

การพัฒนาด้านเทคโนโลยีสารสนเทศนับวันจะเจริญรุดหน้าไปอย่างรวดเร็ว ทำให้โลกเราก้าวเข้าสู่ยุคแห่งข้อมูลสารสนเทศอย่างหลีกเลี่ยงไม่ได้ แต่ละวันข้อมูลข่าวสารถูกผลิตและเผยแพร่ผ่านเครือข่ายอินเทอร์เน็ตที่เชื่อมโยงกันทั่วโลกเป็นจำนวนมหาศาล และในอนาคตก็จะเพิ่มปริมาณมากขึ้น โดยเฉพาะเอกสารบนเว็บ (web document) ที่ให้บริการผ่านเวิร์ลไวด์เว็บ (World Wide Web) ผู้บริโภคข้อมูลข่าวสารต้องใช้เวลามากขึ้นในการคัดเลือกข้อมูลให้ได้ตามความต้องการ ซึ่งก่อให้เกิดปัญหาด้านการสืบค้นข้อมูลที่ใช้ได้ ข้อมูลค้นคืนไม่ตรงตามความต้องการ ประกอบกับที่ผ่านมากการสืบค้นข้อมูลใช้หลักการพื้นฐานของการเปรียบเทียบคำหลัก (keyword – based) เป็นส่วนใหญ่ ซึ่งไม่ได้ให้ความสำคัญของการพิจารณาความหมายในเนื้อหาของเอกสารเท่าที่ควร ปัจจุบันมีการค้นคว้าวิจัยอย่างกว้างขวางเพื่อพัฒนากระบวนการสืบค้นให้มีประสิทธิภาพทั้งในด้านการพัฒนาด้านฐานข้อมูลดัชนีและด้านการค้นคืนสารสนเทศ (information retrieval) งานวิจัยนี้นำเสนอแนวทางหนึ่งสำหรับการพัฒนาฐานข้อมูลดัชนีที่สามารถสืบค้นสารสนเทศบนเว็บเชิงความหมายได้ โดยประยุกต์ใช้ความรู้ด้านการประมวลผลภาษาธรรมชาติ (natural language processing) เพื่อวิเคราะห์การสอดคล้องกันเชิงความหมายของคำศัพท์ (lexical cohesion) ในเนื้อหาของเอกสารแล้วทำการจัดกลุ่มคำหลักเชิงแนวคิด ซึ่งได้เป็นกลุ่มของคำศัพท์ที่มีความหมายในแนวคิดเดียวกันเรียกว่า Lexical chain และในงานวิจัยนี้ยังใช้ความรู้ด้านการวัดความเหมือนระหว่างออนโทโลยี (ontologies similarity measuring) เพื่อวัดความเหมือนระหว่าง Lexical chain สำหรับจัดกลุ่มเอกสารตามความหมาย และนอกจากนั้นในการจัดอันดับเอกสารแต่ละกลุ่มได้วิเคราะห์ถึงการเชื่อมโยงกันของเอกสาร (hyperlink) ด้วยเพื่อให้สามารถจัดกลุ่มเอกสารได้ตามความหมายมากขึ้น ผลที่ได้จากการวิจัยนี้ คือฐานความรู้ที่เหมาะสมสำหรับการสืบค้นข้อมูลสารสนเทศเชิงความหมายอย่างมีประสิทธิภาพ ผลการสืบค้นข้อมูลเชิงความหมายในงานวิจัยนี้มีประสิทธิภาพดีกว่าการสืบค้นด้วยวิธีพิจารณาความถี่ของคำศัพท์เป็นหลัก (Term Frequency/Inverse Document Frequency : TF/IDF) ประมาณ 55.16 % และมีประสิทธิภาพดีกว่าวิธีการสืบค้นอย่างมีความหมาย (Semantic – based : SB) ประมาณ 20.22 %

## ABSTRACT

217109

According to the information technology which is continually developed, we cannot avoid being involved in the world of information technology we are living in. For that reason, an increasing number of information sources produced day by day are thoroughly linked and related all over the cyber world. And of course, they would go up increasingly. Particularly web document found on WWW (World Wide Web) , clients or users have to spend much more time finding the exact and correct information. The search engine used to find information based on keyword has been non-relevant enough. Actually, there have been many researches on how to develop information searching system both for knowledge – based index development and information retrieval. This research focuses on applying natural language processing for lexical cohesion in the document content and grouping in accordance with lexical chain. In addition, its aim is to study how to measure ontologies similarity in clustering documents and how the hyperlink document work. The result gained from the research is sufficient knowledge for effective semantic search. The effectiveness of web document for semantic search comparing with the search based on Term Frequency/Inverse Document Frequency (TF/IDF) is approximately 55.16 % and about 20.22 % comparing with the semantic-based method (SB).