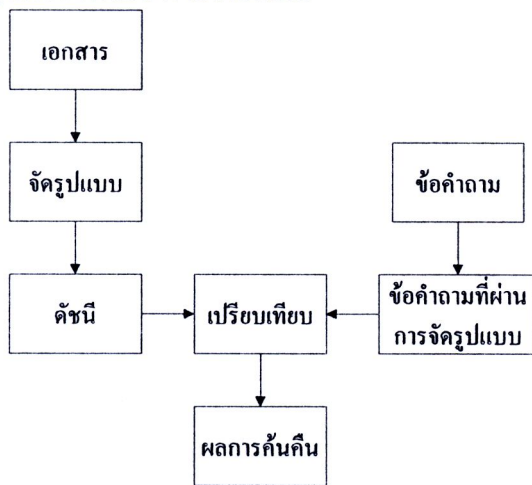


บทที่ 1

บทนำ

1.1 ความสำคัญและที่มาของปัญหา

ระบบค้นคืนสารสนเทศจะเกี่ยวข้องกับการค้นหาข้อมูลที่ตรงตามความต้องการของผู้ใช้ โดยผู้ใช้จะกำหนดข้อความ (Query) ให้แก่ระบบ และระบบจะนำข้อความนั้นมาเปรียบเทียบกับสารสนเทศที่มี แล้วแสดงผลการค้นคืนที่เกี่ยวข้องให้กับผู้ใช้ โดยทั่วไปจะทำการเปรียบเทียบข้อความกับข้อมูลทั้งหน้าเว็บเพจซึ่งผลการค้นคืนที่ได้อาจจะไม่ตรงตามความต้องการของผู้ใช้ เพราะอาจจะพบคำที่ตรงกับข้อความในส่วนอื่นที่ไม่ใช่ส่วนสำคัญ เช่น โฆษณา เมนูหลัก ลิงก์ที่ไม่เกี่ยวข้อง เป็นต้น ซึ่งทำให้ผลการค้นคืนที่ได้ไม่สอดคล้องกับความต้องการ ดังนั้นเพื่อให้ได้ผลการค้นคืนตรงตามความต้องการของผู้ใช้จึงนำอัลกอริทึมในการแบ่งหน้าเว็บเพจ มาทำการศึกษาและทำการจำลองระบบค้นคืนสารสนเทศ เพื่อทดสอบผลที่ได้จากการนำอัลกอริทึมดังกล่าวมาใช้ ซึ่งจะเป็นพื้นฐานในการพัฒนางานวิจัยต่อไป



รูปที่ 1.1 ระบบค้นคืนสารสนเทศ

จากรูปที่ 1.1 เริ่มจากการนำเอกสารที่ต้องการค้นคืนมาจัดรูปแบบให้เหมาะสมเพื่อนำไปสร้างเป็นดัชนีเก็บไว้ เมื่อต้องการค้นคืนเอกสารก็จะใส่ข้อความเข้าสู่ระบบ ระบบก็จะนำข้อความมาทำการจัดรูปแบบให้เหมาะสม แล้วนำมาเปรียบเทียบกับดัชนีที่มีอยู่ ทำให้ได้ผลการค้นคืนเป็นเอกสารที่เกี่ยวข้องกับข้อความจากผู้ใช้งาน

ในโครงงานวิจัยนี้จะมีการนำอัลกอริทึมวีโอพีเอส (VIPS : Vision Based Pages Segmentation) มาแบ่งเว็บเพจออกเป็นส่วนย่อย โดยทำการแบ่งหน้าเว็บเพจออกเป็นบล็อก ซึ่งเมื่อทำการแบ่งหน้าเว็บเพจออกเป็นบล็อกแล้ว จะทำการเปรียบเทียบข้อความที่ได้กับหน้าเว็บเพจที่แบ่งเสร็จแล้ว หลังจากนั้นจะทำการเลือกคำใหม่เพื่อนำมาเพิ่มเข้าไปในข้อความเดิม แล้วจึงจะทำการค้นคืนย้อนกลับ ซึ่งจะศึกษาการทำงานของระบบค้นคืนสารสนเทศบนเว็บโดยใช้อัลกอริทึมวีโอพีเอส นำมารวมกับการค้นคืนย้อนกลับ เพื่อเป็นแนวทางในการพัฒนาการค้นคืนสารสนเทศบนเว็บให้ใช้งานได้จริง

1.2 วัตถุประสงค์

1. เพื่อศึกษาและจำลองการค้นคืนสารสนเทศ โดยใช้อัลกอริทึมวีไอพีเอสและการค้นคืนย้อนกลับ
2. เพื่อศึกษาการทำงานของระบบค้นคืนสารสนเทศ
3. วัดประสิทธิภาพการค้นคืนสารสนเทศที่ใช้อัลกอริทึมวีไอพีเอสและการค้นคืนย้อนกลับ

1.3 ขอบเขตของโครงการวิจัย

1. ศึกษากระบวนการค้นคืนสารสนเทศโดยใช้อัลกอริทึมวีไอพีเอสและการค้นคืนย้อนกลับ
2. พัฒนาและออกแบบระบบจำลองที่มีการรับข้อความเป็นคำสำคัญ (Keyword) เท่านั้น
3. พัฒนาและออกแบบระบบจำลองระบบค้นคืนสารสนเทศที่รองรับเฉพาะภาษาอังกฤษ
4. ทำการวัดประสิทธิภาพการค้นคืนสารสนเทศบนเว็บ ที่ใช้อัลกอริทึมวีไอพีเอสและการค้นคืนย้อนกลับ โดยเปรียบเทียบค่า R-Precision, Recall และ E-Measure

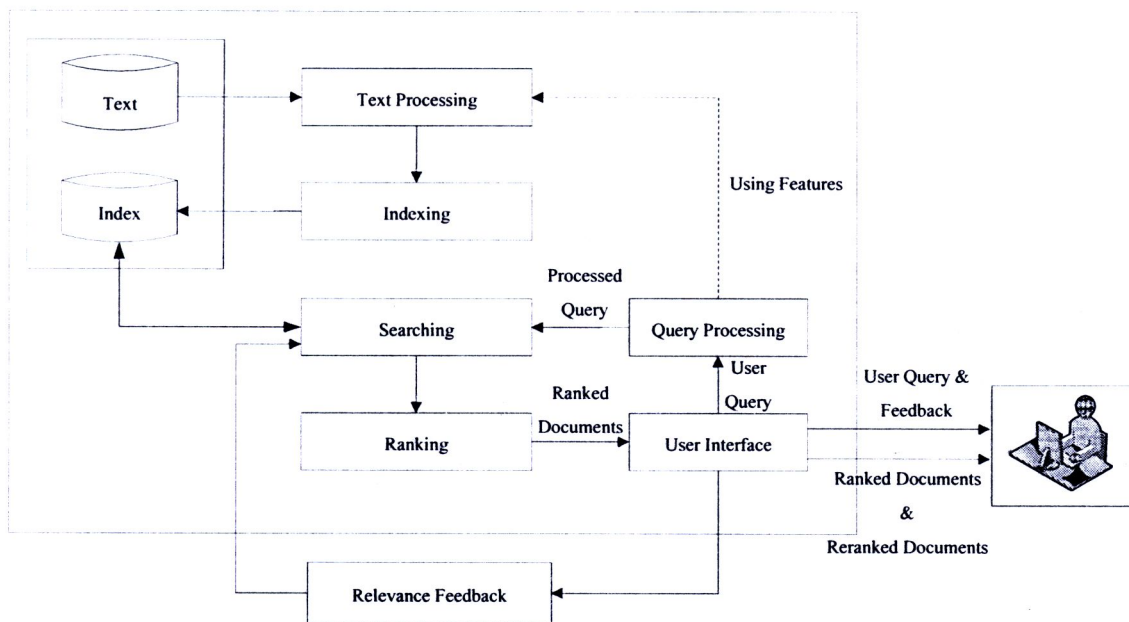
1.4 สมมติฐานของการศึกษา

ในระบบค้นคืนสารสนเทศ ผู้ใช้มักจะมีปัญหาในการตั้งคำถาม คือไม่สามารถตั้งคำถามให้ระบบค้นคืนสารสนเทศค้นคืนเอกสารที่ตรงตามความต้องการของผู้ใช้ได้ บ่อยครั้งที่ผู้ใช้ต้องเสียเวลามากมายในการเลือกชุดเอกสารที่เป็นผลลัพธ์ที่ได้มา ซึ่งมีทั้งเกี่ยวข้องและไม่เกี่ยวข้องกับผู้ต้องการ แนวทางหนึ่งในการแก้ปัญหา เพื่อสร้างความพึงพอใจให้แก่ผู้ใช้ คือการค้นคืนย้อนกลับจากผู้ใช้

การค้นคืนย้อนกลับจากผู้ใช้ หมายถึง การป้อนความเกี่ยวพันย้อนกลับ อันเป็นการใช้ประโยชน์จากข้อมูลย้อนกลับนี้ไปปรับเปลี่ยนข้อความเก่าให้เป็นข้อความใหม่ของการสืบค้นในรอบต่อไป โดยวิธีการนี้สามารถทำได้ทั้งระบบอัตโนมัติและระบบกึ่งอัตโนมัติโดยมีคนร่วมปฏิบัติการด้วย

ระบบค้นคืนสารสนเทศที่มีการค้นคืนย้อนกลับจากผู้ใช้ จะมีการประเมินของตัวอย่างเอกสารที่ถูกค้นคืนออกมา และการประเมินนี้จะถูกนำไปใช้เพื่อปรับปรุงกระบวนการค้นคืน โดยผู้ใช้จะมีการพิจารณาว่าข้อมูลที่ได้รับจากการค้นคืนมานั้น ตรงตามความต้องการของผู้ใช้มากน้อยแค่ไหนหลังจากนั้นจะนำผลลัพธ์ที่ผู้ใช้เลือกมาทำการปรับปรุงกระบวนการค้นคืนแล้วทำการค้นหาใหม่อีกครั้ง โดยการกระทำดังกล่าวจะทำให้จำนวนครั้งที่ได้ผลลัพธ์เป็นที่น่าพอใจ ระบบค้นคืนสารสนเทศที่มีการค้นคืนย้อนกลับจากผู้ใช้แสดงได้ดังรูปที่ 1.2

วีไอพีเอสเป็นอัลกอริทึมที่ใช้ในการแบ่งโครงสร้างทางภาษาของเว็บเพจ ซึ่งในโครงสร้างทางภาษานั้น มีลักษณะเป็นลำดับชั้นและมีโหนดตามลำดับชั้น โดยแต่ละโหนดนั้นจะถูกเรียกว่าบล็อก และจะถูกกำหนดค่าดีไอซี (DoC : Degree of Coherence) เพื่อแสดงว่าแต่ละส่วนในบล็อกนั้นจะถูกรวมกันได้อย่างไร ซึ่งจะมีการนำอัลกอริทึมดังกล่าวมาใช้ในการวิจัย เพื่อวิเคราะห์แนวทางในการที่จะดำเนินการวิจัยต่อไป



รูปที่ 1.2 ระบบค้นคืนสารสนเทศที่มีการค้นคืนย้อนกลับจากผู้ (Relevance Feedback)

1.5 วิธีการดำเนินการวิจัย

ในขั้นตอนการพัฒนางานวิจัยนั้น ผู้วิจัยมีขั้นตอนในการทำวิจัยตามหลักการทางด้านวิศวกรรมซอฟต์แวร์ (Software Engineering) โดยใช้แบบจำลองโมเดลน้ำตก (Waterfall Model) ดังนี้

1. ศึกษาและพัฒนาระบบค้นคืนสารสนเทศ และระบบค้นคืนสารสนเทศย้อนกลับ
2. ศึกษาการทำงานของอัลกอริทึมวีไอพีเอส
3. วิเคราะห์และออกแบบระบบจำลองการค้นคืนสารสนเทศแบบย้อนกลับ
4. วิเคราะห์วิธีการเปรียบเทียบระบบจำลองที่ได้ทำการออกแบบกับระบบค้นคืนสารสนเทศแบบอื่น ๆ
5. พัฒนาระบบตามหลักการที่ได้ออกแบบ
6. ทดสอบการทำงานของระบบจำลองในแต่ละโมดูล จากนั้นทำการทดสอบระบบโดยรวม
7. ทำการเปรียบเทียบระบบจำลองกับระบบค้นคืนสารสนเทศแบบอื่น ๆ
8. สรุปโครงการและทำรายงาน

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1. สามารถพัฒนาระบบจำลองการค้นคืนสารสนเทศที่ใช้อัลกอริทึมวีไอพีเอสและการค้นคืนย้อนกลับ
2. เปรียบเทียบผลการวัดประสิทธิภาพการค้นคืนสารสนเทศบนเว็บที่ใช้อัลกอริทึมวีไอพีเอสและการค้นคืนย้อนกลับ
3. สามารถนำไปต่อยอดในการพัฒนาพื้นฐานของงานวิจัยทางด้านระบบค้นคืนสารสนเทศได้