

บทที่ 1

บทนำ

1.1 ที่มาของปัญหาที่นำไปสู่การค้นคว้าวิจัย และแนวทางการแก้ปัญหา

โดยทั่วไปการค้นหาข้อมูลในฐานข้อมูลเชิงสัมพันธ์ ส่วนใหญ่จะเป็นการค้นหาภายใต้เงื่อนไขของคำหลัก Like หรือเครื่องหมายเชิงการเปรียบเทียบ เช่น $= > <$ เป็นต้น ข้อเสียของการค้นหาข้อมูลลักษณะนี้ คือระบบค้นหาข้อมูลจะไม่สนใจลำดับความสำคัญของผลลัพธ์ที่ได้ ว่าตรงกับความต้องการของผู้ใช้หรือไม่ และหากต้องการเรียงลำดับของผลลัพธ์ที่ได้จากการค้นหา ลักษณะของการจัดเรียงลำดับได้แค่สองรูปแบบ คือ เรียงลำดับจากมากไปหาน้อยหรือจากน้อยไปมากจากคอลัมน์ที่เลือกใช้ในการจัดเรียงลำดับเท่านั้น ทำให้ยากต่อการค้นหาข้อมูลที่ต้องการต้องเสียเวลาค่อนข้างมากในการเลือกใช้ข้อมูลที่ต้องการจากผลลัพธ์ที่ได้จากการค้นหา

ปัจจุบันระบบจัดการฐานข้อมูลเชิงสัมพันธ์บางระบบ อาทิ เช่น Microsoft SQL Server, Oracle, MySQL จะมีความสามารถในการจัดเรียงลำดับความสำคัญของผลลัพธ์ที่ได้จากการค้นหาตามความต้องการของผู้ใช้ได้ โดยใช้ผ่านเทคโนโลยีการค้นหาแบบแยกคำ (Full Text Search)

แต่เทคโนโลยีการค้นหาแบบแยกคำนั้น ใ้ว่าจะมีอยู่ในทุกระบบฐานข้อมูลเชิงสัมพันธ์ ซึ่งส่วนใหญ่จะมีเฉพาะในระบบฐานข้อมูลที่ได้รับความนิยมหรือเวอร์ชันธุรกิจ(Commercial version) อีกทั้งเทคโนโลยีนี้ยังมีรูปแบบการทำงานที่ยังไม่มีความเสถียรภาพเท่าที่ควร ปัญหาหลักของการใช้เทคโนโลยีการค้นหาแบบแยกคำ ในปัจจุบันคือ จะใช้ได้กับข้อมูลที่เก็บในรูปแบบยูนิโคด (Unicode) เท่านั้น ต้องระบุตารางที่จะใช้งานในรูปแบบเฉพาะของเทคโนโลยีการค้นหาแบบแยกคำด้วย ทำให้เกิดความยุ่งยากในการใช้งาน

1.2 หลักการและเหตุผล

จากปัญหาที่กล่าวมาข้างต้น ทำให้ผู้วิจัยมีแนวคิดที่จะศึกษาเกี่ยวกับเทคโนโลยี การค้นคืนสารสนเทศ (Information Retrieval หรือ IR) เพื่อนำมาพัฒนาระบบการค้นหาข้อมูลผ่านระบบฐานข้อมูลเชิงสัมพันธ์ (Relational Database หรือ RDBMS) สำหรับใช้ในการค้นหาข้อมูลต่างๆ ที่ตรงตามความต้องการของผู้ใช้มากที่สุด โดยที่ไม่จำเป็นต้องระบุตารางที่สามารถใช้เทคโนโลยีการค้นหาแบบแยกคำ (Full Text Search) ในการค้นหาข้อมูล และสามารถค้นหาข้อมูลที่ไม่ได้เก็บอยู่ในรูปแบบยูนิโคดได้ด้วย

1.3 วัตถุประสงค์ของการศึกษา

เพื่อพัฒนาระบบค้นหาแบบคำหลักบนระบบจัดการฐานข้อมูลเชิงสัมพันธ์โดยใช้หลักการจัดการค้นคืนสารสนเทศ

1.4 ขอบเขตการศึกษา

- 1.4.1 ระบบรองรับภาษาไทยและภาษาอังกฤษเท่านั้น
- 1.4.2 ระบบตัดคำใช้เทคโนโลยี ICU (International Components for Uni-Code)
- 1.4.3 การเรียงลำดับผลลัพธ์จะใช้การคำนวณน้ำหนักค่าแบบ TF x IDF
- 1.4.4 ระบบรองรับระบบจัดการฐานข้อมูล Microsoft SQL Server เวอร์ชัน 2008 ขึ้นไป และภาษาที่ใช้ในการพัฒนาระบบคือ C#
- 1.4.5 ระบบที่พัฒนาขึ้นใช้แพลตฟอร์ม Microsoft .Net Framework เวอร์ชัน 2.0 รองรับระบบจัดการเว็บบน Internet Information Server Version 7.0 ขึ้นไป และรองรับการทำงานบนระบบปฏิบัติการ Windows Server Version 2008 ขึ้นไป
- 1.4.6 ระบบรองรับการทำงานผ่านโปรแกรม Internet explorer version 7.0 ขึ้นไป หรือเทียบเท่า เช่น Google Chrome, Fire fox เป็นต้น
- 1.4.7 ทดสอบโดยใช้ฐานข้อมูล E-Manage ในด้านงานวิจัยโดยใช้ตารางข้อมูลงานวิจัย
- 1.4.8 ทำ User Test ไม่น้อยกว่า 20 คน โดยเป็นบุคลากรหรือนักศึกษา มหาวิทยาลัยแม่โจ้ที่เข้าใช้งานระบบ E-Manage ด้านงานวิจัยเท่านั้น โดยวิธีการแจกแบบสำรวจ