งานวิจัยนี้นำเสนอ และเปรียบเทียบอัลกอริธึมแบบต่างๆ ในการสร้าง Inverted Index ในสภาพแวดล้อมแบบกระจาย (Distributed Inverted Index) สำหรับการค้นคืนสารสนเทศ (Information Retrieval) พัฒนาโดยใช้ภาษา Erlang ซึ่งเป็นภาษาคอมพิวเตอร์ที่เหมาะสมกับการพัฒนาโปรแกรมในสภาพแวดล้อมแบบกระจาย (Distributed Environment) โดยจุดมุ่งหมายหลักของงานวิจัยนี้ คือ การนำเสนอ และเปรียบเทียบวิธีการสร้าง Inverted Index แบบสมบูรณ์ (Full List Inverted Index) กับ Inverted Index แบบแยกส่วน (Partial List Inverted Index) รวมถึงการเปรียบเทียบวิธีการแบ่งกลุ่มดรรชนีออกเป็นสองรูปแบบคือ การแบ่งกลุ่มดรรชนีตามกลุ่มเอกสาร (Document Partitioning) และการแบ่งกลุ่มตามคำสำคัญ (Term Partitioning) ในการเปรียบเทียบนี้ ใช้วิธีการวัดประสิทธิภาพในด้านความเร็วในการสร้างดรรชนี โดยใช้ข้อมูลเอกสารจาก TREC โดยการเปรียบเทียบความเร็วในการสร้างดรรชนีสำหรับกลุ่มข้อมูลขนาดต่างๆ กัน ในช่วงตั้งแต่ 50 MB ถึง 800 MB และใช้จำนวนเครื่องคอมพิวเตอร์ต่างๆ กันระหว่าง 4 เครื่อง ถึง 16 เครื่อง นอกจากนี้แล้ว ยังเปรียบเทียบความเร็วในการค้นคืนข้อมูลจากการใช้ดรรชนีแบบต่างๆ กัน เพื่อเปรียบเทียบข้อดี และข้อเสียของการใช้ดรรชนีแต่ละแบบ ผลการทดลองแสดงให้เห็นว่า ดรรชนีชนิด Inverted Index แบบแยกส่วน ให้ผลลัพธ์ที่ดีกว่า Inverted Index แบบสมบูรณ์ ทั้งในด้านระยะเวลาที่ใช้ในการสร้างดรรชนี และระยะเวลาที่ใช้ในการค้นคืนข้อมูล ส่วนการแบ่งกลุ่มดรรชนีตามกลุ่มเอกสาร และการแบ่งกลุ่มดรรชนีตามคำสำคัญนั้น ให้ผลในด้านระยะเวลาในการสร้างดรรชนีที่ใกล้เคียงกัน แต่ดรรชนีที่แบ่งตามคำสำคัญนั้นมีความซับซ้อนในการพัฒนาสูงกว่า ดรรชนีที่แบ่งตามกลุ่มเอกสาร

This research work proposes and compares different algorithms for construction of Inverted Index in distributed environment that is used in Information Retrieval. This work is implemented in Erlang programming language, which is suitable for software development to be used in a distributed environment. The main purpose of this work is to propose and compare different techniques for constructing distributed Inverted Index; namely Full-list Inverted Index and Partial-list Inverted Index. This work also proposes and compares two types of Partial-list Inverted Index; namely, Document-partitioning Inverted Index and Term-partitioning Inverted Index. Performance comparison is measured in terms of index construction time. A data collection from TREC is used in the experiments. There are various sizes of data ranging from 50 MB to 800 MB, and there are many computers used in the experiment ranging from 4 Nodes to 16 Nodes. In addition to comparing index construction time, query processing times are also compared to demonstrate the advantages and disadvantages of different types of Inverted Indexes. The experimental results show that Partial-list Inverted Indexes outperform Full-list Inverted Indexes in both index construction time and query processing time whereas Document-partitioning and Term-partitioning approach are as competitive. However, the Term-partitioning approach is more complex to implement than the Document-partitioning approach.