

โครงแบบคลาสลาเทกซ์อัตโนมัติจากภาพเอกสาร

นาย จุพงษ์ พานิชเกรียงไกร

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาเทคโนโลยีทางภาพ ภาควิชาวิทยาศาสตร์ทางภาพถ่ายและเทคโนโลยีทางการพิมพ์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2549
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

AUTOMATED LaTeX-CLASS CONFIGURATION FROM DOCUMENT IMAGE

Chulapong Panichkriangkrai

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Imaging Technology

Department of Imaging and Printing Technology

Faculty of Science

Chulalongkorn University

Academic Year 2006

Copyright of Chulalongkorn University

491598

หัวข้อวิทยานิพนธ์

โครงแบบคลาสลาเทกซ์อัตโนมัติจากภาพเอกสาร

โดย

นาย จุฬพงษ์ พานิชเกรียงไกร

สาขาวิชา

เทคโนโลยีทางภาพ

อาจารย์ที่ปรึกษา

อาจารย์ ดร. กิ่งกาญจน์ สุขคนาภิบาล

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วน
หนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

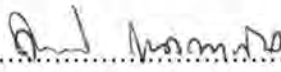


..... คณบดีคณะวิทยาศาสตร์
(ศาสตราจารย์ ดร. เปี่ยมศักดิ์ เมนะเศวต)

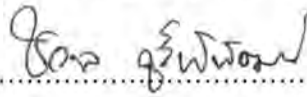
คณะกรรมการสอบวิทยานิพนธ์



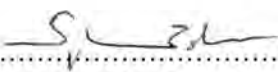
..... ประธานกรรมการ
(รองศาสตราจารย์ ดร. อรรณู หาญสืบสาย)



..... อาจารย์ที่ปรึกษา
(อาจารย์ ดร. กิ่งกาญจน์ สุขคนาภิบาล)



..... กรรมการ
(อาจารย์ ดร. ชวาล คุร์ทิพัฒน์)



..... กรรมการ
(อาจารย์ ดร. สัจจิตรา สือประसार)

จุฬพงษ์ พานิชเกรียงไกร : โครงแบบคลาสลาเทคซ้อัดโนมติจากภาพเอกสาร.
(AUTOMATED LaTeX-CLASS CONFIGURATION FROM DOCUMENT IMAGE)
อ. ที่ปรึกษา : อ.ดร. กิ่งกาญจน์ สุขคณาภิบาล, 103 หน้า.

งานวิจัยนี้มีจุดมุ่งหมายในการผลิต โปรแกรมเพื่อสร้างไฟล์คลาสลาเทคซ้อัดโนมติจากภาพเอกสาร ในปัจจุบันลาเทคซ้อัดโนมติได้รับความนิยมอย่างแพร่หลายเพื่อผลิตสิ่งพิมพ์คุณภาพสูง ซึ่งให้คุณภาพเท่าเทียมกันในทุกระบบงาน มีการเพิ่มประสิทธิภาพการใช้งานลาเทคซ้อัดโนมติด้วยการสร้างไฟล์คลาสลาเทคซ้อัดโนมติซึ่งเป็นการรวบรวมคำสั่งลาเทคซ้อัดโนมติต่าง ๆ ทำให้มีลักษณะเป็นโปรแกรมที่มีความซับซ้อน งานวิจัยนี้จึงทำให้ผู้ใช้การสร้างลาเทคซ้อัดโนมติได้สะดวกยิ่งขึ้น ในขั้นตอนการทดลองมีการทดสอบการวิเคราะห์ภาพเอกสารจากการสร้างขึ้นแก้ฉบับ และ ภาพเอกสารจากสำนักพิมพ์ที่มีชื่อเสียงอีกสามฉบับ โดยโปรแกรมนี้จะนำข้อมูลการจัดหน้าเอกสารที่ได้จากภาพเอกสาร ร่วมกับข้อมูลด้านตัวอักษรที่ผู้ใช้กำหนดขึ้น ผลิตเป็นไฟล์คลาสลาเทคซ้อัดโนมติได้อย่างสมบูรณ์

ภาควิชา วิทยาศาสตร์ทางภาพถ่ายและเทคโนโลยีทางการพิมพ์

ลายมือชื่อนิสิต.....*Wong*.....

สาขาวิชา เทคโนโลยีทางภาพ

ลายมือชื่ออาจารย์ที่ปรึกษา.....*Wong*.....

ปีการศึกษา 2549

4872250223 : MAJOR IMAGING TECHNOLOGY

KEY WORD: LaTeX-class / Document Analysis

CHULAPONG PANICHKRIANGKRAI : AUTOMATED LaTeX-CLASS
CONFIGURATION FROM DOCUMENT IMAGE. THESIS ADVISOR : KINGKARN
SOOKANAPHIBARN, Ph.D. , 103 pp.

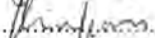
This research aims to develop program for automatically generating a LaTeX-class file from a document image. Recently LaTeX has been widely used to produce high-quality books, which can provide the same results in all systems. LaTeX has been written to extend the power of TeX typesetting system by collecting TeX commands, called LaTeX-class file. Basically, the LaTeX-class files are a complicated program. Automated LaTeX-class configuration for publishing was investigated in this research. The experiments were designed to test against document images of nine synthesized document images and three well-known academic publishers. Parameters corresponding to the layout are carried out from the document image as well as parameters corresponding to fonts defined by users. The program has successfully generated the LaTeX-class file from these parameters.

Department Imaging and Printing Technology

Field of study Imaging Technology

Academic year 2006

Student's signature.....

Advisor's signature.....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้เพราะมีบุคคลที่คอยช่วยเหลือ เป็นกำลังใจ ให้คำปรึกษาและคำแนะนำอยู่เสมอ ผมขอขอบพระคุณทุกคนที่มีส่วนช่วยเหลือตลอดระยะเวลาที่ผ่านมาดังมีรายนามต่อไปนี้

1. ขอขอบพระคุณ อ.ดร.กิงกาญจน์ สุขคณาภิบาล สำหรับคำแนะนำ และการช่วยเหลือในทุกด้าน ทำให้วิทยานิพนธ์ฉบับนี้เสร็จสมบูรณ์ได้
2. ขอขอบพระคุณ ประธานและกรรมการสอบทุกท่าน สำหรับคำแนะนำที่มีประโยชน์
3. ขอขอบพระคุณ อาจารย์ภาควิชาวิทยาศาสตร์ทางภาพถ่ายและเทคโนโลยีทางการพิมพ์ทุกท่าน สำหรับคำสอน ข้อแนะนำ และการช่วยเหลือในทุก ๆ เรื่อง
4. ขอขอบพระคุณ คณาจารย์ทุกท่านที่เคยสอนผมมาตั้งแต่อดีตจนถึงปัจจุบัน
5. ขอขอบคุณ เจ้าหน้าที่ภาควิชาวิทยาศาสตร์ทางภาพถ่ายและเทคโนโลยีทางการพิมพ์ สำหรับการแนะนำและช่วยเหลือในด้านต่าง ๆ
6. ขอขอบใจ เพื่อน ๆ นิสิตปริญญาโทสาขาวิชาเทคโนโลยีทางภาพ ที่ช่วยกันสร้างบรรยากาศที่ดีในการเรียนและทำงานตลอดสองปีที่ผ่านมา
7. ขอขอบใจ เพื่อน ๆ นักเรียนเก่าสาธิตจุฬาฯ รุ่นที่ 36 และ เพื่อน ๆ นิสิตเก่าเทคโนโลยีทางภาพและการพิมพ์รุ่นที่ 29 ทุกคน ที่คอยช่วยเหลือและเป็นกำลังใจให้ตลอดมา
8. ขอขอบพระคุณ ครอบครัวพานิชเกรียงไกร ที่คอยเป็นกำลังใจ ให้โอกาส ยอมรับและเข้าใจในสิ่งที่ผมทำตลอดมา และตลอดไป

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ฅ
สารบัญภาพ.....	ฉุ
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการวิจัย.....	8
1.3 ขอบเขตของการวิจัย.....	8
1.4 ประโยชน์ที่ได้รับ.....	11
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	12
2.1 แนวคิดและทฤษฎี.....	12
2.1.1 เทกซ์.....	12
2.1.2 การวิเคราะห์ภาพเอกสาร.....	13
2.2 เอกสารและงานวิจัยที่เกี่ยวข้อง.....	15
2.2.1 การเขียนไฟล์คลาสลาเทกซ์.....	15
2.2.2 การเขียนโปรแกรมภาษา MATLAB.....	16
2.2.3 การสร้างส่วนต่อประสานลักษณะกราฟิก (GUI).....	17
บทที่ 3 วิธีดำเนินการวิจัย.....	19
3.1 เครื่องมือที่ใช้ในการวิจัย.....	19
3.1.1 ส่วนของการวิเคราะห์ภาพเอกสาร.....	19
3.1.2 ส่วนของการสร้างไฟล์คลาสลาเทกซ์.....	31
3.1.3 ส่วนการติดต่อกับผู้ใช้งาน.....	31
3.2 การเก็บรวบรวมข้อมูล.....	33
3.3 การวิเคราะห์ข้อมูล.....	35
บทที่ 4 ผลการทดลองและวิเคราะห์ข้อมูล.....	36
4.1 ผลการวิเคราะห์.....	36
4.1.1 ค่าที่ได้จากการวิเคราะห์ภาพเอกสาร.....	36

4.1.2 ไฟล์คลาสลาเทกซ์ที่ได้จากการใช้ผลการวิเคราะห์.....	37
4.1.3 ภาพเอกสารใหม่ที่ได้จากการใช้ไฟล์คลาสลาเทกซ์ที่สร้างขึ้น.....	39
4.2 การเปรียบเทียบ.....	40
บทที่ 5 สรุปผลงานวิจัย อภิปรายผล และข้อเสนอแนะ.....	42
5.1 สรุปผลการวิจัย.....	42
5.2 อภิปรายผลการวิจัย.....	42
5.3 ข้อเสนอแนะ.....	42
รายการอ้างอิง.....	44
ภาคผนวก	
ภาคผนวก ก.....	46
ภาคผนวก ข.....	59
ภาคผนวก ค.....	90
ประวัติผู้เขียนวิทยานิพนธ์.....	103

บทที่	หน้า
ตารางที่ 3.1 ค่าการจัดหน้าของเอกสารต้นฉบับที่สร้างขึ้น (มิลลิเมตร).....	34
ตารางที่ 3.2 ค่าการจัดหน้าของเอกสารต้นฉบับที่มีการใช้งานจริง (มิลลิเมตร).....	35
ตารางที่ 4.1 ผลการทดลองการวิเคราะห์ภาพเอกสารต้นฉบับที่สร้างขึ้น (มิลลิเมตร).....	36
ตารางที่ 4.2 ผลการทดลองการวิเคราะห์ภาพเอกสารต้นฉบับที่มีการใช้งานจริง (มิลลิเมตร).....	36
ตารางที่ 4.3 การเปรียบเทียบความสูงและความกว้างของเนื้อหา ของภาพเอกสารที่สร้างขึ้น.....	40
ตารางที่ 4.4 การเปรียบเทียบระยะหัวกระดาษและระยะกั้นหน้ากระดาษ ของภาพเอกสารที่สร้างขึ้น.....	40
ตารางที่ 4.5 การเปรียบเทียบความสูงและความกว้างของเนื้อหา ของภาพเอกสารที่มีการใช้งานจริง.....	41
ตารางที่ 4.6 การเปรียบเทียบระยะหัวกระดาษและระยะกั้นหน้ากระดาษ ของภาพเอกสารที่มีการใช้งานจริง.....	41

บทที่	หน้า
รูปที่ 1.1 ไฟล์พีดีเอฟหน้าที่ 1 แสดงตัวอย่างการจัดหน้าของสำนักพิมพ์ SPRINGER VERLAG....	6
รูปที่ 1.2 ไฟล์พีดีเอฟหน้าที่ 2 แสดงตัวอย่างการจัดหน้าของสำนักพิมพ์ SPRINGER VERLAG....	6
รูปที่ 1.3 ไฟล์นำเข้า ภาพเอกสาร (“.TIF”).....	8
รูปที่ 1.4 ภาพตัวอย่างเอกสารที่ใช้ในการวิเคราะห์.....	9
รูปที่ 1.5 ตัวอย่างการจัดหน้าเอกสาร แสดงตัวแปรที่ศึกษา.....	10
รูปที่ 2.1 หลักการทำงานของไคเลชัน.....	14
รูปที่ 2.2 ตัวอย่างการใช้งานไคเลชัน.....	14
รูปที่ 2.3 หน้าต่างการสร้างส่วนต่อประสานลักษณะกราฟิก.....	17
รูปที่ 2.4 เครื่องมือที่ใช้ในการสร้างส่วนต่อประสานลักษณะกราฟิก.....	18
รูปที่ 3.1 หลักการทำงานของส่วนการวิเคราะห์ภาพเอกสาร.....	20
รูปที่ 3.2 ภาพเอกสารต้นฉบับ.....	21
รูปที่ 3.3 ภาพเอกสารต้นฉบับที่แปลงเป็นภาพขาวดำ 1 บิต.....	22
รูปที่ 3.4 ภาพเอกสารหลังจากการหาขอบเขตการจัดเอกสาร.....	23
รูปที่ 3.5 การหาระยะกั้นหน้าของเอกสาร.....	25
รูปที่ 3.6 การหาระยะกั้นหลังของเอกสาร.....	26
รูปที่ 3.7 การหาระยะหัวกระดาษ.....	27
รูปที่ 3.8 การหาระยะท้ายกระดาษ.....	28
รูปที่ 3.9 การหาระยะห่างระหว่างคอลัมน์.....	29
รูปที่ 3.10 ส่วนต่อประสานงานลักษณะกราฟิก.....	31
รูปที่ 3.11 ส่วนนำเข้าภาพเอกสาร.....	31
รูปที่ 3.12 ส่วนคำนวณและแสดงภาพ.....	32
รูปที่ 3.13 ส่วนตั้งค่าน้ำกระดาษ.....	32
รูปที่ 3.14 ส่วนการตั้งค่าชุดอักษรที่ใช้งาน.....	33
รูปที่ 3.15 ส่วนการบันทึกไฟล์คลาสลาเทกซ์.....	33
รูปที่ 3.16 ลักษณะของเอกสารตัวอย่าง.....	34
รูปที่ 4.1 ตัวอย่างภาพเอกสารที่สร้างไฟล์คลาสลาเทกซ์ จากภาพตัวอย่างที่ 3.....	39