

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

เทคนิคการค้นหาคำสำคัญเป็นที่นิยมอย่างแพร่หลายในการสืบค้นข้อมูลจากอินเทอร์เน็ต ผู้ใช้สามารถค้นหาข้อมูลจากเอกสารบนอินเทอร์เน็ตได้โดยระบุคำสำคัญหรือกลุ่มคำสำคัญที่เกี่ยวข้องกับข้อมูลที่ต้องการค้นหา ในขณะที่การค้นหาข้อมูลจากฐานข้อมูลผู้ใช้อาจค้นหาได้โดยใช้ภาษาสอบถามเชิงโครงสร้างหรือภาษาเอสคิวแอล (Structured Query Language หรือ SQL) ซึ่งผู้ใช้งานจำเป็นต้องมีความรู้เกี่ยวกับโครงสร้างฐานข้อมูลและภาษาที่ใช้ในการสอบถาม หรืออาจค้นหาได้จากแบบฟอร์มที่ผู้ใช้บริการระบบฐานข้อมูลนั้นจัดเตรียมไว้ (Form-Based Interface) [4] การนำเทคนิคการค้นหาคำสำคัญมาใช้กับฐานข้อมูลจึงน่าจะช่วยให้ผู้ใช้ค้นหาข้อมูลได้สะดวกขึ้น แต่เนื่องจากข้อมูลบนอินเทอร์เน็ตและข้อมูลบนฐานข้อมูลเก็บอยู่ในรูปแบบที่ต่างกัน กล่าวคือ บนอินเทอร์เน็ตข้อมูลอยู่ในรูปของเอกสาร ในขณะที่ฐานข้อมูลเก็บข้อมูลแบบมีโครงสร้าง คือ อยู่ในรูปของหลายๆ เอนทิตี (Entity) และมีคีย์เชื่อมโยงความสัมพันธ์ระหว่างข้อมูล (Primary Key-Foreign Key) จึงไม่สามารถนำเทคนิคการค้นหาคำสำคัญบนอินเทอร์เน็ตมาใช้ในการค้นหาข้อมูลจากฐานข้อมูลได้โดยตรง

การค้นหาคำสำคัญในฐานข้อมูล เป็นวิธีหนึ่งที่ใช้ค้นหาข้อมูลจากฐานข้อมูลได้ โดยที่ผู้ใช้ไม่จำเป็นต้องมีความรู้ด้านโครงสร้างฐานข้อมูลและภาษาเอสคิวแอล ผู้ใช้สามารถทำการค้นหาได้โดยระบุคำสำคัญที่เกี่ยวข้องกับข้อมูลที่ต้องการค้นหา คำสำคัญดังกล่าวอาจปรากฏในค่าข้อมูลหรือข้อมูลฐานข้อมูล (ชื่อเอนทิตี หรือชื่อแอทริบิว) ผลลัพธ์ที่ได้จากการค้นหา คือ ระเบียบข้อมูลที่ประกอบด้วยคำสำคัญ และ/หรือชื่อเอนทิตีหรือชื่อแอทริบิวตรงกับคำสำคัญตามที่ผู้ใช้ต้องการ โดยระเบียบเหล่านั้นมีความสัมพันธ์กันทางโครงสร้างฐานข้อมูล

ปัจจุบันมีงานวิจัยที่เกี่ยวกับการค้นหาคำสำคัญในฐานข้อมูล โดยมีการสร้างแบบจำลองที่ใช้ในการจัดเก็บโครงสร้างฐานข้อมูลรวมทั้งค่าข้อมูลต่างๆ ที่เปรียบได้กับเอกสารบนอินเทอร์เน็ต และนำดัชนีผกผัน (Inverted Index) มาใช้ในการระบุตำแหน่งของคำสำคัญในแบบจำลอง แล้วนำตำแหน่งที่ได้มาค้นหาผลลัพธ์ตามอัลกอริทึมที่แตกต่างกันออกไป ซึ่งสามารถแบ่งระบบการค้นหาคำสำคัญในฐานข้อมูลตามประเภทของคำดัชนีได้ 2 ประเภท คือ ระบบที่ใช้คำดัชนีตรงกับค่าข้อมูล [3][5][7][11] และระบบที่ใช้คำดัชนีตรงกับค่าข้อมูลและข้อมูลฐานข้อมูล [1][10][13] ซึ่งระบบที่ใช้

คำดัชนีตรงกับค่าข้อมูลไม่สามารถค้นหาคำสำคัญที่ตรงกับข้อมูลฐานข้อมูลได้ ในขณะที่ระบบที่ใช้คำดัชนีตรงกับค่าข้อมูลและข้อมูลฐานข้อมูลบางระบบไม่สามารถค้นหาด้วยชื่อเอนทิตีได้ [13] และบางระบบไม่สามารถค้นหาคำสำคัญที่มาจากเอนทิตีมากกว่า 2 เอนทิตีได้ [10] นอกจากนี้ ระบบส่วนใหญ่ไม่อนุญาตให้ผู้ใช้ค้นหาคำสำคัญด้วยคำพ้องความหมายของค่าข้อมูล หรือข้อมูลฐานข้อมูลได้

งานวิจัยนี้จึงได้เสนอวิธีการค้นหาคำสำคัญในฐานข้อมูล ซึ่งผู้ใช้สามารถใช้คำสำคัญหรือกลุ่มคำสำคัญที่ตรงกับค่าข้อมูล ข้อมูลฐานข้อมูล (ชื่อเอนทิตี และชื่อแอทริบิว) และคำพ้องความหมายในการค้นหา คำสำคัญเหล่านี้จะถูกแปลงให้อยู่ในรูปของภาษาเอสคิวแอลโดยใช้แบบจำลองข้อมูลที่ประกอบด้วยข้อมูล 3 ประเภท คือ โครงสร้างฐานข้อมูล คำพ้องความหมาย และเคส เรียกว่า การค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นหาข้อมูลฐานข้อมูล

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

งานวิจัยนี้เป็นการพัฒนาวิธีการค้นหาคำสำคัญในฐานข้อมูล โดยใช้การค้นหาข้อมูลฐานข้อมูลซึ่งประกอบด้วยข้อมูล 3 ประเภท คือ โครงสร้างฐานข้อมูล คำพ้องความหมาย และเคส ทำให้ผู้ใช้สามารถค้นหาข้อมูลจากฐานข้อมูลได้โดยใช้คำสำคัญที่ตรงกับค่าข้อมูล ข้อมูลฐานข้อมูล (ชื่อเอนทิตี และชื่อแอทริบิว) และคำพ้องความหมาย

1.3 สมมติฐานของการศึกษา

การค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นหาข้อมูลฐานข้อมูลให้ผลลัพธ์ที่ถูกต้องแม่นยำกว่าการค้นหาคำสำคัญที่ตรงกับค่าข้อมูลเพียงอย่างเดียว

1.4 ขอบเขตของการศึกษา

ออกแบบแบบจำลองข้อมูลและดัชนีผกผัน ซึ่งแบบจำลองข้อมูลใช้ในการเก็บโครงสร้างฐานข้อมูล คำพ้องความหมาย และเคส รวมทั้งพัฒนาวิธีการค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นหาข้อมูลฐานข้อมูล

1.5 ส่วนประกอบของวิทยานิพนธ์

ส่วนที่เหลือของวิทยานิพนธ์ฉบับนี้แบ่งเนื้อหาออกเป็น 4 บท ดังต่อไปนี้

บทที่ 2 กล่าวถึงทฤษฎีและงานวิจัยที่เกี่ยวข้อง ซึ่งแบ่งเนื้อหาออกเป็น 3 ส่วน คือ ส่วนที่ 1 กล่าวถึงข้อมูลพื้นฐานของการค้นหาคำสำคัญในฐานข้อมูล ส่วนที่ 2 กล่าวถึงวิธีการค้นหาคำสำคัญจากฐานข้อมูลในวิธีต่างๆ ซึ่งแบ่งเป็นระบบที่ใช้คำสำคัญที่ตรงกับคำข้อมูล และระบบที่ใช้คำสำคัญที่ตรงกับคำข้อมูลและข้อมูลฐานข้อมูล และสุดท้าย ส่วนที่ 3 เป็นการสรุปผลการวิเคราะห์วิธีการค้นหาคำสำคัญในฐานข้อมูล

บทที่ 3 กล่าวถึงการค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นหาข้อมูลฐานข้อมูล ซึ่งแบ่งเนื้อหาเป็น 4 ส่วนหลักๆ คือ ส่วนแรกกล่าวถึงแบบจำลองข้อมูล ส่วนที่ 2 กล่าวถึงสถาปัตยกรรมของระบบ ส่วนที่ 3 กล่าวถึงอัลกอริทึมที่ใช้ในการค้นหาผลลัพธ์ และส่วนที่ 4 กล่าวถึง การจัดการเอกสาร

บทที่ 4 เป็นการประเมินผล ซึ่งแบ่งเนื้อหาเป็น 5 ส่วน คือ ส่วนแรก กล่าวถึงวิธีการประเมินผลของระบบ ส่วนที่ 2 กล่าวถึงตัวอย่างฐานข้อมูลที่ใช้ในการประเมินผล ส่วนที่ 3 กล่าวถึงตัวอย่างการค้นหาคำสำคัญในฐานข้อมูลโดยใช้การค้นหาข้อมูลฐานข้อมูล ส่วนที่ 4 กล่าวถึงตัวอย่างการพิจารณาเคสที่เหมาะสมกับการสอบถามเพื่อนำมารียูส และส่วนที่ 5 กล่าวถึงการสรุปผลการประเมินผล

บทที่ 5 เป็นบทสรุปผลการวิจัยและข้อเสนอแนะ