บทที่ 6

สรุปผลการวิจัยและข้อเสนอแนะ

1. สรุปผลการวิจัย

การวิจัยในครั้งนี้มีจุดประสงค์เพื่อ สร้างโปรแกรมที่ใช้ในการเก็บ และการค้นคืนสารสนเทศที่อยู่ในรูปของเอกสาร ซึ่งสามารถใช้ได้ทั้งเอกสารที่เป็นภาษาไทย และเอกสารที่เป็นภาษาอังกฤษ โดยเอกสารที่เป็นภาษาไทย จะต้องมีการแบ่งข้อความในเอกสารออกเป็นคำๆ ด้วยช่องว่าง

ในการประมวลผลจะเก็บคำที่เป็นดัชนีไว้ในแฟ้มข้อมูลดิคชันนารี และเก็บรายการหมายเลขอ้างอิงตำแหน่งเอกสาร ที่สัมพันธ์กับดัชนีไว้ในแฟ้มข้อมูลผกผัน ในการค้นคืนเอกสารจะระบุโดยดัชนีที่ต้องการค้นหา และเอกสารที่ถูกระบุโดยหมายเลขอ้างอิงที่สัมพันธ์กับดัชนีก็จะถูกอ่านมาจากแฟ้มข้อมูลเอกสาร

ในการออกแบบโครงสร้างของแฟ้มข้อมูลที่สำคัญ คือ แฟ้มข้อมูลดิคชันนารี ใช้โครงสร้างข้อมูลแบบบีทรี เนื่องจากโครงสร้างข้อมูลแบบบีทรี เป็นทรีที่มีความสูงสมดุล คือ โหนดใบทุกโหนดจะอยู่ในระดับเดียวกัน ทำให้สามารถค้นหาข้อมูลได้รวดเร็ว และจบลงด้วยจำนวนครั้งในการเข้าถึงแฟ้มข้อมูลที่เท่ากัน และยังสามารถใช้เนื้อที่ในการเก็บข้อมูลได้อย่างมีประสิทธิภาพ เนื่องจากแต่ละโหนดจะมีจำนวนคีย์อย่างน้อยที่สุดครึ่งหนึ่งของจำนวนคีย์ที่สามารถมีได้มากที่สุด และในการออกแบบแฟ้มข้อมูลผกผันใช้โครงสร้างข้อมูลแบบลิงค์ลิส ซึ่งเหมาะสำหรับการเก็บรายการหมายเลขอ้างอิงตำแหน่งของเอกสารของดัชนีแต่ละตัว ที่มีจำนวนแตกต่างกัน

จากการที่มีการสร้างแฟ้มข้อมูล ในหน่วยความจำสำรองโดยตรง ทำให้การทำงานในส่วนของการสร้างแฟ้มข้อมูล ทำได้ช้ากว่าการสร้างแฟ้มข้อมูล ในหน่วยความจำหลัก แต่จะสามารถเก็บข้อมูลได้ในปริมาณที่มากกว่า ซึ่งในการประยุกต์ใช้งาน ข้อมูลที่ต้องการเก็บมีปริมาณมาก ดังนั้นเมื่อเปรียบเทียบปริมาณข้อมูลที่ต้องการเก็บ กับความเร็วในการทำงานของโปรแกรม จึงนับได้ว่าระบบที่สร้างขึ้นมาสามารถทำงานได้ดีพอสมควร และในการค้นหาข้อมูล สามารถค้นหาได้รวดเร็ว โปรแกรมสามารถแสดงข้อความได้ทันที ที่ผู้ใช้ป้อนข้อความในการค้นหาข้อมูลเสร็จเรียบร้อย

ในการทดสอบโปรแกรม โดยการสร้างบีทรีที่มีลำดับ 2M+1 ต่างๆ กัน โดยที่ M คือเลขจำนวนเต็มบวก ที่แทนจำนวนคีย์ที่น้อยที่สุด ที่สามารถมีได้ในแต่ละโหนด พบว่า การทดสอบโดยใช้ค่า M ที่มีขนาดใหญ่ ทำให้ระดับของบีทรีมีจำนวนน้อยลง ซึ่งจะทำให้ลดเวลาในการเข้าถึงแฟ้มข้อมูลได้มาก แต่จะเสียเวลาในการเรียงลำดับข้อมูลในแต่ละโหนดมากขึ้น การทดสอบโดยใช้ค่า M ที่มีขนาดเล็ก ทำให้ไม่ต้องเสียเวลาในการเรียงลำดับข้อมูลในแต่ละโหนด แต่ทำให้ระดับของบีทรีมีจำนวนเพิ่มมากขึ้น ซึ่งจะทำให้เสียเวลาในการเข้าถึงแฟ้มข้อมูลมากขึ้นด้วย และจากการทดสอบโปรแกรม โดยใช้แฟ้มข้อมูลข้อความที่มีขนาด 20 กิโลไบต์ พบว่า ค่า M ที่เหมาะสมจะมีค่าอยู่ในช่วง 10-30

ภาษาคอมพิวเตอร์ที่ใช้ในการพัฒนาโปรแกรมใช้ภาษาซี บนเครื่องไมโครคอมพิวเตอร์ภายใต้ระบบปฏิบัติการ DOS รุ่น 3.2

2. ปัญหาและข้อเสนอแนะ

1) เนื่องจากในระบบภาษาไทย ไม่มีมาตรฐานของคำหยุด (stop words) ดังนั้นในการประมวลผล จึงเก็บคำทุกคำเป็นดัชนี ซึ่งจะทำให้เปลืองเนื้อที่ในการเก็บคำบางคำที่ไม่มีนัยสำคัญ เช่น จะ ที่ และ หรือ เป็นต้น

2) การแบ่งข้อความในเอกสารภาษาไทยออกเป็นคำๆ ขึ้นอยู่กับผู้ใช้แต่ละคน จึงทำให้คำที่ได้จากการแบ่ง ไม่มีมาตรฐานที่แน่นอน