หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

# บทที่ 1
# บทนำ

ในยุคปัจจุบันเป็นยุคที่ข่าวสารข้อมูลมีความสำคัญและมีบทบาทต่อชีวิตเราอย่างมาก การเก็บข้อมูลต่าง ๆ ในด้านธุรกิจ ด้านอุตสาหกรรม ด้านการแพทย์ ด้านกฎหมาย จำเป็นที่จะต้องใช้หน่วยเก็บข้อมูลจำนวนมาก หน่วยเก็บข้อมูลที่มีใช้ในปัจจุบันได้แก่ จานแม่เหล็ก (Floppy Disk) เทปแม่เหล็ก (Magnetic Tape) จานแสง (Optical Disk) เป็นต้น

ถึงแม้ว่าเทคโนโลยีทางด้านอุปกรณ์เก็บข้อมูลได้มีการพัฒนาให้มีประสิทธิภาพในการเก็บข้อมูลได้มากขึ้น แต่ผู้ใช้มักจะประสบปัญหาอุปกรณ์เก็บข้อมูลที่มีขนาดไม่เพียงพอกับความต้องการในการใช้งานด้านต่าง ๆ ที่มีแนวโน้มเพิ่มขึ้นอยู่ตลอดเวลา

การแก้ปัญหาที่สามารถทำได้คือ การเพิ่มอุปกรณ์เก็บข้อมูลให้กับเครื่องคอมพิวเตอร์แต่ก็อาจมีปัญหาเรื่องขีดจำกัดด้านความสามารถของเครื่อง หรือปัญหาเรื่องค่าใช้จ่ายที่เพิ่มขึ้นทำให้ไม่สามารถเพิ่มอุปกรณ์เก็บข้อมูลให้กับเครื่องคอมพิวเตอร์ได้อีก การแก้ปัญหาสามารถทำได้อีกทางหนึ่งคือ การลดขนาดของข้อมูลที่จัดเก็บในหน่วยเก็บข้อมูลให้มีขนาดเล็กลง เพื่อให้หน่วยเก็บข้อมูลสามารถเก็บข้อมูลได้มากขึ้น วิธีการนี้จะมีผลเท่ากับการเพิ่มความจุให้กับหน่วยเก็บข้อมูล และมีผลพลอยได้ที่ตามมาคือ อัตราการรับ-ส่งของข้อมูลที่ถูกลดขนาดลงแล้วระหว่างหน่วยความจำกับหน่วยเก็บข้อมูลจะสูงกว่าอัตราการรับ-ส่งของข้อมูลปกติ ดังนั้น ถ้านำวิธีการลดขนาดข้อมูลมาใช้ในลักษณะเวลาจริง (Real Time) จะทำให้ความเร็วในการรับ-ส่งข้อมูลของหน่วยเก็บข้อมูลเพิ่มขึ้น และช่วยลดปริมาณงานช่องด้านเข้า/ด้านออก (Input/Output Channel) ของระบบคอมพิวเตอร์อีกด้วย

นอกจากนี้ การลดขนาดข้อมูลยังสามารถนำไปใช้ประโยชน์เกี่ยวกับงานด้านการสื่อสารข้อมูล ซึ่งปัจจุบันได้มีการขยายเครือข่ายกว้างขึ้น จึงทำให้มีการรับส่งข้อมูลภายในเครือข่ายเพิ่มมากขึ้น ดังนั้นถ้ามีการลดขนาดข้อมูลก่อนส่งผ่านเครือข่าย จะมีผลทำให้ค่าใช้จ่ายของการสื่อสารลดลง เพราะใช้เวลาในการส่งข้อมูลน้อยลงและเป็นการรักษาความลับของข้อมูลอีกด้วยเนื่องจากข้อมูลจะถูกเปลี่ยนรูปแบบรหัสก่อนส่งผ่านเครือข่าย

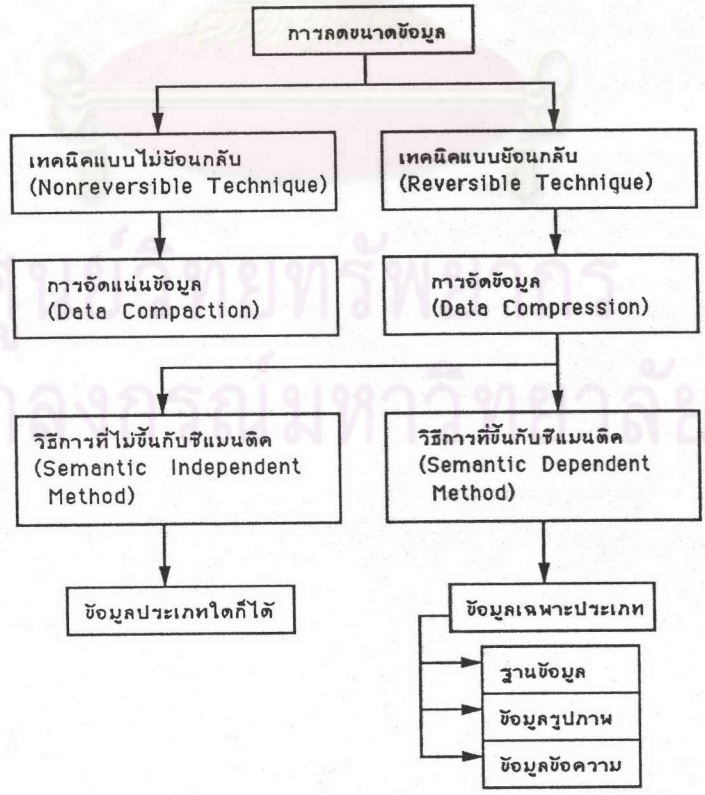
จากแนวโน้มที่เพิ่มขึ้นของข่าวสารข้อมูลและการสื่อสารข้อมูลระหว่างผู้ใช้ ทำให้การลดขนาดข้อมูลมีความสำคัญและมีบทบาทมากขึ้นเรื่อย ๆ ปัจจุบันจึงได้มีการพัฒนาอัลกอริทึมเดิม หรือพยายามคิดค้นอัลกอริทึมใหม่ ๆ เพื่อเพิ่มความสามารถในการลดขนาดข้อมูล

## 1.1 การลดขนาดข้อมูล

การลดขนาดข้อมูลสามารถทำได้ 2 เทคนิค คือ เทคนิคแบบไม่ย้อนกลับ

(Nonreversible Techniques) และเทคนิคการแบบย้อนกลับ (Reversible Techniques) เทคนิคแบบไม่ย้อนกลับ ข้อมูลจะถูกลดขนาดลงโดยการตัดสิ่งที่ไม่จำเป็นทิ้งไปเช่น การตัดช่องว่างหลาย ๆ ตัวให้เหลือเพียงตัวเดียว ที่เรียกว่าเทคนิคแบบไม่ย้อนกลับ เพราะข้อมูลที่ถูกลดขนาดลงจะอยู่ในรูปแบบใหม่ตลอดไปโดยไม่สามารถย้อนกลับคืนเป็นข้อมูลเดิมได้อีก บางครั้งจะเรียกว่าการอัดแน่นข้อมูล (Data Compaction) สำหรับเทคนิคแบบย้อนกลับเป็นเทคนิคที่เก็บทุกส่วนของข้อมูลเดิมไว้ทั้งหมด แต่เป็นการเก็บที่ลดขนาดลงโดยการเข้ารหัส (Encoding) การลดขนาดข้อมูลโดยลักษณะนี้เรียกว่า การอัดข้อมูล (Data Compression) เมื่อต้องการข้อมูลเดิมกลับมาจะต้องทำการถอดรหัส (Decoding) ขั้นตอนในการย้อนกลับนี้เรียกว่า การขยายข้อมูล (Data Decompression)

การอัดข้อมูลแบ่งออกเป็น 2 วิธีคือ วิธีการที่ไม่ขึ้นกับซีแมนติค (Semantic Independent Method) และวิธีการที่ขึ้นกับซีแมนติค (Semantic Dependent Method) วิธีการที่ไม่ขึ้นกับซีแมนติคเป็นวิธีการที่พัฒนาขึ้นโดยไม่สนใจลักษณะของข้อมูล สามารถนำไปใช้กับข้อมูลใด ๆ ก็ได้ ส่วนวิธีการที่ขึ้นกับซีแมนติค จะพัฒนาวิธีการจากการศึกษาลักษณะเฉพาะที่มีในข้อมูลและนำมาใช้ให้เป็นประโยชน์ในการลดขนาดข้อมูล ข้อมูลที่เหมาะสมสำหรับวิธีการอัดข้อมูลแบบนี้ ได้แก่ ข้อมูลรูปภาพ (Image Data) ฐานข้อมูล (Data Base) และข้อมูลข้อความ (Text Data) (Reghbati 1981 : 71)




รูปที่ 1.1 วิธีการลดขนาดข้อมูล

การพัฒนาวิธีการอัดข้อมูลที่ขึ้นกับชีแมนติค อาจพัฒนาวิธีการที่เป็นเทคนิคเฉพาะสำหรับใช้ในการอัดข้อมูลประเภทนั้นเท่านั้น หรืออาจนำวิธีการใดวิธีการหนึ่งของวิธีการอัดข้อมูลที่ไม่ขึ้นกับชีแมนติคมาพัฒนาต่อเพื่อให้มีประสิทธิภาพเพิ่มขึ้น นอกจากนี้วิธีการดังกล่าวที่ได้รับการพัฒนาแล้วก็ยังสามารถนำไปใช้ในการอัดข้อมูลอื่น ๆ ได้ตามปกติ

## 1.2 ลำดับชั้นของวิธีการอัดข้อมูล (Classification of Data Compression Methods)

การอัดข้อมูลวิธีแรกเริ่มต้นจากแนวความคิดของ ชานนอน ในปี ค.ศ.1948 และฟาโน ในปี ค.ศ.1949 ต่อมาในปี ค.ศ. 1952 ฮัฟแมนได้ปรับปรุงวิธีการบางส่วนใหม่ให้เป็นวิธีการที่รับประกันได้ว่า สามารถลดความเหลือเฟือได้ต่ำสุด (Minimum Redundancy) หมายความว่า สามารถกำจัดส่วนที่ไม่จำเป็นออกไปจากข้อมูลได้มากที่สุด วิธีการนี้จึงได้รับการยอมรับอย่างมาก แม้ในปัจจุบันนี้ก็ยังมีการนำไปใช้งานในโปรแกรมอัดข้อมูลหลาย ๆ โปรแกรมเช่น โปรแกรมจัดเก็บแฟ้มข้อมูลพีเคแวร์ (PKWARE File Archives) โปรแกรมอรรถประโยชน์คอมเพลส (Compress Utility) ของโปรแกรมจัดการระบบยูนิกซ์ เป็นต้น วิธีการที่คิดขึ้นต่อมาเป็นวิธีการเชิงคำนวณที่นำหลักการของอัลกอริทึมฮัฟแมนมาปรับใหม่ ให้ใช้ช่วงค่าของเลขจำนวนจริงระหว่าง 0 ถึง 1 ในการสร้างรหัสใหม่แทนการใช้ทรี (Tree)

วิธีการอัดข้อมูลที่มีในระยะแรก ๆ นั้น เป็นวิธีการที่คิดค้นขึ้นโดยใช้หลักการตามทฤษฎีของสารสนเทศ มีการอ่านข้อมูล 2 รอบ อ่านข้อมูลรอบแรกเพื่อนับจำนวนครั้งในการใช้งานของรหัสแล้วนำค่าที่นับได้มาใช้ในการสร้างรหัสใหม่ และอ่านข้อมูลรอบที่สองเพื่อเข้ารหัสข้อมูลด้วยรหัสใหม่ที่สร้างขึ้น รหัสใหม่ที่ใช้จะมีรูปแบบเดิมไม่เปลี่ยนแปลงทุกครั้งที่ปรากฏในการเข้ารหัสข้อมูล เราจึงเรียกวิธีการอัดข้อมูลในลักษณะนี้ว่า วิธีการคงที่ (Static Method) วิธีการนี้เรียกอีกชื่อหนึ่งว่า วิธีการกำหนดรูปแบบคงที่ (Static Defined-word Scheme Method)

สำหรับวิธีการอัดข้อมูลวิธีการใหม่ ๆ มักเป็นวิธีการที่อ่านข้อมูลเพียงรอบเดียว ในขณะอ่านข้อมูล อัลกอริทึมการเข้ารหัสจะศึกษาลักษณะของข้อมูลและสร้างรหัสใหม่ไปพร้อม ๆ กับเข้ารหัสข้อมูล รหัสใหม่ที่ใช้สำหรับการเข้ารหัสจะมีการเปลี่ยนแปลงตลอดเวลา เราเรียกวิธีการอัดข้อมูลในลักษณะนี้ว่า วิธีการอแดปทีฟ (Adaptive Method) วิธีการที่จัดอยู่ในกลุ่มนี้ ได้แก่ วิธีการของอแดปทีฟฮัฟแมน วิธีการของแลมเพล-ซิฟ วิธีการของบีเอสทีดับบลิว

อัลกอริทึมอแดปทีฟฮัฟแมน โดยหลักการแล้วยังเป็นวิธีการที่ใช้หลักการตามทฤษฎีของสารสนเทศ เพียงแต่เปลี่ยนวิธีการจากการอ่านข้อมูล 2 รอบเป็นการอ่านข้อมูลรอบเดียว สำหรับอัลกอริทึมแลมเพล-ซิฟ และอัลกอริทึมบีเอสทีดับบลิวเป็นวิธีการอัดข้อมูลที่สนใจความเหลือเฟือของข้อมูลในลักษณะอื่น ๆ เช่น กลุ่มรหัสที่ใช้งานบ่อย การเกิดซ้ำรหัส เป็นต้น แทนที่จะสนใจความเหลือเฟือในลักษณะการใช้งานมากน้อยเพียงอย่างเดียว

หลักการของวิธีการอัดข้อมูลแต่ละวิธีโดยย่อ ๆ ได้แก่

1. วิธีการของชานนอน-ฟาโน (Shannon 1948 and Fano 1949 quoted in Lynch 1985 : 53) การสร้างรหัสใหม่จะได้จากการแบ่งกลุ่มรหัสที่มีการจัดเรียงอันดับตามค่าความน่าจะเป็นของการเกิดรหัส การอัดข้อมูลวิธีนี้ไม่รับประกันความเหลือเฟือต่ำสุด (Lelewer and Hirschberg 1987 : 270)

2. วิธีการของฮัฟแมน (Huffman 1952 quoted in Lynch 1985 : 55) การสร้างรหัสใหม่จะได้จากทรีที่สร้างจากความน่าจะเป็นของรหัสที่จัดเรียงอันดับแล้ว รหัสที่มีความน่าจะเป็นสูงจะได้รหัสใหม่ที่สั้น และรหัสที่มีความน่าจะเป็นต่ำจะได้รหัสใหม่ที่ยาวกว่า วิธีนี้รับประกันความเหลือเฟือต่ำสุด (Lynch 1985 : 55)

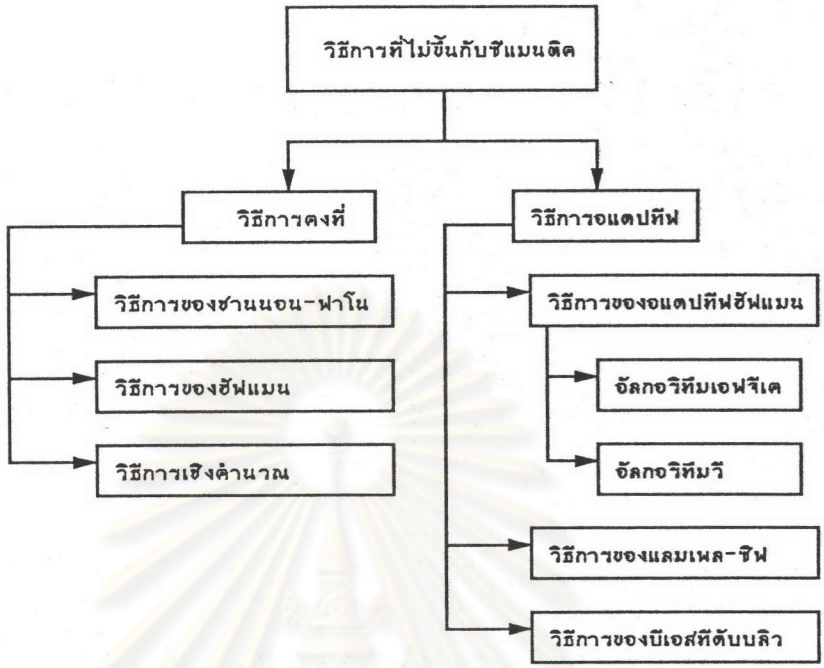
3. วิธีการเชิงคำนวณ วิธีการนี้ไม่ต้องจัดเรียงอันดับความน่าจะเป็น ค่าความน่าจะเป็นของรหัสที่อยู่ในช่วงค่าจำนวนจริง [0,1) จะถูกนำมาใช้ในการเข้ารหัสข้อมูล เนื่องจากการใช้เลขจำนวนจริงเป็นรหัส ดังนั้นวิธีการนี้มักจะมีปัญหาเรื่องความเที่ยง (Precision) (Blahut 1987 : 85-87) จึงมีการนำไปใช้กับการรับส่งข้อมูล และเปลี่ยนช่วงค่าของรหัสให้ใช้เลขจำนวนเต็มแทนเลขจำนวนจริง (Witten 1987 : 520-540)

4. วิธีการของอแดปทีฟฮัฟแมน มี 2 อัลกอริทึมคือ อัลกอริทึมเอฟจีเค (Algorithm FGK) (Knuth quoted in Lelewer and Hirschberg 1987 : 278) และอัลกอริทึมวี (Algorithm V) (Vitter 1987 : 825-845) ใช้ค่าความน่าจะเป็นของการเกิดรหัสสร้างทรีในขณะนั้น ๆ ทำให้ทรีเปลี่ยนรูปร่างอยู่ตลอดเวลา อัลกอริทึมจึงซับซ้อนมากกว่าหลักการเดิมของอัลกอริทึมฮัฟแมน ต่อมาวิทเตอร์ได้พัฒนาอัลกอริทึมวี ให้ใช้งานสำหรับการรับส่งข้อมูล (Vitter 1989 : 158-167)

5. วิธีการของแลมเพล-ซิฟ (Ziv and Lempel 1977 : 337-343) ใช้หลักการการกระจายอิสระ (Free Parsing) ในการสร้างตารางรหัสสำหรับการเข้ารหัส วิธีการนี้มีอัลกอริทึมไม่ซับซ้อนทำงานได้รวดเร็ว สามารถลดความเหลือเฟือได้ทั้งลักษณะของการเกิดซ้ำอักษรและรูปแบบที่ใช้งานมาก วิธีการนี้ได้นำมาพัฒนาต่อโดยเวล และตั้งชื่อใหม่ว่า อัลกอริทึมแอลแซดดับบลิว (Algorithm LZW) (Welch 1984 : 8-19)

6. วิธีการของบีเอสทีดับบลิว (Bentley, Sleator, Tarjan and Wei 1986 : 320-330) ใช้หลักการย้ายมาหน้า (Move to Front) ของรหัสในลิสต์ที่มีการจัดรูปแบบเอง (Self Organizing List) รหัสใดปรากฏบ่อยจะอยู่ต้นลิสต์ และรหัสใดที่ไม่ค่อยปรากฏจะอยู่ท้ายลิสต์ ซึ่งจะมีผลทำให้รหัสที่อยู่ต้นลิสต์มีรหัสใหม่ที่สั้นและรหัสที่อยู่ท้าย ๆ ลิสต์จะมีรหัสใหม่ที่ยาว

วิธีการอัดข้อมูลที่กล่าวข้างต้นเป็นวิธีการอัดข้อมูลที่ไม่ขึ้นกับชีแมนติค เราสามารถนำไปใช้ในการอัดข้อมูลประเภทใด ๆ ก็ได้ วิธีการดังกล่าวสามารถลดขนาดของข้อมูลลงได้โดยการรวบรวมลักษณะความเหลือเฟือของข้อมูลที่กำลังใช้งานอยู่ในขณะนั้น ๆ




รูปที่ 1.2 วิธีการอัดข้อมูลที่ไม่ขึ้นกับซีแมนติค

เริ่มแรกวิธีการอัดข้อมูลมีใช้เฉพาะในการสื่อสารข้อมูล เพื่อเพิ่มประสิทธิภาพของการรับ-ส่งข้อมูล ต่อมาได้มีการนำมาใช้ในงานด้านอื่น ๆ ด้วย มักเป็นงานที่จะต้องใช้หน่วยเก็บข้อมูลจำนวนมาก เช่น ข้อมูลรูปภาพ (Image Data) ฐานข้อมูล (Database) หรือข้อความ (Text) เอกสารต่าง ๆ (Document) ซึ่งในปัจจุบันมีแนวโน้มในการใช้งานเพิ่มขึ้นอย่างมาก จึงได้มีวิธีการอัดข้อมูลอีกประเภทหนึ่งคือ วิธีการที่ขึ้นกับซีแมนติค เป็นวิธีที่พัฒนาวิธีการจากรูปแบบความเหลือเฟือเฉพาะอย่างซึ่งจะมีลักษณะแตกต่างกันไปตามชนิดของข้อมูล

การอัดข้อมูลที่ขึ้นกับซีแมนติคสำหรับข้อมูลเฉพาะอย่างที่ได้รับความสนใจ และได้มีการพัฒนาวิธีการอัดข้อมูลไว้แล้ว ได้แก่

1. การอัดข้อมูลรูปภาพ วิธีการที่ใช้กัน ได้แก่ การเข้ารหัสแบบนับความยาว (Run-length Encoding) (Pountain 1987 : 317-320) อีกวิธีการหนึ่งที่ใช้กับข้อมูลรูปภาพ คือ การใช้โครงสร้างข้อมูลควอดทรี (Quadtree Data Structure) (Samet 1985 : 973-993)

2. การอัดข้อมูลฐานข้อมูล งานประมวลผลข้อมูลธุรกิจ (Business Data Processing) เป็นงานที่ต้องใช้แฟ้มข้อมูลจำนวนมากและแฟ้มข้อมูลมักจะมีขนาดใหญ่ ถ้านำการอัดข้อมูลมาใช้งานด้านนี้จะสามารถลดค่าใช้จ่ายของหน่วยเก็บข้อมูล ค่าใช้จ่ายในการรับ-ส่งข้อมูล และมีผลทำให้การติดต่อกับดิสก์ลดลงอีกด้วย ลักษณะความเหลือเฟือของงานประเภทนี้ ได้แก่ เลข 0 ในฟิลด์ตัวเลข ช่องว่างในฟิลด์อักษร เป็นต้น

วิธีการอัดข้อมูลอาจใช้วิธีการนับความยาวในการลดขนาดข้อมูลของกลุ่มเลข หรือกลุ่มช่องว่าง วิธีการสร้างสมนัยบิต (Bit Map) ไว้ต้นระเบียน (Record) ซึ่งถูกจะกำหนดเป็น 1 สำหรับเขต (Field) มีการใช้งานในระเบียนนั้นหรืออาจใช้สัญญลักษณ์พิเศษ (Special Sign) แทนกลุ่มข้อความที่มีการใช้งานในฐานข้อมูล (Ruth and Kreutzer 1972 : 62-66) ส่วนวิธีการอื่น ๆ ได้แก่ วิธีการแทนค่าดิกชันนารี (Dictionary Substitution) (Regbati 1981 : 73) วิธีการของครอแมค (Cormack 1985 : 1336-1342) เป็นวิธีที่นำไปใช้งานร่วมกับระบบไอเอ็มเอส (IMS : IBM 's Information Management System) สามารถลดการใช้เนื้อที่เก็บข้อมูลได้ถึง 50 เปอร์เซนต์

3. การอัดข้อมูลข้อความ วิธีการอัดข้อมูลข้อความที่มีกล่าวถึงในบทความทางวิชาการหรือหนังสือวิชาการนั้นเป็นวิธีการอัดข้อมูลสำหรับข้อความภาษาอังกฤษ แต่ละบทความได้เสนอวิธีการต่าง ๆ หลายวิธี ครอติซี (Cortesi 1982 : 397-403) ได้พัฒนาวิธีการจากสถิติคู่อักษรที่พบบ่อย แล้วเข้ารหัสคู่อักษรที่พบบ่อย 2 ไบท์ด้วยรหัส 1ไบท์ ถ้าคู่อักษรใดไม่ใช่คู่อักษรที่พบบ่อยก็จะเข้ารหัสด้วยรหัสของอักษรแต่ละตัว วิธีการนี้ลดขนาดข้อมูลลงได้ 35 เปอร์เซนต์ แต่สามารถลดได้สูงสุดเพียง 50 เปอร์เซนต์เท่านั้น ซอลตัน (Salton 1989 : 104-129) ได้พัฒนาวิธีการได้จากสถิติกลุ่มอักษรลงท้ายที่พบบ่อยที่รวมช่องว่างด้วยอีก 1 ตัวเช่น tion_ the_ and_ เป็นต้น วิธีการนี้ให้อัตราส่วนการอัด 50 เปอร์เซนต์ ถึง 55 เปอร์เซนต์ ฮาน (Hahn 1974 : 434-436) ได้เสนอแนวคิดในการอัดข้อมูลโดยการแทนคำทุกในแฟ้มด้วยค่าตัวเลข นอกจากนี้ยังมีวิธีการอื่นอีกหลายวิธีที่มีกล่าวถึงในบทความของรูบิน (Rubin 1976 : 617-623) บทความของทรอปเปอร์ (Tropper 1982 : 398-413)

## 1.3 แนวทางการวิจัย

การวิจัยครั้งนี้เป็นการนำวิธีการอัดข้อมูลที่ไม่ขึ้นกับซีแมนติคที่มีอยู่มาพัฒนาต่อ โดยให้มีการพิจารณาซีแมนติคภาษาไทย (Thai Semantics) ด้วย ซึ่งซีแมนติคของภาษาไทยที่ใช้ในการวิจัยจะนำมาจากงานวิจัยเรื่อง การวิเคราะห์ข้อมูลคำไทย (Thai Word Analysis) ของห้องปฏิบัติการวิจัยไมโครคอมพิวเตอร์ มหาวิทยาลัยเกษตร ที่ได้รวบรวมคำใช้บ่อยในชีวิตประจำวัน (ยืน ภู่วรวรรณ 2527 : 290-299)

วิธีการอัดข้อมูลที่นำมาใช้ในการวิจัยมี 2 วิธี คือวิธีการอัดข้อมูลของฮัฟแมน และวิธีการอัดข้อมูลของแลมเพล-ซิฟ วิธีการฮัฟแมนเป็นวิธีการอัดข้อมูลที่รับประกันความเหลือเฟือต่ำสุด สำหรับวิธีการแลมเพล-ซิฟ เป็นวิธีการอัดข้อมูลที่สามารถลดความเหลือเฟือในลักษณะของการเกิดซ้ำอักษรหรือรูปแบบที่ใช้บ่อย อัลกอริทึมของวิธีการนี้ไม่ซับซ้อนและทำงานได้รวดเร็ว

การวัดประสิทธิภาพของวิธีการการอัดข้อมูล จะใช้ค่าอัตราส่วนการอัดเป็นค่าเปรียบเทียบประสิทธิภาพระหว่างวิธีการแบบเดิมกับแบบใหม่ที่ได้รับการปรับเปลี่ยนแล้ว

## 1.4 ขอบเขตการวิจัย

1.4.1 ในการทำวิทยานิพนธ์นี้จะพัฒนาบนเครื่องไมโครคอมพิวเตอร์

1.4.2 ใช้ภาษาระดับสูงในการพัฒนาโปรแกรมการอัดข้อมูลและขยายข้อมูล

1.4.3 การทดสอบการอัดข้อมูล จะทดสอบกับข้อมูล 2 ลักษณะคือ ข้อมูลภาษาไทยล้วน และข้อมูลภาษาไทยทั่ว ๆ ไปที่อาจมีลักษณะอื่นรวมอยู่ด้วยเช่น ภาษาอังกฤษ การตีตาราง เป็นต้น โดยที่ข้อมูลแต่ละกลุ่มจะประกอบด้วยข้อมูลหลาย ๆ ขนาด

1.4.4 ทดสอบการขยายข้อมูล เพื่อตรวจสอบความถูกต้องของโปรแกรมการอัดข้อมูลและโปรแกรมการขยายข้อมูล

1.4.5 ใช้รหัสภาษาไทยของสำนักงานมาตรฐานผลิตภัณฑ์อุตสาหกรรม (ส.ม.อ.)

## 1.5 ขั้นตอนการดำเนินการวิจัย

1.5.1 ศึกษาและค้นคว้าวิธีการอัดข้อมูลวิธีต่าง ๆ ที่มีอยู่ในปัจจุบัน

1.5.2 ศึกษาและค้นคว้าลักษณะความเหลือเฟือ (Redundancy) ของภาษาไทย

1.5.3 จัดเตรียมข้อมูลภาษาไทยเพื่อใช้ในการทดสอบ

1.5.4 พัฒนาโปรแกรมการอัดข้อมูล (Compression Program) และโปรแกรมการขยายข้อมูล (Decompression Program)

1.5.5 ทดสอบโปรแกรม

1.5.6 สรุปผลการวิจัย และข้อเสนอแนะ

## 1.6 ประโยชน์ที่ได้รับจากการวิจัย

1.6.1 ได้รับความรู้เกี่ยวกับการอัดข้อมูลวิธีต่าง ๆ ที่มีอยู่ในปัจจุบัน

1.6.2 สามารถใช้เป็นแนวทางสำหรับการพัฒนาโปรแกรมสำเร็จรูป หรือเป็นพื้นฐานสำหรับโปรแกรมประยุกต์อื่น ๆ ที่ต้องการนำวิธีการอัดข้อมูลไปใช้งาน

1.6.3 ได้วิธีการอัดข้อมูลที่เหมาะสมที่มีการพัฒนาให้สอดคล้องกับข้อมูลภาษาไทย

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย