



บทที่ 1

บทนำ

1. ความเป็นมาของปัญหา

ในยุคแรกๆ คอมพิวเตอร์ ถูกใช้สำหรับการประมวลผลข้อมูล ที่เป็นตัวเลขเพียงอย่างเดียว แต่ต่อมาคอมพิวเตอร์ถูกพัฒนา ให้สามารถใช้กับการประมวลผลข้อมูลที่เป็นตัวอักษรได้ด้วย และในปัจจุบัน การประยุกต์ใช้คอมพิวเตอร์ ในการประมวลผลข้อมูลที่เป็นตัวอักษร มีมากขึ้นกว่าการประมวลผลข้อมูล ที่เป็นตัวเลข ทั้งนี้เพราะมนุษย์ ใช้ตัวอักษรในการสื่อสารกัน อย่างแพร่หลายมากกว่าตัวเลข และการประมวลผลข้อมูลที่เป็นตัวเลข มักจะมีการใช้อยู่ในวงจำกัด ด้วยเหตุนี้ สารสนเทศในลักษณะของตัวอักษร จึงถูกผลิตออกมามากมายหลายรูปแบบ เช่นหนังสือ บทความ เอกสาร วารสาร รายงานการวิจัย รายงานการประชุม และสิ่งตีพิมพ์ต่างๆ เป็นต้น การเก็บและการค้นคืนสารสนเทศโดยใช้คอมพิวเตอร์ จึงมีความสำคัญและได้รับความสนใจอย่างกว้างขวางมากขึ้น เนื่องจากการใช้คอมพิวเตอร์ ในการจัดเก็บสารสนเทศ ที่มีปริมาณมากมาย มีความสะดวก และการค้นคืนสารสนเทศ สามารถทำได้โดยง่าย และรวดเร็ว ทั้งยังสามารถดำเนินการได้ในหลายๆ รูปแบบ

ปัจจุบันวิธีการที่ใช้ในการเก็บ และการค้นคืนสารสนเทศ มี 2 ระบบใหญ่ๆ คือ [Salton, 1989]

1.1 ระบบการค้นคืนสารสนเทศแบบสัญญาณนิยม (Conventional Information Retrieval System)

เป็นระบบการค้นคืนสารสนเทศ ที่อาศัยแนวคิดของแฟ้มข้อมูลผกผัน (Inverted File) เป็นพื้นฐานในการค้นคืนสารสนเทศ

1.2 ระบบการค้นคืนสารสนเทศขั้นสูง (Advanced Information Retrieval System)

เป็นระบบการค้นคืนสารสนเทศ ที่อาศัยแนวคิดของเวกเตอร์สเปซโมเดล (Vector Space Model) เป็นพื้นฐานในการค้นคืนสารสนเทศ

ในงานวิจัยนี้ จะศึกษาการเก็บและการค้นคืนสารสนเทศ โดยอาศัยแนวคิดของแฟ้มข้อมูลผกผันเป็นพื้นฐาน

2. วัตถุประสงค์ของการวิจัย

สร้างโปรแกรมเพื่อใช้ในการเก็บและการค้นคืนสารสนเทศ โดยอาศัยแนวคิดของแฟ้มข้อมูลผกผัน กับข้อมูลที่เป็นภาษาไทย

3. ขอบเขตของการวิจัย

3.1 ข้อมูลที่ใช้ในการทดสอบ จะต้องเป็นเอกสารที่ประกอบด้วย ตอน ประโยค และคำ มารวมกัน

3.2 ระบบที่พัฒนาขึ้นสามารถใช้ได้กับเอกสารที่เป็นภาษาไทย

3.3 ในส่วนของเอกสารภาษาไทย ได้รับความอนุเคราะห์ให้ใช้ โปรแกรมสำหรับตัดคำ จากบริษัทซิมมิตคอมพิวเตอรื

3.4 ภาษาคอมพิวเตอร์ที่ใช้ ในการพัฒนาโปรแกรม จะใช้ภาษาซี และ เครื่องคอมพิวเตอร์ที่ใช้ เป็นเครื่องไมโครคอมพิวเตอร์

4. ขั้นตอนของการวิจัย

4.1 ศึกษาแนวคิดของแฟ้มข้อมูลผกผัน และโครงสร้างข้อมูลแบบทรี ที่ใช้ในการเก็บและการค้นคืนสารสนเทศ

4.2 ออกแบบโครงสร้างของข้อมูล และโครงสร้างของแฟ้มข้อมูล ที่ใช้ประยุกต์ตามแนวคิดนี้

4.3 ออกแบบโครงสร้างของโปรแกรม

4.4 เขียนโปรแกรม และทดสอบโปรแกรม

4.5 สรุป และ ประเมินผล

5. ประโยชน์ที่คาดว่าจะได้รับ

5.1 อำนวยความสะดวก ในการเก็บและการค้นคืนสารสนเทศที่เป็นภาษาไทย

5.2 เพื่อเป็นแนวทาง ในการพัฒนาโปรแกรม สำหรับการเก็บ และการค้นคืนสารสนเทศที่เป็นภาษาไทย โดยอาศัยแนวคิดของแฟ้มข้อมูลผกผัน ต่อไป