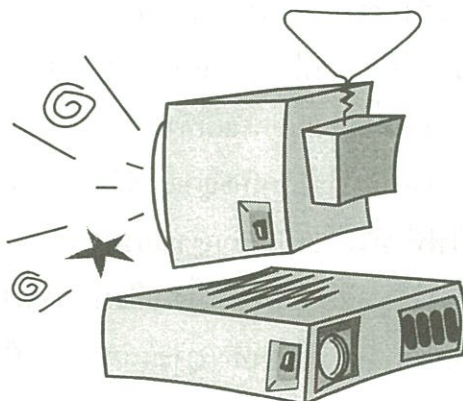


คอมพิวเตอร์จำตัวอักษรได้อย่างไร



*ศิริภัตรา เหมือนมัลัย



ปัจจุบันการป้อนข้อมูลไม่ว่าจะเป็นภาษาไทย หรือภาษาใด ๆ ก็ตามลงในเครื่องคอมพิวเตอร์ ส่วนใหญ่ก็ต้องใช้การเคาะแป้นพิมพ์ซึ่งเป็นการเสียเวลา เสียแรงงานและยังเสียค่าใช้จ่ายสูงมากอีกด้วย การแก้ปัญหาดังกล่าวสามารถทำได้วิธีหนึ่ง คือ การทำให้คอมพิวเตอร์จำตัวอักษรที่ใช้ในภาษานั้น ๆ โดยให้คอมพิวเตอร์สามารถอ่านตัวอักษรที่พิมพ์หรือเขียนบนกระดาษได้โดยตรง และให้ความเร็วในการอ่านพอกับการถ่ายเอกสาร

ก็จะเป็นการประหยัดแรงงาน และค่าใช้จ่ายในการป้อนข้อมูลเข้าคอมพิวเตอร์ได้อย่างมากมาย

คอมพิวเตอร์เป็นเครื่องอิเล็กทรอนิกส์ ดังนั้น ข้อมูลที่จะส่งเข้าไปในคอมพิวเตอร์ ต้องเป็นสัญญาณไฟฟ้า ซึ่งมีอยู่ ๒ แบบ คือ สัญญาณแบบ Analog มีลักษณะเป็นสัญญาณต่อเนื่องติดต่อกันไปเรื่อยๆ กับสัญญาณแบบ Digital ที่มีลักษณะเป็นสัญญาณไม่ต่อเนื่อง การเปลี่ยนรูปร่าง

* หัวหน้าแผนกบริการคอมพิวเตอร์ สำนักวิจัยและพัฒนา มหาวิทยาลัยธุรกิจบัณฑิตย์ :
วท.ม. จุฬาลงกรณ์มหาวิทยาลัย

ลักษณะของตัวอักษรที่ปรากฏอยู่บนกระดาษจะเป็นสัญญาณแบบ Digital ซึ่งอาศัยหลักการเดียวกับเครื่องถ่ายภาพเอกสาร คือ ใช้แสงส่องลงบนกระดาษที่มีข้อความ ถ้าแสงตกกระทบบนตัวอักษรที่เป็นสีดำ แสงสะท้อนกลับจะลดปริมาณลงมากกว่าแสงที่ตกกระทบบนพื้นกระดาษสีขาว จากนั้นจะมีอุปกรณ์ที่ทำหน้าที่เปลี่ยนปริมาณของแสงให้เป็นสัญญาณไฟฟ้า แต่สัญญาณไฟฟ้าที่ได้จะเป็นสัญญาณแบบ Analog จึงต้องใช้อุปกรณ์อีกชิ้นหนึ่งเปลี่ยนสัญญาณแบบ Analog ให้เป็นสัญญาณแบบ Digital อีกต่อหนึ่ง เพื่อแทนลักษณะรูปร่างของตัวอักษร แล้วนำไปเก็บไว้ในหน่วยความจำของคอมพิวเตอร์ เพื่อให้คอมพิวเตอร์จำตัวอักษรนั้น ๆ ต่อไป

ส่วนอุปกรณ์ที่ใช้ในการเปลี่ยนลักษณะรูปร่างตัวอักษรหรือรูปร่างของรูปภาพใด ๆ ให้เป็นสัญญาณไฟฟ้า มีอยู่ด้วยกันหลายชนิด ได้แก่ เครื่องโทรสาร (Facimile), (Image Scanner) กล้องวิดิทัศน์ เป็นต้น นอกจากนี้ยังมีอุปกรณ์ที่ใช้เขียนเพื่อให้สัญญาณไฟฟ้าแก่คอมพิวเตอร์อีก ได้แก่ ปากกาแสง หรือเรียกว่า Light Pen, Mouse เป็นต้น ซึ่งอุปกรณ์ดังกล่าวราคาถูกลงและหาซื้อได้ง่ายในปัจจุบัน

จากอดีตจนถึงปัจจุบันการพัฒนาเพื่อ

ให้คอมพิวเตอร์จำรูปแบบจะแบ่งออกได้เป็น ๒ ลักษณะ คือ คอมพิวเตอร์ที่จำตัวอักษร (Character Recognition) และคอมพิวเตอร์จำรูปภาพ (Picture Recognition) ขั้นตอนการทำงานของคอมพิวเตอร์ทั้ง ๒ ลักษณะสามารถแบ่งออกได้ดังนี้ ขั้นแรกจะเป็นการใช้เทคโนโลยีในการเปลี่ยนลักษณะของภาพไปเป็นสัญญาณไฟฟ้า เช่น เทคโนโลยีของกล้องโทรทัศน์ จากนั้นขั้นที่สองจะเป็นการใช้เทคโนโลยีในการปรับปรุงคุณภาพสัญญาณไฟฟ้าของภาพที่ได้ แล้วจัดเก็บไว้ในหน่วยความจำของคอมพิวเตอร์เพื่อให้ใช้เนื้อที่ของหน่วยความจำให้น้อยที่สุด เช่น เทคโนโลยีของการทำ data reduction เพื่อลดข้อมูล และเทคโนโลยีของการทำ data construction เพื่อสร้างข้อมูลกลับมาใหม่ แล้วทำการประมวลผลสัญญาณ หรือที่เรียกว่า Signal Processing จึงมาทำขั้นที่สามซึ่งเป็นการทำให้คอมพิวเตอร์จำหรือเข้าใจสัญญาณภาพในการทำให้คอมพิวเตอร์จำตัวอักษรนั้นสามารถแบ่งให้เป็นสองรูปแบบ คือ จำแบบทันทีทันใด (on-line recognition) และจำแบบไม่ทันทีทันใด (off-line recognition) การจำแบบทันทีทันใดคอมพิวเตอร์จะวิเคราะห์ว่าสัญญาณข้อมูลนั้นเป็นตัวอักษรอะไร โดยทำไปพร้อม ๆ กับการรับสัญญาณข้อมูลเข้ามา ส่วนการจำแบบ

ไม่ทันที่ทันใด คอมพิวเตอร์จะรับสัญญาณไฟฟ้า ซึ่งเป็นข้อมูลที่แทนลักษณะรูปร่างของตัวอักษร นำมาเก็บไว้ในหน่วยความจำ แล้วจึงทำการวิเคราะห์สัญญาณเหล่านั้นว่าเป็นตัวอักษรอะไรต่อไป การจำทั้งสองแบบจะใช้กับตัวอักษรตัวพิมพ์และตัวอักษรลายมือเขียน แต่การจำแบบทันทีทันใด มักใช้กับการจำตัวอักษรลายมือเขียนเท่านั้น นอกจากนี้แล้ว เทคโนโลยีการเปลี่ยนลักษณะภาพไปเป็นสัญญาณไฟฟ้า แล้วจัดเก็บในหน่วยความจำของคอมพิวเตอร์ที่มีให้เห็นเพิ่มขึ้นเรื่อย ๆ ได้แก่ เครื่อง Image Scanner, เครื่องโทรสาร (Facimile), ปากกาแสง ตลอดจนเครื่องตรวจสอบแบบปรนัย, เครื่องอ่านรหัสแถบ (Bar Code Reader) โดยเฉพาะอย่างยิ่งการจำตัวอักษร ก็มีเครื่องป้อนข้อมูลอัตโนมัติที่รู้จักกันในชื่อของ OCR หรือ Optical Character Reader ซึ่งเครื่อง OCR นี้ใช้กับลายมือเขียนนิยมใช้ในประเทศญี่ปุ่น อาจเป็นเพราะภาษาญี่ปุ่น นำมาสร้างเครื่องพิมพ์ดีดไฟฟ้าได้ยาก และราคาสูงมาก คนญี่ปุ่นจะต้องใช้การเขียนเป็นหลัก

สำหรับประเทศไทยมีการสั่งซื้อผลิตภัณฑ์เกี่ยวกับเรื่องนี้มาใช้มากขึ้น เช่น เครื่องตรวจสอบแบบปรนัย (Optical Mark Reader) หรือที่เรียกว่า OMR, เครื่อง Image

Scanner, เครื่องอ่าน Bar Code, เครื่องโทรสาร และปากกาแสง เป็นต้น

ผู้เขียนจำได้ว่าศูนย์อิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ หรือที่รู้จักกันในนามของ NECTEC ได้เคยทำการวิจัยเพื่อสร้าง card ดัชนีแบบในการให้คอมพิวเตอร์จำตัวพิมพ์อักษรภาษาไทย ที่ใช้ร่วมกับไมโครคอมพิวเตอร์ โดยข้อมูลตัวอักษรภาษาไทยที่พิมพ์บนกระดาษธรรมดาเข้าไปผ่านเครื่อง Image Scanner เพื่อตรวจกวาดภาพ จากนั้นข้อมูลที่ได้อาจจะถูกจัดเก็บไว้ในหน่วยความจำของคอมพิวเตอร์ แล้วจะมีระบบการทำงานในโปรแกรมบน card จะเรียกมาประมวลผลวิเคราะห์ว่าเป็นตัวอักษรอะไร แล้วให้รหัส ASC II ส่งออกมานำไปเก็บในหน่วยความจำของคอมพิวเตอร์เพื่อนำไปใช้งานต่อไป

เห็นได้ว่าเครื่อง Image Scanner เครื่องโทรสาร กล้องวิดีโอ ตลอดจนอุปกรณ์ที่ใช้ในการอ่านข้อมูล ล้วนมีราคาถูกลงมาก มีผู้นำมาใช้กันมากขึ้น เพื่อให้อุปกรณ์ดังกล่าวมีประโยชน์มากขึ้น การนำเทคโนโลยีการจำตัวอักษรมาผนวกเข้าไป จึงทำให้การใช้อุปกรณ์ดังกล่าวแพร่หลายมากขึ้นอย่างรวดเร็ว เทคโนโลยีการจำตัวอักษรจึงเป็นเทคโนโลยีของการสร้างเครื่องตรวจสอบ (Optical Mark Reader) และเครื่องอ่านรหัส

แถบ (Bar Code Reader) ประเทศไทย อนาคตของอุตสาหกรรมที่ใช้เทคโนโลยีมี
 สามารถสร้างเครื่องเหล่านี้ขึ้นใช้เอง ทำให้ เพิ่มขึ้น จึงน่าจะมีผลให้เศรษฐกิจของประเทศ
 ราคาเครื่องถูกลงและมีผู้นิยมใช้กันมากขึ้น เจริญขึ้นตามไปด้วย □□
 สำหรับประเทศไทย แนวโน้มใน

บรรณานุกรม

- Gore, Marvin R, and Stubbe, John W. Computers and data processing. NewYork : McGraw-Hall, 1979.
- Orilia, Lawrence S. Computers and information. Singapore : McGraw-Hill, 1982.