไม่ว่าจะอยู่ในระบบใด จะต้องมียุคสมบัติที่เหมือนกัน ทั้งในแง่ของการจัดการแสดงผล และการวิเคราะห์ความถูกต้องของสายอักขระโดยวิธีการตรวจสอบความเป็นไปได้ หัวข้อนี้เป็นการบรรยายคุณสมบัติด้านต่าง ๆ ของอักขระต่าง ๆ ซึ่งจะจำแนกเป็นประเด็นต่าง ๆ ได้แก่ การกำหนด “เลขบอกประเภทอักขระ (chtype)” การเรียกชื่อ ระดับการแสดงผลอักขระ(chlevel) การสอบถามคุณสมบัติ *ความเป็นตัวหยุดอยู่กับที่(isdead)* *ความเป็นพยัญชนะ(iscons)* *ความเป็นสระ(isvowel)* *ความเป็นวรรณยุกต์(istone)* *ความเป็นตัวเลข(isdigit)* *ความเป็นเครื่องหมายวรรคตอน(ispunct)* ฯลฯ ดังรายละเอียดต่อไปนี้

4.3.1 เลขบอกประเภทอักขระ (character type) ให้ยึดถือตามค่าจำกัดความตามตารางที่ 2 ในการเขียนโปรแกรม เพื่อหาค่า character type ของอักขระใด ๆ ให้เรียกใช้ API ฟังก์ชันชื่อ TACchtype ซึ่งมีข้อกำหนดการเรียกใช้เป็นภาษา C ดังนี้

int TACchtype (unsigned char ch) ;

โดยฟังก์ชันนี้จะคืนค่า chtype ที่มีค่าระหว่าง 0-16 ตามข้อกำหนดในตารางที่ 2

ตารางที่ 2 เลขบอกประเภทอักขระ

(ข้อ 4.3.1)

ชื่อประเภท	เลขบอกประเภท
CTRL	0
NON	1
CONS	2
LV	3
FV1	4
FV2	5
FV3	6
BV1	7
BV2	8
BD	9
TONE	10
AD1	11
AD2	12
AD3	13
AV1	14
AV2	15
AV3	16

- 4.3.2 การเรียกชื่อระดับการแสดงผลอักษร (chlevel) มาตรฐานนี้ได้จัดระดับการแสดงผลอักษรไทย-อังกฤษ เป็น 5 ระดับ (แสดงผล 4 ระดับ และ ไม่แสดงผล 1 ระดับ) ดังนี้
- 4.3.2.1 ระดับไม่แสดงผล (NONDISP) ได้แก่สัญลักษณ์ควบคุมต่าง ๆ (กลุ่ม CTRL)
- 4.3.2.2 ระดับฐาน (BASE) ได้แก่สัญลักษณ์ที่วางในระดับบรรทัดตามปกติ ได้แก่ สัญลักษณ์กลุ่ม NON, CONS และกลุ่ม FV, LV
- 4.3.2.3 ระดับเหนือฐาน (ABOVE) ได้แก่ สัญลักษณ์ไทยที่วางซ้อนกับพยัญชนะตัวสุดท้ายโดยอยู่เหนือพยัญชนะ นั้น ๆ ได้แก่สัญลักษณ์กลุ่ม BV1, BV2 และ BD
- 4.3.2.4 ระดับต่ำกว่าฐาน (BELOW) ได้แก่ สัญลักษณ์กลุ่ม AD และ AV
- 4.3.2.5 ระดับบนสุด (TOP) ได้แก่รรณยุกต์(กลุ่ม TONE) กลุ่ม AD1 และ AD2 ซึ่งเป็นสัญลักษณ์ที่อาจจะปรากฏควบคู่กับสัญลักษณ์ในระดับเหนือฐานอื่น ๆ ได้ในช่องเดียวกัน ดังนั้นจึงต้องแสดงอยู่ในระดับที่สูงกว่าสัญลักษณ์ระดับเหนือฐานอย่างไรก็ตามสัญลักษณ์ในกลุ่มนี้อาจจะนำมาวางในระดับที่ต่ำลงได้ในกรณีที่ไม่มีสัญลักษณ์ระดับเหนือฐานอยู่ร่วมช่องเดียวกัน
- นอกจากนี้สัญลักษณ์ที่แสดงผลในระดับสูงสุดนี้ยังเป็นตัวระบุจุดสิ้นสุดของเซลล์แสดงผลนั้น ๆ ด้วย
- ดังจะอธิบายต่อไปในข้อ 4.4

รูปที่ 3 สรุปความหมายของระดับการแสดงผลอักษรตามมาตรฐานพร้อมกับค่าตัวเลขกำกับระดับการแสดงผลที่จะใช้ในระบบภาษาไทยทั่วไป ซึ่งมีข้อกำหนดการเรียกใช้ฟังก์ชันหาระดับการแสดงผลของอักษร ดังนี้

int TACchlevel (unsigned char ch);

โดยฟังก์ชันนี้จะคืนค่า chlevel ที่มีค่าระหว่าง 0-4 ตามมาตรฐานโดยค่า 0-4 จะตรงกับชื่อค่าคงตัวใน การเขียนโปรแกรมว่า NONDISP, TOP, ABOVE, BASE และ BELOW ตามลำดับ

Ag>>ปู่	1	TOP
	2	ABOVE
	3	BASE
	4	BELOW

- หมายเหตุ 1. ระดับ 0 คือการไม่แสดงผล (NONDISP)
2. ไม่เอ็อกของคำว่า ปู่ เป็นตัวอักษรระดับ TOP แต่อาจถูกเลื่อนระดับให้ต่ำลงมาเพื่อความสวยงาม

รูปที่ 3 ความหมายของระดับการแสดงผลอักษร

(ข้อ 4.3.2)

- 4.3.3 การสอบถามคุณสมบัติอื่น ๆ เกี่ยวกับอักษรแต่ละตัว อักษรในแต่ละประเภทมีคุณสมบัติประจำตัวด้านอื่น ๆ ที่สามารถกำหนดให้ชัดเจนได้โดยข้อกำหนดต่าง ๆ ดังสรุปในตารางที่ 3 โดยให้กำหนดเชิงกลุ่มประเภทของอักษร ตามความหมายต่อไปนี้

- 4.3.3.1 **ตัวหยุดอยู่กับที่** ได้แก่สัญลักษณ์ที่แสดงผล ณ ช่องแสดงผลที่ผ่านมาในลักษณะการรวมกับอักขระประเภทพยัญชนะ การแสดงตัวหยุดอยู่กับที่ ในสภาพปกติจะไม่ให้ผลให้ตัวชี้ตำแหน่งเคลื่อนย้ายตำแหน่งแต่อย่างใดอักขระที่ไม่มีหยุดอยู่กับที่จะมีชื่อเรียกว่า ตัวเดินหน้า ซึ่งได้แก่ สัญลักษณ์ภาษา อังกฤษ พยัญชนะ สระประเภท LV และ FV ในขณะที่ตัวหยุดจะได้แก่สัญลักษณ์ประเภท CTRL, BV, AV, BD,AD และ TONE ทั้งหมด
- 4.3.3.2 **พยัญชนะ** ได้แก่อักขระประเภท CONS ทั้งหมด
- 4.3.3.3 **สระ** ได้แก่อักขระประเภท LV, FV, BV, AV ทั้งหมด
- 4.3.3.4 **วรรณยุกต์** ได้แก่อักขระประเภท TONE ทั้งหมด
- 4.3.3.5 **ตัวเลข** ได้แก่อักขระ 0 1 2 3 4 5 6 7 8 9 ๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙ รวม 20 ตัว

ตารางที่ 8 สรุปคุณสมบัติเบื้องต้นของอักขระต่างๆพร้อมชื่อย่อของฟังก์ชันที่ใช้ในการสอบถามคุณสมบัติเหล่านั้น (ชื่อเต็มต้องนำด้วย TAC)
(ข้อ 4.3.3)

chtype	ประเภท	chlevel	istcad	iscons	isvowel	istone	isdigit	ispunct
0	CTRL	NONDISP	dead					
1	NON	BASE		cons			0-9,๐-๙	ตามที่ระบุในตารางที่ 4
2	CONS	BASE						
3	LV	BASE			vowel			
4	FV1	BASE			vowel			
5	FV2	BASE			vowel			
6	FV3	BASE			vowel			
7	BV1	BELOW	dcad		vowel			
8	BV2	BELOW	dcad		vowel			
9	BD	BELOW	dcad					
10	TONE	TOP	dcad			tone		
11	AD1	TOP	dcad					
12	AD2	TOP	dcad					
13	AD3	TOP	dcad					
14	AV1	TOP	dcad		vowel			
15	AV2	TOP	dcad		vowel			
16	AV3	TOP	dcad		vowel			

4.3.4 ตารางแจกแจงลักษณะเฉพาะตัวของอักขระ

ตารางที่ 4 สรุปคุณสมบัติต่างๆที่เกี่ยวข้องกับอักขระแต่ละตัวในชุดอักขระพื้นฐานเพื่อใช้ในการกำหนด อักขระวิธีเชิงสัมพันธ์ระหว่างอักขระต่าง ๆ ในสายอักขระเดียวกัน ดังจะบรรยายในข้อ 4.4 เป็นต้นไป ใน ตารางนั้นนอกจากจะเป็นการสรุปคุณสมบัติต่าง ๆ ของอักขระแล้ว ยังได้แสดงรายละเอียดเกี่ยวกับอักขระ ได้แก่

รหัสฐานสิบ หมายถึง รหัสอักขระเป็นเลขฐานสิบ (Dec.)

รหัสฐานสิบหก หมายถึง รหัสอักขระเป็นเลขฐานสิบหก (Hex.)

ตำแหน่งรหัสตาม ISO หมายถึง ตำแหน่งของรหัสตามวิธีการเรียกหมายเลขแถวและสดมภ์ (Pos.) ตามวิธีการของ ISO

ชื่ออักขระเป็นภาษาไทย (ชื่อ)

ชื่ออักขระเป็นภาษาอังกฤษ (Name)

ภาพของสัญลักษณ์ (Char.)

ประเภท (Class)

ระดับการแสดงผล (Level)

คุณสมบัติ

สำหรับอักขระในชุดพื้นฐานทุกตัว เฉพาะรหัสในช่วง 161-251 (0xA1 ถึง 0xFB) ชื่ออักขระไทยที่เป็นภาษาไทยและภาษาอังกฤษที่ยึดถือเป็นชื่อเรียกค่าคงตัวได้ในภาษาโปรแกรมทั่วไปดังแสดงไว้ในแฟ้ม TIS.H ในภาคผนวก ก.

ตารางที่ 4 ลักษณะเฉพาะตัวของอักขระต่าง ๆ ในชุดพื้นฐาน
(ข้อ 4.3.4)

Dec.	Hex.	Oct.	ชื่อ	Name	Char.	Class	Level	Properties
0	00	00/00		NULL	NULL	CTRL	NONDISP	
1	01	00/01		START OF HEADING	SOH	CTRL	NONDISP	
2	02	00/02		START OF TEXT	STX	CTRL	NONDISP	
3	03	00/03		END OF TEXT	ETX	CTRL	NONDISP	
4	04	00/04		END OF TRANSMISSION	EOT	CTRL	NONDISP	
5	05	00/05		ENQUIRY	ENQ	CTRL	NONDISP	
6	06	00/06		ACKNOWLEDGE	ACK	CTRL	NONDISP	
7	07	00/07		BELL	BEL	CTRL	NONDISP	
8	08	00/08		BACKSPACE	BS	CTRL	NONDISP	
9	09	00/09		HORIZONTAL TABULATION	HT	CTRL	NONDISP	
10	0A	00/10		LINE FEED	LF	CTRL	NONDISP	
11	0B	00/11		VERTICAL TAB	VT	CTRL	NONDISP	
12	0C	00/12		FORM FEED	FF	CTRL	NONDISP	
13	0D	00/13		CARRIAGE RETURN	CR	CTRL	NONDISP	
14	0E	00/14		SHIFT OUT	SO	CTRL	NONDISP	
15	0F	00/15		SHIFT IN	SI	CTRL	NONDISP	
16	10	01/00		DATA LINK ESCAPE	DLE	CTRL	NONDISP	
17	11	01/01		DEVICE CONTROL 1,XON	DC1	CTRL	NONDISP	
18	12	01/02		DEVICE CONTROL 2	DC2	CTRL	NONDISP	
19	13	01/03		DEVICE CONTROL 3,XOFF	DC3	CTRL	NONDISP	
20	14	01/04		DEVICE CONTROL 4	DC4	CTRL	NONDISP	
21	15	01/05		NEGATIVE ACKNOWLEDGE	NAK	CTRL	NONDISP	
22	16	01/06		SYNCHRONOUS IDLE	SYN	CTRL	NONDISP	
23	17	01/07		END OF TRANSMISSION BLOCK	ETB	CTRL	NONDISP	
24	18	01/08		ANCEL	CAC	CTRL	NONDISP	
25	19	01/09		END OF MEDIUM	EM	CTRL	NONDISP	
26	1A	01/10		SUBSTITUTE CHARACTER	SUB	CTRL	NONDISP	
27	2B	01/11		ESCAPE	ESC	CTRL	NONDISP	
28	1C	01/12		INFORMATION SEPARATOR 4,(FILE SEPARATOR)	FS	CTRL	NONDISP	
29	1D	01/13		INFORMATION SEPARATOR 3,(GROUP SEPARATOR)	GS	CTRL	NONDISP	
30	1E	01/14		INFORMATION SEPARATOR 2,(RECORD SEPARATOR)	RS	CTRL	NONDISP	
31	1F	01/15		INFORMATION SEPARATOR 1,(UNIT SEPARATOR)	US	CTRL	NONDISP	
32	20	02/00	เว้นวรรค	SPACE	SP	NON	BASE	
33	21	02/01	อัฒจันทร์	EXCLAMATION MARK	!	NON	BASE	punct
34	22	02/02	พาดพิง	QUATATION MARK	"	NON	BASE	punct
35	23	02/03		NUMBER SIGN	#	NON	BASE	punct
36	24	02/04		DOLLAR SIGN	\$	NON	BASE	punct
37	25	02/05		PERCENT SIGN	%	NON	BASE	punct
38	26	02/06		AMPERSAND	&	NON	BASE	punct
39	27	02/07		APOSTROPHE	'	NON	BASE	punct
40	28	02/08	วงเล็บเปิด	LEFT PARENTHESIS	(NON	BASE	punct

ตารางที่ 4 อักขระเฉพาะตัวของอักขระต่างๆ ในชุดพื้นฐาน (ต่อ)

Dec.	Hex.	Pos.	ชื่อ	Name	Char.	Class	Level	Properties
41	29	02/09	วงเล็บปิด	RIGHT PARENTHESIS)	NON	BASE	punct
42	2A	02/10	ดอกจัน	ASTERISK	*	NON	BASE	punct
43	2B	02/11	บวก	PLUS SIGN	+	NON	BASE	punct
44	2C	02/12	จุลภาค	COMMA	,	NON	BASE	punct
45	2D	02/13	ขีดลบ	HYPHEN, MINUS SIGN	-	NON	BASE	punct
46	2E	02/14	มหัพภาค	FULL STOP	.	NON	BASE	punct
47	2F	02/15	ทับ	SOLIDUS	/	NON	BASE	punct
48	30	03/00	ศูนย์	DIGIT ZERO	0	NON	BASE	digit
49	31	03/01	หนึ่ง	DIGIT ONE	1	NON	BASE	digit
50	32	03/02	สอง	DIGIT TWO	2	NON	BASE	digit
51	33	03/03	สาม	DIGIT THREE	3	NON	BASE	digit
52	34	03/04	สี่	DIGIT FOUR	4	NON	BASE	digit
53	35	03/05	ห้า	DIGIT FIVE	5	NON	BASE	digit
54	36	03/06	หก	DIGIT SIX	6	NON	BASE	digit
55	37	03/07	เจ็ด	DIGIT SEVEN	7	NON	BASE	digit
56	38	03/08	แปด	DIGIT EIGHT	8	NON	BASE	digit
57	39	03/09	เก้า	DIGIT NINE	9	NON	BASE	digit
58	3A	03/10	วงเล็บปิด	COLON	:	NON	BASE	punct
59	3B	03/11	อัฒภาค	SEMICOLON	;	NON	BASE	punct
60	3C	03/12	น้อยกว่า	LESS-THAN SIGN	<	NON	BASE	punct
61	3D	03/13	เท่ากับ	EQUAL SIGN	=	NON	BASE	punct
62	3E	03/14	มากกว่า	GREATER-THAN SIGN	>	NON	BASE	punct
63	3F	03/15	ประจักษ์	QUESTION MARK	?	NON	BASE	punct
64	40	04/00		COMMERCIAL AT	@	NON	BASE	punct
65	41	04/01		CAPITAL LETTER A	A	NON	BASE	xdigit alpha
66	42	04/02		CAPITAL LETTER B	B	NON	BASE	xdigit alpha
.....								
90	5A	05/10		CAPITAL LETTER Z	Z	NON	BASE	alpha
91	5B	05/11	วงเล็บ เหลี่ยมเปิด	LEFT SQUARE BRACKET	[NON	BASE	alpha
92	5C	05/12	ทับกลับ	REVERSE SOLIDUS	\	NON	BASE	punct
93	5D	05/13	วงเล็บ เหลี่ยมปิด	RIGHT SQUARE BRACKET]	NON	BASE	punct
94	5E	05/14		CIRCUMFLEX ACCENT	^	NON	BASE	punct
95	5F	05/15	ขีดเส้นใต้	LOW LINE, UNDERLINE	_	NON	BASE	punct
96	60	06/00		GRAVE ACCENT	`	NON	BASE	punct
97	61	06/01		SMALL LETTER A	a	NON	BASE	alpha
98	62	06/02		SMALL LETTER B	b	NON	BASE	alpha
.....								
122	7A	07/10		SMALL LETTER Z	z	NON	BASE	alpha
123	7B	07/11	วงเล็บปีก กาเปิด	LEFT CURLY BRACKET	{	NON	BASE	punct
124	7C	07/12	เส้นตั้ง	VERTICAL LINE		NON	BASE	punct

ตารางที่ 4 ลักษณะเฉพาะตัวของอักขระต่าง ๆ ในชุดพื้นฐาน (ต่อ)

Dec.	Hex.	Pos.	ชื่อ	Name	Char.	Class	Level	Properties
125	7D	07/13	วงเล็บปีกกาคิด	RIGHT CURLY BRACKET		NON	BASE	punct
126	7E	07/14		TILDE	~	NON	BASE	punct
127	7F	07/15		DELETE	DEL	CTRL	NONDISP	
160	A0	10/00	อักขระห้ามแยก	NO-BREAK SPACE	NON	NONDISP		
161	A1	10/01	ก-ไก	KO KAI	ก	CONS	BASE	cons alpha
162	A2	10/02	ข-ไค	KHO KHAI	ข	CONS	BASE	cons alpha
163	A3	10/03	ช-ชวต	KHO KHUAT	ช	CONS	BASE	cons alpha
164	A4	10/04	ค-ควย	KHO KHWAI	ค	CONS	BASE	cons alpha
165	A5	10/05	ด-ดพ	KHO KHON	ด	CONS	BASE	cons alpha
166	A6	10/06	ร-รคัง	KHO RAKHANG	ร	CONS	BASE	cons alpha
167	A7	10/07	ง-งู	NGO NGU	ง	CONS	BASE	cons alpha
168	A8	10/08	จ-จัน	CHO CHAN	จ	CONS	BASE	cons alpha
169	A9	10/09	ฉ-ฉัง	CHO CHING	ฉ	CONS	BASE	cons alpha
170	AA	10/10	ช-ชัง	CHO CHANG	ช	CONS	BASE	cons alpha
171	AB	10/11	ซ-โซ	SO SO	ซ	CONS	BASE	cons alpha
172	AC	10/12	ฃ-ฃอ	CHO CHOE	ฃ	CONS	BASE	cons alpha
173	AD	10/13	ญ-ญิง	YO YING	ญ	CONS	BASE	cons alpha
174	AE	10/14	ด-ชฎา	DO CHADA	ด	CONS	BASE	cons alpha
175	AF	10/15	ฏ-ปฎัก	TO PATAK	ฏ	CONS	BASE	cons alpha
176	B0	11/00	ฐ-ฐาน	THO THAN	ฐ	CONS	BASE	cons alpha
177	B1	11/01	ท-นวมณฑล	THO NAMMONTHO	ท	CONS	BASE	cons alpha
178	B2	11/02	ณ-ผู้เฒ่า	CONS	ณ	CONS	BASE	cons alpha
179	B3	11/03	น-เนน	NO NEN	ณ	CONS	BASE	cons alpha
180	B4	11/04	ด-เดก	DO DEK	ด	CONS	BASE	cons alpha
181	B5	11/05	ด-เต้า	TO TAO	ด	CONS	BASE	cons alpha
182	B6	11/06	ด-ตุง	THO THUNG	ด	CONS	BASE	cons alpha
183	B7	11/07	ท-ทหา	THO THAHAN	ท	CONS	BASE	cons alpha
184	B8	11/08	ธ-ธง	THO THONG	ธ	CONS	BASE	cons alpha
185	B9	11/09	น-นญ	NO NY	น	CONS	BASE	cons alpha
186	BA	11/10	บ-ใบไม้	BO BAIMAI	บ	CONS	BASE	cons alpha
187	BB	11/11	ป-ปถา	CONS	ป	CONS	BASE	cons alpha
188	BC	11/12	ผ-ผัง	PHO PHUNG	ผ	CONS	BASE	cons alpha
189	BD	11/13	ฟ-ฟา	FO FA	ฟ	CONS	BASE	cons alpha
190	BE	11/14	พ-พาน	PHO PHAN	พ	CONS	BASE	cons alpha
191	BF	11/15	ฟ-ฟัน	FO FAN	ฟ	CONS	BASE	cons alpha
192	C0	12/00	ภ-สัมภา	PHO SAMPHAO	ภ	CONS	BASE	cons alpha
193	C1	12/01	ม-ม้า	MO MA	ม	CONS	BASE	cons alpha
194	C2	12/02	ย-ยัก	YO YAK	ย	CONS	BASE	cons alpha
195	C3	12/03	ร-เรือ	RO RUA	ร	CONS	BASE	cons alpha
196	C4	12/04	ร	RU	ร	PV3	BASE	vowel alpha
197	C5	12/05	ล-ลิง	LO LING	ล	CONS	BASE	cons alpha

ตารางที่ 4 ลักษณะเฉพาะตัวของอักขระต่าง ๆ ในชุดพื้นฐาน (ต่อ)

Doc.	Hex.	Pos.	ชื่อ	Name	Char.	Class	Level	Properties
198	C8	12/06	ลี	LI	1	CONS	BASE	vowel alpha
199	C7	12/07	ว-เววม	WO WAEN	ว	CONS	BASE	cons alpha
200	C8	12/08	ศ-ศาลา	SO SALA	ศ	CONS	BASE	cons alpha
201	C9	12/09	ซ-ซุสึ	SO RUSI	ซ	CONS	BASE	cons alpha
202	CA	12/10	ส-เสื่อ	SO SUA	ส	CONS	BASE	cons alpha
203	C8	12/11	ห-หับ	HO HIP	ห	CONS	BASE	cons alpha
204	CC	12/12	หล-จุลา	LO CHULA	หล	CONS	BASE	cons alpha
205	CD	12/13	อ-อ้ง	O ANG	อ	CONS	BASE	cons alpha
206	CE	12/14	ฮ-นุกฮุก	HO NOK HUK	ฮ	CONS	BASE	cons alpha
207	CF	12/15	ไปยาลน้อย	PAI YAN NOI	๗	CONS	BASE	punct
208	DD	13/00	สระอะ	SARA A	ะ	NON	BASE	vowel alpha
209	D1	13/01	ไม้หันอากาศ	MAI HAN-AKAT	๘	AV2	ABOVE	vowel alpha dead
210	D2	13/02	สระอา	SARA AA	า	FV1	BASE	vowel alpha
211	D3	13/03	สระอำ	SARA AM	ำ	FV1	BASE	vowel alpha
212	D4	13/04	สระอิ	SARA I	ิ	AV1	ABOVE	vowel alpha dead
213	D6	13/05	สระอี	SARA II	ี	AV3	ABOVE	vowel alpha dead
214	D6	13/06	สระเอ	SARA UE	เ	AV2	ABOVE	vowel alpha dead
215	D7	13/07	สระเออ	SARA UEE	เอ	AV3	ABOVE	vowel alpha dead
216	D8	13/08	สระอุ	SARA U	ู	BV1	BELOW	vowel alpha dead
217	D9	13/09	สระอู	SARA UU	ู	BV2	BELOW	vowel alpha dead
218	DA	13/10	พินทุ	PHINTHU	๘	BD	BELOW	vowel alpha dead
219	DB	13/11		(This position must not be used)	.	NON	NONDISP	
220	DC	13/12	แยกคำไทย	THAI WORD BREAK	.	NON	NONDISP	punct
221	DD	13/13		(This position must not be used)	.	NON	NONDISP	
222	DE	13/14		(This position must not be used)	.	NON	NONDISP	
223	DF	13/15	บาท	BAHT	฿	NON	BASE	punct
224	FD	14/00	สระเอ	SARA E	เ	LV	BASE	vowel alpha
225	E1	14/01	สระแอม	SARA AE	แ	LV	BASE	vowel alpha
226	E2	14/02	สระโอ	SARA O	โ	LV	BASE	vowel alpha
227	E3	14/03	สระอไมมีวัน	SARA AI MAI MIJAN	ไ	LV	BASE	vowel alpha
228	E4	14/04	สระอ	SARA AI MAI MALAI	ใ	LV	BASE	vowel alpha
			ไม้หันตาย					
229	E5	14/05	ลาขางยาว	LAK KHANG YAO	๗	FV2	BASE	vowel alpha
230	E6	14/06	ไม้ยมก	MAI YAMOK	๗	NON	BASE	punct
231	E7	14/07	ไม้ไผ่หัว	MAI TAI KHU	๘	AD2	TOP	alpha dead
232	E8	14/08	ไม้เอก	MAI EK	.	TONE	TOP	tone alpha dead
233	E9	14/09	ไม้โท	MAI THO	.	TONE	TOP	tone alpha dead
234	EA	14/10	ไม้ตรี	MAI TRI	.	TONE	TOP	tone alpha dead
235	EB	14/11	ไม้จัตวา	MAI CHATTAWA	.	TONE	TOP	tone alpha dead
236	EC	14/12	พินทุฆค	THANTHAKHAT	๘	AD1	TOP	alpha dead
237	ED	14/13	นิกหิต	NIKHAHT	.	AD1	TOP	alpha dead
238	EE	14/14	ยามักการ	YAMAKKAN	.	AD3	TOP	alpha dead
239	EF	14/15	ฟองมัน	FONGMAN	๘	NON	BASE	punct
240	FO	15/00	เลขศูนย์	LEK SUN	๐	NON	BASE	digit

ตารางที่ 4 อักขระเฉพาะตัวของอักขระต่าง ๆ ในชุดพื้นฐาน (ต่อ)

Dec.	Hex.	Pos.	ชื่อ	Name	Char.	Class	Level	Properties
241	F1	15/01	เลขหนึ่ง	LEK NUNG	๑	NON	BASE	digit
242	F2	15/02	เลขสอง	LEK SONG	๒	NON	BASE	digit
243	F3	15/03	เลขสาม	LEK SAM	๓	NON	BASE	digit
244	F4	15/04	เลขสี่	LEK SI	๔	NON	BASE	digit
245	F5	15/05	เลขห้า	LEK HA	๕	NON	BASE	digit
246	F6	15/06	เลขหก	LEK HOK	๖	NON	BASE	digit
247	F7	15/07	เลขเจ็ด	LEK CHET	๗	NON	BASE	digit
248	F8	15/08	เลขแปด	LEK FAET	๘	NON	BASE	digit
249	F9	15/09	เลขเก้า	LEK KAO	๙	NON	BASE	punct
250	FA	16/10	อักขระคั่น	ANGKHANKHU	๙	NON	BASE	punct
251	FB	15/11	โหน้	KHOMUT	๙	NON	BASE	punct
252	FC	15/12		(This position must not be used)		NON	NONDISP	
253	FD	15/13		(This position must not be used)		NON	NONDISP	
254	FE	15/14		(This position must not be used)		NON	NONDISP	
255	FF	15/15		(This position must not be used)		CTRL	NONDISP	

ข้อมูลในตารางที่ 4 สามารถนำไปสร้างเป็นฟังก์ชันทดสอบคุณสมบัติของอักขระในภาษาโปรแกรมต่าง ๆ ได้ และมีความหมายดังนี้

- TACisalpha (c) ใช้ตามตารางที่ 4
- TACisalnum (c) คือกรณีที่ isalpha (c) หรือ isdigit (c) เป็น true
- TACisctrl (c) ใช้ตามตารางที่ 4
- TACisdigit (c) ใช้ตามตารางที่ 4
- TACisgraph (c) ใช้ตามตารางที่ 4 กรณี chlevel ไม่เป็น NONDISP ยกเว้น space
- TACislower (c) a-z (เทียบเท่า islower มาตรฐาน)
- TACisupper (c) A-Z (เทียบเท่า isupper มาตรฐาน)
- TACisprintx (c) ใช้ตามตารางที่ 4 กรณี chlevel ไม่เป็น NONDISP
- TACisSPACE (c) เหมือน ANSI C ได้แก่ space, formfeed, newline, carriage return, tab, vertical tab และ แยกคำ (wordbreak)
- TACisxdigit (c) 0-9, A-F, a-f ไม่ใช่เลขไทย (เทียบเท่า isxdigit มาตรฐาน)

4.4 ข้อกำหนดทั่วไปเกี่ยวกับการแสดงผลภาษาไทย

แม้ว่าคำวาทหลักภาษาไทยต่าง ๆ จะบัญญัติอักษรวิธภาษาไทยไว้ชัดเจนว่าวิธีการเขียน (แสดงผล) ภาษาไทยที่ถูกต้องเป็นอย่างไร แต่ก็ยังไม่เคยมีการกำหนดวิธีการแสดงผลของคอมพิวเตอร์ไว้อย่างเป็นทางการว่าเมื่อใช้กับชุดอักษรมาตรฐานแล้วจะให้ปฏิบัติอย่างไรจึงจะถูกต้องหรือตรงกันตามข้อตกลงใด ๆ ที่ชัดเจน วัตถุประสงค์ของมาตรฐานข้อนี้จึงได้กำหนดวิธีการแสดงผลให้กับความเป็นไปได้ของสายอักขระที่อาจจะเกิดขึ้นได้ทุกกรณี โดยใช้หลักการกว้าง ๆ ดังนี้

- ๑ กำหนดวิธีการแสดงผลถูกต้องตรงตามหลักภาษาไทย สำหรับกรณีที่มีข้อมูลถูกต้องตามหลักภาษาไทย
- ๒ กำหนดวิธีการแสดงผลที่ถูกต้องตามหลักภาษาไทยให้กับสัญลักษณ์ที่ใช้น้อยหรือมักใช้กันในระบบคอมพิวเตอร์ภาษาไทยต่าง ๆ อย่างไม่ถูกต้อง เช่น (พิณฑ) (นิคหิต) (ไม้ไค้) (ยามักกร)
- ๓ กำหนดวิธีการแสดงผลที่ชัดเจนและง่ายต่อการค้นหาที่ผิดพลาด สำหรับกรณีที่มีข้อมูลเข้าระบบแสดงผลผิดหลักอักษรวิธภาษาไทย (เช่น สายอักขระ ' ' จะถูกแสดงเป็น ' ' ซึ่งต่างกับคำว่า ' ' ซึ่งต้องป้อนตามลำดับ ' ' เสมอ เป็นต้น)

จากการวิเคราะห์หลักภาษาไทย พบว่าการกำหนดวิธีแสดงผลอักขระตัวปัจจุบัน (C_n) สามารถจัดแสดงได้เพียง 2 วิธีการคือ

- (ก) วางในเซลล์แสดงผลตำแหน่งถัดไป (next position)
- (ข) นำไปรวมกับอักขระในเซลล์แสดงผลเซลล์ที่แล้ว

โดยการตัดสินใจว่าจะเป็นวิธีการใดขึ้นอยู่กับคุณสมบัติของตัวปัจจุบัน (C_n) และคุณสมบัติของอักขระตัวที่แล้ว (C_{n-1}) หรือ ตัวตั้ง เท่านั้น คุณสมบัติที่ว่านี้

ได้แก่ ประเภทของอักขระ (chtype) ของอักขระคู่ประชิดทั้งสองนั่นเอง

สำหรับเนื้อหาของสูตรการตัดสินใจว่าจะวางอักขระตัวปัจจุบัน(หรือตัวตาม)ไว้ ณ ตำแหน่งใดเมื่อพิจารณาตามคุณสมบัติของอักขระตัวที่แล้ว(ตัวตั้ง)นั้น พิจารณาได้ดังนี้

- ๑ หากตัวตามเป็นอักขระประเภท ตัวเดินหน้า ให้แสดงผลในเซลล์ต่อไปได้ทันที
- ๒ หากตัวตามเป็นอักขระประเภท ตัวหยุด ให้พิจารณาดูประเภท (chtype) ของตัวนำก่อน และตัดสินใจตามผลของการวางแสดงวิธีการแสดงผลภาษาไทย (รูปที่ 4) ถ้าคำตอบเป็น C แปลว่าให้นำตัวตามไปรวมกับข้อความในเซลล์ที่ผ่านมา แต่ถ้าคำตอบเป็น N แสดงว่าเกิดปัญหาในการแสดงผลตัวหยุดนั้น จำเป็นต้องแสดงผลอักขระตัวตามในเซลล์ถัดไป

ตัวตั้ง C _n	ตัวตาม C _n	FORWARD						DEAD										
		2 CONS	3 LV	4 FV1	5 FV2	6 FV3	0 CTRL	7 BV1	8 BV2	9 BD	10 TONE	11 AD1	12 AD2		13 AD3	14 AV1	15 AV2	16 AV3
FORWARD	1 NON	N	N	N	N	N	X	N	N	N	N	N	N	N	N	N	N	NON COMPOSIBLE
	2 CONS	N	N	N	N	N	X	C	C	C	C	C	C	C	C	C	C	พยัญชนะ
	3 LV	N	N	N	N	N	X	N	N	N	N	N	N	N	N	N	N	โน้ต
	4 FV1	N	N	N	N	N	X	N	N	N	N	N	N	N	N	N	N	สระ
	5 FV2	N	N	N	N	N	X	N	N	N	N	N	N	N	N	N	N	ฯ
	6 FV3	N	N	N	N	N	X	N	N	N	N	N	N	N	N	N	N	ฯฯ
DEAD	0 CTRL	N	N	N	N	N	X	N	N	N	N	N	N	N	N	N	N	
	7 BV1	N	N	N	N	N	X	N	N	N	C	C	N	N	N	N	N	
	8 BV2	N	N	N	N	N	X	N	N	N	C	N	N	N	N	N	N	
	9 BD	N	N	N	N	N	X	N	N	N	N	N	N	N	N	N	N	
	10 TONE	N	N	N	N	N	X	N	N	N	N	N	N	N	N	N	N	
	11 AD1	N	N	N	N	N	X	N	N	N	N	N	N	N	N	N	N	(บังคับ) (กำหนด)
	12 AD2	N	N	N	N	N	X	N	N	N	N	N	N	N	N	N	N	(ไม่บังคับ)
	13 AD3	N	N	N	N	N	X	N	N	N	N	N	N	N	N	N	N	(ยามักการ)
	14 AV1	N	N	N	N	N	X	N	N	N	C	C	N	N	N	N	N	
	15 AV2	N	N	N	N	N	X	N	N	N	C	N	N	N	N	N	N	
16 AV3	N	N	N	N	N	X	N	N	N	C	N	C	N	N	N	N		

- N = เมื่อตัวตามตามหลังตัวตั้ง กำหนดให้แสดงตัวตามในช่องถัดไป
- C = เมื่อตัวตามตามหลังตัวตั้ง กำหนดให้แสดงตัวตามรวมกับตัวตั้งในช่องเดิมของตัวตั้ง
- X = ไม่แสดงผลใด

รูปที่ 4 วิธีการแสดงอักขระพื้นฐานโดยพิจารณาจากประเภทของตัวตามและตัวตั้ง (ข้อ 4.4)

4.5 วิธีการป้อนข้อมูล

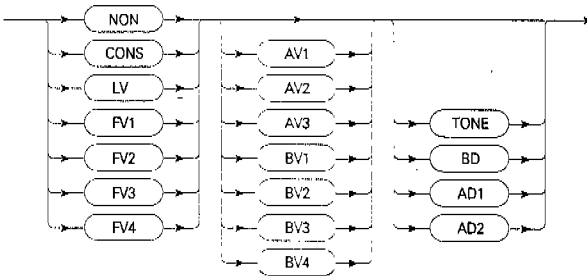
4.5.1 ข้อกำหนดพื้นฐาน

การป้อนข้อมูลภาษาไทยเป็นการป้อนอักขระทางแป้นพิมพ์ภาษาไทยเป็นละ 1 อักขระ (direct mapping method) เช่นเดียวกับภาษาอังกฤษ ไม่ว่าอักขระนั้น ๆ จะเป็นตัวเดินหน้า หรือตัวหยุดอยู่กับที่ ทั้งนี้ป้อนสัญลักษณ์แต่ละตัวจากซ้ายไปขวาโดยไม่คำนึงถึงวิธีการออกเสียงใด ๆ เช่น คำว่า เกาะ จะถูกป้อนตามลำดับ ก-ก-า-ะ (แม้ว่ากลุ่มสัญลักษณ์ ก-ะ รวมกันจะมีเสียงเป็น สระเอาะ โดยยังไม่มีการอื่นใดสำหรับป้อนสระเอาะ) สำหรับกรณี ที่สัญลักษณ์ที่ต้องรวมกับพยัญชนะในเซลล์เดียวกัน เช่น BV, AV, AD, BD หรือ TONE ให้ปฏิบัติตามลำดับดังนี้

- ก. ทุกเซลล์ต้องเริ่มด้วยสัญลักษณ์ในระดับฐานก่อนเสมอ และเฉพาะกลุ่ม CONS เท่านั้นที่มีสิทธิให้สัญลักษณ์อื่นมาประสม
- ข. สัญลักษณ์ในข้อ ก. อยู่ในกลุ่ม CONS อาจจะถูกตามด้วยอักขระประเภท AV1-AV3 และ BV1-BV2 หรือไม่มีอักขระประเภทนี้

ค. ถ้าสัญลักษณ์ในข้อ ก. อยู่ในกลุ่ม CONS (โดยอาจจะมีหรือไม่มีสัญลักษณ์ในข้อ ข. ก็ได้) อาจตามด้วยวรรณยุกต์ (TONE) หรือเครื่องหมายกำกับเสียงประเภท BD, AD1, AD2 ลำดับขั้นตอนที่กล่าวมานี้ เป็นสูตรพื้นฐานที่ช่วยในการกลั่นกรองความผิดพลาดเบื้องต้นของข้อมูล และสามารถสรุปได้เป็นผังวากยสัมพันธ์ตามรูปที่ 5

คุณสมบัติของการป้อนข้อความภาษาไทยนี้สามารถรับประกันได้ว่าลำดับอักขระที่ผิดไปจากข้อกำหนดนี้ผิดหลักภาษาไทยแน่นอน แต่ไม่ได้รับประกันว่าการป้อนตามลำดับนี้มีความถูกต้องเพียงใด ในทำนองเดียวกับระบบการป้อนภาษาอังกฤษซึ่งไม่มีการตรวจสอบความถูกต้องใด ๆ ของข้อความที่ถูกป้อนเข้าไป ตัวอย่างเช่น ระบบการรับข้อมูลจะอนุญาตให้ป้อนข้อความที่ไร้ความหมาย เช่น กบาวธรรฯ ๆ เข้าไปได้ แต่จะระงับการป้อนข้อความ " ฅ ฅ จังหวะที่กำลังป้อน " เพื่อให้ผู้ใช้ระบบ ดำเนินการแก้ไขความผิดพลาดโดยการโต้ตอบกับโปรแกรมประยุกต์เสียก่อน เช่น ลบไม้เอกด้วย backspace และป้อน " ตามด้วย ' เข้าไปแทนที่เป็นต้น



รูปที่ 5 ผังวากยสัมพันธ์ของการป้อนข้อความภาษาไทยลงในหนึ่งเซลล์
(ข้อ 4.5.1 และข้อ 4.6.2)

เนื่องจากส่วนรับข้อมูลอาจจะปฏิบัติงานเป็นอิสระจากส่วนแสดงผลพร้อมและอาจจะเป็นโปรแกรมแยกจากส่วนแสดงผล ดังนั้นจึงกำหนดให้ระบบรับข้อมูลกระทำการเพียงการระงับ (block) การป้อนอักขระที่ผิดวากยสัมพันธ์ตามรูปที่ 5 และหาทางรายงานความผิดพลาดนี้ให้แก่ผู้ใช้ทราบโดยการส่งเสียงหรือวิธีการอื่นใดตามความเหมาะสม (ไม่กำหนดเป็นข้อบังคับ)

4.5.2 การตรวจสอบความถูกต้องของการป้อนข้อมูล

นอกเหนือไปจากข้อกำหนดเกี่ยวกับวิธีการ "ทั่วไป" ในการป้อนข้อมูลข้างต้นแล้ว การใช้เงื่อนไขทางตรรกะ (logical condition) มาช่วยกลั่นกรองให้ข้อมูลภาษาไทยถูกต้องยิ่งขึ้น นับว่าเป็นสิ่งจำเป็นสำหรับการรักษาความถูกต้องของข้อมูลต่าง ๆ ดังนั้นมาตรฐานนี้จึงกำหนดให้โปรแกรมประยุกต์เมื่อทำงาน (ร่วมกับส่วนควบคุมภาษาไทย) แล้วมีคุณสมบัติในการตรวจสอบความถูกต้องของการป้อนข้อมูลได้อย่างเคร่งครัดขึ้น โดยให้มี API ฟังก์ชันสำหรับกำหนดระดับการตรวจสอบความถูกต้องของการป้อนข้อมูลได้ 3 ระดับคือ

(1) การตรวจสอบระดับ 0

ได้แก่ การรับการตรวจสอบความถูกต้อง (passthrough) เพื่อให้โปรแกรมประยุกต์ทำหน้าที่รับผิดชอบในการตรวจสอบทั้งหมด ทั้งโปรแกรมประยุกต์จะต้องส่งผ่านฟังก์ชันการควบคุมแบบพิมพ์เพื่อให้ระบบเข้าสู่ สภาพนี้ และโปรแกรมประยุกต์เริ่มทำหน้าที่ควบคุมการตรวจสอบตัวเอง อนึ่ง เมื่อสิ้นสุดการทำงานของโปรแกรมประยุกต์ โปรแกรมจะต้องเปลี่ยนระดับการตรวจสอบกลับคืนเป็นสภาพปกติ (ระดับ 1)

(2) การตรวจสอบระดับ 1

ได้แก่การตรวจสอบความถูกต้องขั้นพื้นฐาน(basic check)ตามข้อ 4.5.1 โดยโปรแกรมประยุกต์อาจ จะเสริมการทดสอบเพิ่มเติมได้ตามความเหมาะสม ระดับนี้เป็น default ของระบบภาษาไทยทั่วไป

(3) การตรวจสอบระดับ 2

ได้แก่ การตรวจสอบที่เคร่งครัดขึ้น โดยใช้เงื่อนไขทางตรรกของภาษาไทยที่กำหนดเพิ่มเติมจากที่กำหนดไว้ในข้อ 4.5.1 โดยเพิ่มเงื่อนไขต่าง ๆ เท่าที่จะจัดทำได้ในเชิงกลุ่มประเภทอักขระดังสรุปไว้ในรูปที่ 6 และคำอธิบายในรูปที่ 7 การตรวจสอบในระดับนี้จะช่วยเพิ่มความถูกต้องของข้อมูลขึ้นอีกมากกล่าวคือส่วนที่ถูกระงับการป้อนเป็นส่วนที่ผิดพลาดแน่นอน ส่วนที่รับเข้ามาอาจยังมีข้อความที่ไร้ความหมายได้เช่นเดียวกับระดับ 1 แต่จะมีสถิติความผิดพลาดน้อยลง รูปที่ 8 แสดงผังวากยสัมพันธ์ของการป้อนอักขระไทยที่ตรงกับการตรวจสอบในระดับนี้

ในการสอบถามลักษณะความถูกต้องของอักขระคู่ประชิด (ตัวตั้ง/ตัวตาม) สามารถสอบถามได้จากฟังก์ชัน TAC_composable (chn, chn1) ซึ่งมีค่ารหัสของตัวตาม (chn) และตัวตั้ง (chn1) เป็นพารามิเตอร์ โดยจะคืนค่าเป็น char มีค่าเป็น "A", "C", "S", "R" หรือ "X" ตามความหมายในรูปที่ 6 และ 7 ตามลำดับ

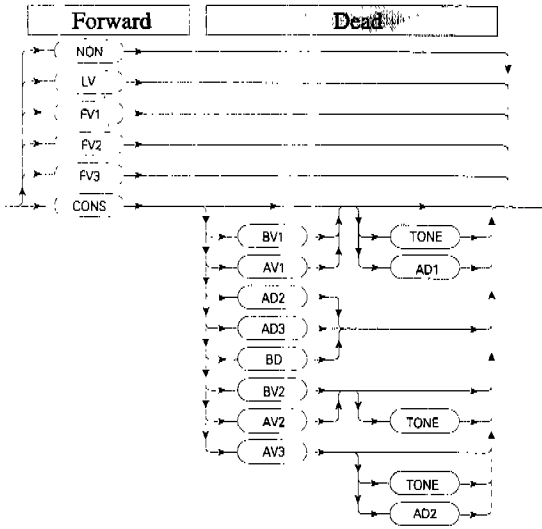
ตัวตั้ง C _{n-1}	ตัวตาม C _n	FORWARD									DEAD									
		2 CONS	3 LV	4 FV1	5 FV2	6 FV3	0 CTRL	7 BV1	8 BV2	9 BD	10 TONE	11 AD1	12 AD2	13 AD3	14 AV1	15 AV2		16 AV3		
FORWARD	1 NON	A	A	A	S	S	A	X	R	R	R	R	R	R	R	R	R	R	R	NON COMPOSIBLE
	2 CONS	A	A	A	A	S	A	X	C	C	C	C	C	C	C	C	C	C	C	พยางค์
	3 LV	S	A	S	S	S	S	X	R	R	R	R	R	R	R	R	R	R	R	แน้ว
	4 FV1	S	A	A	S	S	A	X	R	R	R	R	R	R	R	R	R	R	R	ะว่า
	5 FV2	A	A	A	A	S	A	X	R	R	R	R	R	R	R	R	R	R	R	ำ
	6 FV3	A	A	A	S	A	S	X	R	R	R	R	R	R	R	R	R	R	R	ำ ๆ
DEAD	0 CTRL	A	A	A	A	A	A	X	R	R	R	R	R	R	R	R	R	R	R	
	7 BV1	A	A	A	S	S	A	X	R	R	R	C	C	R	R	R	R	R	R	
	8 BV2	A	A	A	S	S	A	X	R	R	R	C	R	R	R	R	R	R	R	
	9 BD	A	A	A	S	S	A	X	R	R	R	R	R	R	R	R	R	R	R	(พันธุ)
	10 TONE	A	A	A	A	A	A	X	R	R	R	R	R	R	R	R	R	R	R	
	11 AD1	A	A	A	S	S	A	X	R	R	R	R	R	R	R	R	R	R	R	(นิคคิด) (หันขมาค)
12 AD2	A	A	A	S	S	A	X	R	R	R	R	R	R	R	R	R	R	R	(ไม้ค)	
13 AD3	A	A	A	S	S	A	X	R	R	R	R	R	R	R	R	R	R	R	(ยามักกา)	
14 AV1	A	A	A	S	S	A	X	R	R	R	C	C	R	R	R	R	R	R		
15 AV2	A	A	A	S	S	A	X	R	R	R	C	R	R	R	R	R	R	R		
16 AV3	A	A	A	S	S	A	X	R	R	R	C	R	C	R	R	R	R	R		

รูปที่ 6 หลักการกลั่นกรองความถูกต้องของกระบวนการป้อนข้อมูล โดยพิจารณาจากอักขระปัจจุบัน (ตัวตาม) และอักขระตัวที่แล้ว (ตัวตั้ง) (ข้อ 4.5.2(3))

ผล	ความหมาย	ความหมายของผลลัพธ์ในระดับการตรวจสอบต่าง ๆ			ลักษณะการแสดงผลของระบบแสดงผล
		ระดับ 0	ระดับ 1	ระดับ 2	
A	accept	รับ	รับ	รับ	ในเซลล์ถัดไป
C	accept (compose)	รับ	รับ	รับ	ในเซลล์เดียวกับตัวตั้ง
S	strict mode reject	รับ	รับ	ไม่รับ	ในเซลล์ถัดไป
R	reject	รับ	ไม่รับ	ไม่รับ	ในเซลล์ถัดไป
X	accept (non-display)	รับ	รับ	รับ	ไม่แสดงผลอักขระ

รูปที่ 7 ความหมายของผลลัพธ์จากรูปที่ 6 (ข้อ 4.5.25(3))

ข้อกำหนดในข้อนี้ไม่ได้ครอบคลุมถึงการจัดผังตัวอักษรบนแป้นพิมพ์ ซึ่งได้กำหนดไว้ใน มอก.820 และไม่ได้กำหนดลักษณะการป้อนข้อมูลวิธีอื่นที่นอกเหนือไปจากวิธี direct mapping method ซึ่งอาจจะเป็นวิธีการป้อนข้อความภาษาไทยแบบใหม่ ๆ ซึ่งอาจจะมีประสิทธิภาพสูงกว่าปัจจุบันก็ได้



รูปที่ 8 ผังวางสัมพันธ์ของการป้อนข้อความภาษาไทยระดับที่ 2
(ข้อ 4.5.2(3))

4.6 วิธีแสดงผล

4.6.1 สายอักขระไทย-อังกฤษ

สายอักขระไทย-อังกฤษที่สามารถนำมาแสดงผลตามมาตรฐานนี้ให้มีค่าจำกัดความตรงตามรหัสของชุดอักขระพื้นฐานดังแสดงในรูปที่ 1 เท่านั้น ในการแสดงผลจากสายอักขระรหัสอื่นใด เช่น รหัสเลขศร หรือ EBCDIC-มอก. ฯลฯ ให้ผ่านกระบวนการแปลงรหัสเป็นรหัสชุดอักขระพื้นฐานเสียก่อน สายอักขระตามรหัสชุดนี้ให้ถือว่าเป็นรหัสมาตรฐานสำหรับการแลกเปลี่ยนข้อมูลตามมาตรฐาน

4.6.2 การรวมสัญลักษณ์ระดับอื่น ๆ เข้ากับระดับฐาน

อักขระที่จัดอยู่ในประเภท 'ตัวหยุดอยู่กับที่' สามารถนำมารวมแสดงผลในเซลล์เดียวกับอักขระประเภทพยัญชนะ (consonant) โดยมีหลักการใกล้เคียงกับผังวางสัมพันธ์ของการป้อนข้อมูลใน 1 คอลัมน์ไทย (ดูรูปที่ 5) อย่างไรก็ตามจากการค้นหาความเป็นไปได้ทั้งหมดของการรวมสัญลักษณ์ประเภท DEAD กับ CONS พบว่า

- 4.6.2.1 ในเซลล์แสดงผลภาษาไทยมีสัญลักษณ์ได้ไม่เกิน 3 ตัว นั่นคือในหนึ่งเซลล์อาจจะมีสัญลักษณ์ 1 หรือ 2 หรือ 3 ตัว ประสมกัน
- 4.6.2.2 ในเซลล์หนึ่งเซลล์จะต้องมีสัญลักษณ์จากกลุ่ม NON, CONS หรือ LV, FV อย่างน้อย 1 ตัวและหากเริ่มด้วย CONS อาจประสมกับอักขระจากกลุ่ม dead อีก 0 ตัว 1 ตัวหรือ 2 ตัวก็ได้

- 4.6.2.3 สำหรับส่วนที่มาจากกลุ่ม dead จำนวน 0-2 ตัว มีการรวมตัวกันได้เพียง 51 แบบ ดังแสดงในรูปที่ 9 ซึ่งในมาตรฐานนี้ถือว่าระบบแสดงผลทั่วไปควรมีความสามารถในการแสดงผลได้ไม่น้อยกว่า 47 แบบ (คือ อนุโลมให้ยกเว้นการแสดงผล 4 แบบ ได้แก่ สระอ+นิคหิต สระอ+นิคหิต สระอ+ไม้ได้คู่ และ สระอ+ไม้ได้คู่) สำหรับผู้ที่พัฒนาระบบภาษาไทยขึ้นมาใหม่ หรือใช้วิธีการทางซอฟต์แวร์ล้วน ต้องมีความสามารถในการแสดงครบทั้ง 51 แบบ

ความสามารถในการแสดงผลอย่างครบถ้วนในยุคหลังการทดลองใช้มาตรฐานนี้อาจจะเป็นสิ่งจำเป็นสำหรับประเทศไทย ซึ่งในอดีตเราได้ถูกข้อจำกัดของเทคโนโลยีเครื่องพิมพ์ตีตกทำให้สัญลักษณ์ที่ใช้กันมาถึงช่วงรัตนโกสินทร์ตอนกลางถูกยกเลิกไปโดยปริยาย (เช่น โคมุตร ฟองมัน อังคั่นคู่ ฟินทุ ยามักการ) เมื่อมีคอมพิวเตอร์ที่ดีขึ้น ก็ได้มีผู้พัฒนาให้สามารถใช้ประโยชน์ได้กับภาษาไทยที่สมบูรณ์ เช่น การบันทึกพระไตรปิฎกโดยมหาวิทยาลัยมหิดล การบันทึกกฎหมายตราสามดวงโดยมหาวิทยาลัยเกียวกโด และการบันทึกข้อความ รวมชื่อบาลีของพระสงฆ์ในระบบข้อมูลของธนาคารแห่งประเทศไทย เป็นต้น ดังนั้น การกำหนดลักษณะการแสดงผลที่ถูกต้องสมบูรณ์ตาม มอก.620พร้อมหลักการป้อนข้อมูลที่ถูกต้องจะช่วยให้ข้อมูลภาษาไทยในฐานะข้อมูลต่าง ๆ ได้มาตรฐานที่ถูกต้องและตรงกัน

- 4.6.3 การเข้ารหัสของเซลล์แสดงผลภาษาไทย-อังกฤษ

ในมาตรฐานนี้ได้มีการกำหนดโครงสร้างข้อมูลชื่อ THAI_CELL สำหรับใช้สื่อสารระหว่างระบบแสดงผล และโปรแกรมประยุกต์โดยข้อมูลนี้ประกอบด้วยส่วนต่าง ๆ 4 ไบต์ ได้แก่ mid, hilo, attr และ style ซึ่ง mid และ hilo มีค่าจำกัดความดังนี้

mid หมายถึง รหัสอักษรของสัญลักษณ์ใด ๆ ที่แสดงในระดับฐาน (baselevel) ตามข้อกำหนดของรหัสชุดอักขระพื้นฐาน

hilo หมายถึง มีข้อกำหนดวิธีการเข้ารหัสสำหรับส่วนที่เป็น dead character จำนวน 0-2 ตัว รวมกันเป็น 1 ไบต์ ตามความหมายและสูตรการเข้ารหัสตามรูปที่ 9

TACVloadfrom(id) ;
TACVsaveto(mode, id) .

4.6.4 การแสดงผลในกรณีที่สายอักขระผิดอักขรวิธีมาตรฐาน

ข้อกำหนดส่วนนี้เป็นการบรรยายพฤติกรรมของคอมพิวเตอร์เมื่อมีการนำสายอักขระภาษาไทยที่ไม่ตรงตามอักขรวิธีมาแสดงบนจอภาพหรือทางเครื่องพิมพ์ ซึ่งถือว่าเป็นกรณีที่ไม่ปกติ ซึ่งแม้จะไม่ได้เกิดขึ้นบ่อย แต่ก็มีผลกระทบต่อความรู้สึกของผู้ใช้คอมพิวเตอร์ ดังนั้นจึงต้องมีการกำหนดให้ระบบภาษาไทยมีแนวปฏิบัติที่ตรงกันด้วย

คำว่า "ผิดอักขรวิธี" ในที่นี้หมายถึงสายอักขระซึ่งมีสัญลักษณ์คู่ประชิดซึ่งเข้าข่ายการตีความเป็น R หรือ S ตามรูปที่ 6 และรูป 7 ซึ่งต้องแสดงผลโดยการวางลงในเซลล์ถัดไปทางขวามือพร้อมกับเคลื่อนตัวชี้ตำแหน่งไปทางขวามือ 1 เซลล์ หลักการที่กำหนดนี้จะช่วยให้ระบบภาษาไทยทุกระบบแสดงผลของข้อความที่ผิดอักขรวิธีออกมาเหมือนกัน เพราะใช้สูตรเดียวกัน นอกจากนี้การกำหนดให้แสดงอักขระที่ผิดในคอลัมน์ถัดไปจะช่วยให้ผู้ใช้สามารถมองเห็นความผิดพลาดดังกล่าวบนจอภาพ (หรือบนกระดาษพิมพ์) ได้อย่างชัดเจนเสมอ ดังตัวอย่างในรูปที่ 10

สายอักขระ										
ท	ล	ล	ล	ล	ล	ล	ล	ล	ล	ล
ผลลัพธ์ (จะเป็นแบบ ก. หรือ ข. ก็ได้)										
ก. วางอักขระที่ผิดพลาดในระดับบรรทัด										
ท	ล	ล	ล	ล	ล	ล	ล	ล	ล	ล
ข. วางอักขระที่ผิดพลาดในระดับปกติของอักขระนั้น										
ท	ล	ล	ล	ล	ล	ล	ล	ล	ล	ล

รูปที่ 10 ตัวอย่างการตีความตามรูปที่ 6 และรูปที่ 7

(ข้อ 4.6.4)

4.7 การตีความอักขระควบคุมในสายอักขระไทยที่ถูกส่งมาทางจอภาพที่กำลังปฏิบัติหน้าที่เป็นเครื่องพิมพ์ดีด หรือ โทรพิมพ์ (teletype หรือ TTY)

อักขระควบคุมบางตัวจะต้องได้รับการตีความอย่างถูกต้องตามข้อกำหนดในตารางที่ 5

ความหมายของโทรพิมพ์ที่ไม่ได้ครอบคลุมถึงการทำงานของจอภาพแบบเต็มจอ (full screen editing) หรือ block mode แต่อย่างใด

ตารางที่ 5 การตีความอักขระควบคุมขณะที่เป็นเครื่องพิมพ์ดีดหรือโทรพิมพ์
(ข้อ 4.7)

รหัสควบคุม	การปฏิบัติ
BEL (bell)	ไม่มีการเปลี่ยนแปลงบนจอภาพ แต่มีเสียงบีบลิ้น ทุออกทางลำโพง
BS (backspace)	เลื่อนตัวชี้ตำแหน่งไปทางซ้ายมือ 1 ช่อง ยกเว้นกรณีที่ช่องทางซ้ายมือเป็นขอบซ้ายของหน้าต่างแสดงผลตัวชี้ตำแหน่งจะหยุดกับที่ และกรณีที่มีข้อความในเซลล์ที่ผ่านมามีสัญลักษณ์บนล่างของไทย ก็ให้หยุดอยู่กับที่และนับขีดเขยจำนวนสัญลักษณ์บน-ล่าง จนครบถ้วน (ขึ้นอยู่กับว่ามี BS เข้ามากี่ตัว) จึงจะเคลื่อนตัวชี้ตำแหน่งได้ในการตีความ BS ไม่ต้องลบสัญลักษณ์ได้ออกจากจอภาพ (non-destructive backspace)
HT (horizontal tab)	เติม space ไปจนกระทั่งตัวชี้ตำแหน่งเลื่อนไปยังคอลัมน์ตำแหน่งหยุด tab อันถัดไป ตำแหน่งหยุด tab มาตรฐานคือทุก ๆ 8 คอลัมน์นับจากขอบซ้ายของช่องแสดงผล เช่น 8, 16, 24, 32... (คอลัมน์แรกสุดหมายเลข 0)
LF (line feed)	เลื่อนตัวชี้ตำแหน่งลงไปยังล่าง 1 บรรทัด หากอยู่ที่บรรทัดล่างสุดอยู่แล้วให้เลื่อนภายในหน้าต่างแสดงผล 1 บรรทัด และตัวชี้ตำแหน่งอยู่ตำแหน่งเดิม
FF (form feed)	ลบหน้าต่างแสดงผลและเคลื่อนตัวชี้ตำแหน่งไปยังตำแหน่ง home (มุมซ้ายบนของช่องแสดงผล)
CR (carriage return)	เลื่อนตัวชี้ตำแหน่งไปยังช่องทางซ้ายสุดของหน้าต่างแสดงผลอักขระควบคุมอื่น ๆ ละทิ้งไปโดยไม่มีการเปลี่ยนแปลงใด ๆ บนจอภาพ

4.8 การตรวจแก้ข้อความ (text editing)

การทำงานขั้นพื้นฐานของการตรวจแก้ข้อความ ที่จะต้องตรงกันมีเพียงส่วนของการเลื่อนตัวชี้ตำแหน่งและการลบอักษรเท่านั้น โดยโปรแกรมประยุกต์ที่มีการแก้ไขข้อความในเขตของข้อมูลหรือในการประมวลค่า ต้องจัดทำหน้าที่การเลื่อนตัวชี้ตำแหน่งไปมาที่ละเซลล์ แต่ถ้ามีการลบข้อความทางซ้ายมือของตัวชี้ตำแหน่งให้จัดการลบอักขระบนจอที่ละอักขระ หากลบข้อความ ณ ตำแหน่งของตัวชี้ตำแหน่งให้ลบที่ละเซลล์ ตามตารางที่ 6

ตารางที่ 6 ข้อกำหนดพื้นฐานสำหรับการตรวจแก้ข้อความ
(ข้อ 4.8)

หน้าที่	การปฏิบัติ
forward cell	เคลื่อนตัวชี้ตำแหน่งไปทางขวาทีละเซลล์
backward cell	เคลื่อนตัวชี้ตำแหน่งไปทางซ้ายทีละเซลล์
delete current cell	ลบอักขระในเซลล์ที่มีตัวชี้ตำแหน่งอยู่ครั้งละหนึ่งเซลล์
*(delete next cell)	(ไม่สนใจว่ามีอักขระในเซลล์นั้น)
delete previous char.	ลบอักขระในเซลล์ที่อยู่ทางซ้ายมือของตัวชี้ตำแหน่งไปครั้งละหนึ่งอักขระ

หมายเหตุ * หากตัวชี้ตำแหน่งแสดงเป็นขีดแนวดิ่ง (vertical bar insertion point) มีความหมายเป็นการลบเซลล์ถัดไป (delete next cell)

ในกรณีที่จัดทำโปรแกรมประยุกต์สำหรับตรวจแก้ข้อความโดยทำงานกับอุปกรณ์แป้นพิมพ์แบบ enhanced keyboard หรือเทียบเท่า ให้กำหนดความหมายของปุ่ม (keytop) ตรงกับหน้าที่ตามตารางที่ 7

ตารางที่ 7 ข้อกำหนดแป้นพิมพ์สำหรับการตรวจแก้ข้อความ
(ข้อ 4.8)

ปุ่ม	หน้าที่
→	forward cell
←	backward cell
delete	delete current cell
backspace	delete previous char.

4.9 ประเด็นสำหรับการพิจารณา ร่วมกับมาตรฐานอื่น

เอกสาร *การกำหนดหลักเกณฑ์โครงสร้างของตัวอักษรไทย* โดยคณะกรรมการพิจารณาร่าง “หลักเกณฑ์โครงสร้างของตัวอักษรไทย” ที่เสนอต่ออนุกรรมการบัญญัติศัพท์วิชาการพิมพ์ ราชบัณฑิตยสถาน และจัดพิมพ์ในลักษณะร่างเพื่อเผยแพร่ก่อนการประกาศในหนังสือ “วันการพิมพ์ไทย 2534” ซึ่งจัดพิมพ์โดยมูลนิธิเงินทุนงานแสดงการพิมพ์แห่งประเทศไทย ได้มีข้อกำหนดต่าง ๆ เกี่ยวกับลักษณะของตัวอักษรไทย (Thai glyphs) ซึ่งจะมีประโยชน์อย่างยิ่งในการออกแบบชุดแบบอักษรไทย (Thai font) สำหรับคอมพิวเตอร์และการพิมพ์ต่าง ๆ ในเอกสารฉบับนี้ มีประเด็นที่น่าสนใจและควรแก่การพิจารณาดังต่อไปนี้

4.9.1 อักษรไทยที่ใช้ในการพิมพ์แต่ยังไม่มีการกำหนดรหัส

ได้แก่ ตัวฝนทองพองมัน (หรือพองมันพินหนุ) มีรูปร่างเป็นพองมัน (©) เสริมด้วยขีดตรง ๆ 2 ขีด ในแนวตั้งในระดับสระบน (๖) อักษรนี้หากจัดเข้าในรหัส TACTIS แล้วจะเป็นอักขระในประเภท NON และมีสถานะเป็นอักขระพิเศษ แสดงผลในระดับ BASE

- 4.9.2 อักขระอื่น ๆ ที่ใช้ในการพิมพ์ ซึ่งอาจกำหนดรหัสในแบบอักษรหรือ Code Page อื่น ๆ ได้แก่อักขระต่อไปนี้
- อัญประกาศคู่เปิด (“)
 - อัญประกาศคู่ปิด (”)
 - เครื่องหมายไม่เท่ากับ (≠)
 - เครื่องหมายเพราะฉะนั้น (∴)
 - เครื่องหมายเพราะว่า (∵)
 - เครื่องหมายบุพសัญญา (,,)

อักขระเหล่านี้ไม่ใช่อักขระไทย และมีใช้งานในระบบการพิมพ์ด้วยคอมพิวเตอร์โดยอาศัยกลไกการเปลี่ยนแบบอักษร ดังนั้นจึงไม่จำเป็นต้องกล่าวถึงในมาตรฐานนี้

- 4.9.3 การประสมสระล่างเข้ากับพยัญชนะที่มีเชิงหรือหางล่าง

ให้วางได้พยัญชนะเหล่านั้นได้ในกรณีของ ญ หญิง และ รุ ฐาน อนุญาตให้ตัดเชิงได้แต่ไม่เป็นการบังคับดังตัวอย่าง

ญ ญ ญ จุ รุ กตัญญูตา วัญญู

- 4.9.4 การประสมพินทุเข้ากับพยัญชนะที่มีเชิงหรือหางล่าง

ให้ปฏิบัติเช่นเดียวกับข้อ 4.9.3 ดังตัวอย่าง ญ จุ

ภาคผนวก ก.

แฟ้ม TIS.H

(ข้อ 3.3.2 และข้อ 4.3.4)

```
/* NECTEC THAI SOFTWARE STANDARD PROJECT 1989-1991
TIS.H
```

This include file defines the names and character codes of
 (1) Standard ASCII control codes,
 (2) Thai Industrial Standards characters (ASCII TIS 620),
 (3) Dot Matrix Printer Code Extension (TIS 988),
 The names are written as defined in the future announcement of ISO
 8859-XX for Thailand.

LAST UPDATE: 18 FEBRUARY 1991

*/

```
/* PART 1 - STANDARD ASCII CODES (ISO-646) */
```

```
#define NULL 0
#define NUL 0
#define SOH 1 /* 01 START OF HEADING */
#define STX 2 /* 02 START OF TEXT */
#define ETX 3 /* 03 END OF TEXT */
#define EOT 4 /* 04 END OF TRANSMISSION*/
#define ENQ 5 /* 05 ENQUIRE */
#define ACK 6 /* 06 ACKNOWLEDGE */
#define BEL 7 /* 07 BELL */
#define BS 8 /* 08 BACKSPACE */
#define HT 9 /* 09 HORIZONTAL TAB */
#define LF 10 /* 0A LINE FEED */
#define VT 11 /* 0B VERTICAL TAB */
#define FF 12 /* 0C FORM FEED */
#define CR 13 /* 0D CARRIAGE RETURN */
#define SO 14 /* 0E SHIFT OUT */
#define SI 15 /* 0F SHIFT IN */
#define DLE 16 /* 10 DATA LINK ESCAPE */
#define DC1 17 /* 11 DEVICE CONTROL 1 */
#define XON DC1 /* 11 X-ON */
#define DC2 18 /* 12 DEVICE CONTROL 2 */
#define DC3 19 /* 13 DEVICE CONTROL 3 */
#define XOFF DC3 /* 13 X-OFF */
#define DC4 20 /* 14 DEVICE CONTROL 4 */
#define NAK 21 /* 15 NEGATIVE ACKNOWLEDGE */
#define SYN 22 /* 16 SYNCHRONOUS IDLE */
#define ETE 23 /* 17 END OF TRANSMISSION BLOCK */
#define CAN 24 /* 18 CANCEL */
#define EM 25 /* 19 END OF MEDIUM */
```

```
#define SUB 26 /* 1A SUBSTITUTE CHARACTER */
#define ESC 27 /* 1B ESCAPE */
#define FS 28 /* 1C FILE SEPARATOR */
#define GS 29 /* 1D GROUP SEPARATOR */
#define RS 30 /* 1E RECORD SEPARATOR */
#define US 31 /* 1F UNIT SEPARATOR */
#define IS4 28 /* 1C INFORMATION SEPARATOR 4 */
#define IS3 29 /* 1D INFORMATION SEPARATOR 3 */
#define IS2 30 /* 1E INFORMATION SEPARATOR 2 */
#define IS1 31 /* 1F INFORMATION SEPARATOR 1 */
#define SP 32 /* 20 SPACE */
#define DEL 127 /* 7F DELETE */
```

/* PART 2 - THAI INDUSTRIAL STANDARD - TIS 620
NAMING CONVENTION AS TISI PROPOSED TO ISO/ECMA FOR DRAFT ISO 8859-XX FOR
THAILAND.

UPDATED 18 FEB 91

```
/*
#define KOKAI 161 /* n A1 */
#define KHOKHAI 162 /* v A2 */
#define KHOKHUAT 163 /* v A3 */
#define KHOKHWAI 164 /* n A4 */
#define KHOKHON 165 /* n A5 */
#define KHORAKHANG 166 /* u A6 */
#define NGONGU 167 /* i A7 */
#define CHOCHAN 168 /* v A8 */
#define CHOCHING 169 /* n A9 */
#define CHOCHANG 170 /* r AA */
#define SOSO 171 /* y AB */
#define CHOCHOE 172 /* w AC */
#define YOYING 173 /* y AD */
#define DOCHADA 174 /* g AE */
#define TOPATAK 175 /* g AF */
#define THOTHAN 176 /* j B0 */
#define THONANGMONTHO 177 /* w B1 */
#define THOPHUTHAO 178 /* n B2 */
#define NONEN 179 /* w B3 */
#define DODEK 180 /* n B4 */
#define TOTAO 181 /* n B5 */
#define THOTHUNG 182 /* n B6 */
#define THOTHAHAN 183 /* n B7 */
#define THOTHONG 184 /* t B8 */
#define NONU 185 /* u B9 */
#define BOBAIMAI 186 /* u BA */
#define POPLA 187 /* j BB */
#define PHOPHUNG 188 /* n BC */
```

```

#define FOFA 189 /* # BD */
#define PHOPHAN 190 /* # BE */
#define FOFAN 191 /* # BF */
#define PHOSAMPHAO 192 /* # C0 */
#define MOMA 193 /* # C1 */
#define YOYAK 194 /* # C2 */
#define RORUA 195 /* # C3 */
#define RU 196 /* # C4 */
#define LOLING 197 /* # C5 */
#define LU 198 /* # C6 */
#define WOWAEN 199 /* # C7 */
#define SOSALA 200 /* # C8 */
#define SORUSI 201 /* # C9 */
#define SOSUA 202 /* # CA */
#define HOHIP 203 /* # CB */
#define LOCHULA 204 /* # CC */
#define OANG 205 /* # CD */
#define HONOKHUK 206 /* # CE */
#define PAIYANNOI 207 /* # CF */
#define SARA_A 208 /* # D0 */
#define MAIHUNAKAT 209 /* # D1 */
#define SARA_AA 210 /* # D2 */
#define SARA_AM 211 /* # D3 */
#define SARA_I 212 /* # D4 */
#define SARA_II 213 /* # D5 */
#define SARA_UE 214 /* # D6 */
#define SARA_UEE 215 /* # D7 */
#define SARA_U 216 /* # D8 */
#define SARA_UU 217 /* # D9 */
#define PHINTHU 218 /* # DA */
#define BAHT 223 /* # DB */
#define SARA_E 224 /* # E0 */
#define SARA_AE 225 /* # E1 */
#define SARA_O 226 /* # E2 */
#define MAIMUAN 227 /* # E3 */
#define MAIMALAI 228 /* # E4 */
#define LAKKHANGYAO 229 /* # E5 */
#define MAIYAMOK 230 /* # E6 */
#define MAITAIKHU 231 /* # E7 */
#define MAIEK 232 /* # E8 */
#define MAITHO 233 /* # E9 */
#define MAITRI 234 /* # EA */
#define MAICHATTAWA 235 /* # EB */
#define THANTHAKHAT 236 /* # EC */
#define NIKHAHIT 237 /* # ED */
#define YAMAKKAN 238 /* # EE */

```

NO. 1566-2541

```
#define FONGMAN      239 /* e EF */
#define LEKSUN       240 /* o F0 */
#define LEKNUNG      241 /* a F1 */
#define LEKSONG      242 /* w F2 */
#define LEKSAM       243 /* a F3 */
#define LEKSI        244 /* d F4 */
#define LEKHA        245 /* z F5 */
#define LEKHOK       246 /* b F6 */
#define LEKCHET      247 /* n F7 */
#define LEKPAET      248 /* z F8 */
#define LEKKAO       249 /* t F9 */
#define ANGKHANKHU   250 /* u FA */
#define KHOMUT       251 /* w FB */
```

```
/* PART 3 - THAI INDUSTRIAL STANDARDS EXTENDED FOR PRINTER */
/* TIS 988 */
```

```
#define MAIHUNAKAT_EK 0x80 /* TIS 988 */
#define SARA_I_EK     0x84 /* TIS 988 */
#define SARA_I_THANTHAKHAT 0x88 /* TIS 988 */
#define SARA_II_EK    0x89 /* TIS 988 */
#define SARA_UE_EK    0x8D /* TIS 988 */
#define SARA_UEE_EK   0x91 /* TIS 988 */
#define PR_AM_EK      0x95 /* TIS 988 */
#define PR_BOTTOM_O   0xDB /* TIS 988 */
#define PR_TOP_O      0xDC /* TIS 988 */
#define PR_TOP_MAIMUAN 0xDD /* TIS 988 */
#define PR_TOP_MAIMALAI 0xDE /* TIS 988 */
```

```
/*.....END OF TIS.H.....*/
```