การแสดงผลภาพบิตแม็บสำหรับข้อมูลเอกสารดิจิทัล

นายสัณห์ชัย นักรบ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2549

VISUALIZATION BITMAPS FOR DIGITAL DOCUMENT COLLECTION

Mr. Sunchai Nakrob

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science Program in Computer Science

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2006

หัวข้อวิทยานิพนธ์	การแสดงผลภาพบิตแม็บสำหรับข้อมูลเอกสารดิจิทัล
โดย	นายสัณห์ชัย นักรบ
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษา	อาจารย์ ดร.โชติรัตน์ รัตนามหัทธนะ

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.ดิเรก ลาวัณย์ศิริ)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(รองศาสตราจารย์ ดร.พรศิริ หมื่นไชยศรี)

..... อาจารย์ที่ปรึกษา
(อาจารย์ ดร.โชติรัตน์ รัตนามหัทธนะ)

..... กรรมการ
(อาจารย์ ดร.อติวงศ์ สุชาโต)

..... กรรมการ
(อาจารย์ ธงชัย โรจน์กังสดาล)

สัณห์ชัย นักรบ : การแสดงผลภาพบิตแม็บสำหรับข้อมูลเอกสารดิจิทัล. (VISUALIZATION BITMAPS FOR DIGITAL DOCUMENT COLLECTION) อาจารย์ ที่ปรึกษา : อ.ดร.โชติรัตน์ รัตนามหัทธนะ, 62 หน้า.

วิทยานิพนธ์นี้มีวัตถุประสงค์ในการแสดงผลภาพสำหรับข้อมูลเอกสารดิจิทัล โดยทำการแปลงข้อมูลในเอกสารจากตัวอักษรให้เป็นรูปภาพ เพื่อช่วยในการพิจารณาเปรียบเทียบความเหมือนและความแตกต่างของประเภทหรือหมวดหมู่เอกสาร ทำให้ผู้ใช้สามารถ จัดการ และจำแนกรูปแบบหรือประเภทของเอกสารได้ง่ายและรวดเร็วมากยิ่งขึ้น โดยไม่จำเป็นต้องเข้าไปพิจารณาเนื้อความในเอกสาร โดยการแสดงผลภาพมีแนวทางในการพัฒนาจากแนวคิดของทฤษฎีเคออสเกม ประยุกต์ร่วมกับการแสดงผลภาพบิตแม็บของข้อมูลอนุกรมเวลาโดยใช้วิธีการแบบแซ็ค

งานวิจัยนี้ได้ทำการวิเคราะห์รูปแบบและลักษณะต่างๆของเอกสาร โดยการปรับข้อมูลในเอกสารและกำหนดพารามิเตอร์ที่สำคัญต่างๆ เพื่อให้การแสดงผลภาพบิตแม็บจากข้อมูลในเอกสารมีความชัดเจนและมีประสิทธิภาพ ซึ่งได้มาจากการทดลองด้วยข้อมูลจริง นอกจากนี้ยังได้ทำการทดสอบประสิทธิภาพของการแสดงผลภาพจากการพิจารณาเปรียบเทียบภาพบิตแม็บของข้อมูลเอกสาร ทั้งจากการสังเกตและใช้วิธีการจัดกลุ่มภาพบิตแม็บโดยอัตโนมัติ ซึ่งได้ผลสรุปจากการทดสอบว่า การแสดงผลภาพสำหรับข้อมูลเอกสารดิจิทัล สามารถช่วยในการพิจารณาเปรียบเทียบความเหมือนและความแตกต่างของประเภทหรือหมวดหมู่เอกสารดิจิทัลได้อย่างมีประสิทธิภาพ

ภาควิชา.....วิศวกรรมคอมพิวเตอร์..... ลายมือชื่อนิสิต.....สัณห์ชัย นักรบ.....

สาขาวิชา.....วิทยาศาสตร์คอมพิวเตอร์..... ลายมือชื่ออาจารย์ที่ปรึกษา.....

ปีการศึกษา.....2549.....

4871543121 : MAJOR COMPUTER SCIENCE
KEY WORD: BITMAP VISUALIZATION / DOCUMENT BITMAPS / VISUALIZATION

SUNCHAI NAKROB : VISUALIZATION BITMAPS FOR DIGITAL DOCUMENT COLLECTION. THESIS ADVISOR : CHOTIRAT RATANAMAHATANA, Ph.D., 62 pp.

The objective of this research is to visualize digital documents by converting text data in the digital documents to a bitmap image to help compare the similarities and differences of document types or categories so that the document can be easily and more conveniently clustered and managed. Users do not need to read details in the document. This visualization technique combines together the advance in Chaos Game Theory and SAX representation in Time Series bitmap visualization.

By experimenting with real data, this research analyzes the feature and format of digital documents and later adjusts document data and defines important parameters so that bitmap visualization of the document data is well-defined and effective. Moreover, this research also tests the visualization efficiency by comparing the bitmaps of the digital document through both users' observation and automatic clustering. The result shows that the bitmap visualization technique for digital document data can effectively help differentiate the documents types or categories.

Department.....Computer Engineering..... Student's signature.....
Field of study.....Computer Science..... Advisor's signature.....
Academic year .....2006.....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้ด้วยความอนุเคราะห์ และความช่วยเหลืออย่างยิ่งจาก อ.ดร.โชติรัตน์ รัตนามหัทธนะ อาจารย์ที่ปรึกษา ซึ่งให้ข้อคิด แนวทาง และคำปรึกษาตลอดจนเป็น ผู้ตรวจทานแก้ไข ทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วง ขอขอบพระคุณเป็นอย่างสูง

ขอขอบพระคุณ รศ.ดร.พรศิริ หมื่นไชยศรี อ.ดร.อติวงศ์ สุชาโต และ อ.ธงชัย โรจน์กังสดาล ประธานกรรมการและกรรมการสอบวิทยานิพนธ์ ที่กรุณาให้คำแนะนำในการแก้ไขวิทยานิพนธ์ให้มีคุณภาพยิ่งขึ้น ขอขอบพระคุณคณาจารย์ในภาควิชาวิศวกรรมคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัยทุกท่านที่ประสิทธิ์ประสาทความรู้อันมีค่ายิ่งแก่ผู้วิจัย

ที่สำคัญที่สุดขอบพระคุณ คุณพ่อ คุณแม่ และเพื่อนๆ ที่เป็นแรงผลักดัน เป็นกำลังใจ ที่สำคัญให้ตลอดการศึกษาครั้งนี้

สารบัญ

หน้า

บทที่ หน้า

สารบัญตาราง

หน้า

สารบัญภาพ

หน้า

หน้า