

การสร้างระบบการจัดรูปแบบข้อความภาษาไทยภายใต้ระบบปฏิบัติการยูนิกซ์



นางสาว นิษฐิตา นวลศรี

ศูนย์วิทยทรัพยากร

จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

ภาควิชาวิศวกรรมคอมพิวเตอร์

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

พ.ศ. 2531

ISBN 974-568-513-5

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

013801

013801

I15983845

A Construction of Thai Text Formatting System
under UNIX Operating System



Miss Nittida Nuansri

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science
Department of Computer Engineering
Graduate School
Chulalongkorn University

1988

ISBN 974-568-513-5

หัวข้อวิทยานิพนธ์ การสร้างระบบการจัดรูปแบบข้อความภาษาไทยภายใต้ระบบปฏิบัติการยูนิคซ์
โดย นางสาว นิษจิตา นวลศรี
ภาควิชา วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษา ดร.วิเศษ นครชัย
ผู้ช่วยศาสตราจารย์ สุ่มเมธ วัชรชัยสุรพล



บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัยรับเป็น ส่วนหนึ่งของ
การศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

.....
(ศาสตราจารย์ ดร.ถาวร วัชรชัย)

คณะกรรมการสอบวิทยานิพนธ์

.....
(ดร.ยรรยง เต็งอำนาจ)

.....
(ดร.วิเศษ นครชัย)

.....
(ผู้ช่วยศาสตราจารย์ สุ่มเมธ วัชรชัยสุรพล)

.....
(ผู้ช่วยศาสตราจารย์ วิษัญญ์ เลิศวิภาตระกูล)



นิษฐิศา นวลศรี การสร้างระบบการจัดรูปแบบข้อความภาษาไทยภายใต้ระบบปฏิบัติการยูนิกซ์
(A CONSTRUCTION OF THAI TEXT FORMATTING SYSTEM UNDER UNIX OPERATING SYSTEM) อ.พีรภิกษา คร.วิเศษ นครชัย, ผศ. สุเมธ วัชรชัยสุรพล, 109 หน้า

การผลิตเอกสารในรูปแบบต่างๆ เช่น การพิมพ์จดหมาย บันทึกข้อความ การพิมพ์บทความ ตำรานับเป็นงานที่ต้องใช้เวลา การนำคอมพิวเตอร์มาช่วยในการผลิตเอกสารดังกล่าว จะช่วยให้การทำงานสะดวก รวดเร็ว และมีประสิทธิภาพมากขึ้น ปัจจุบันมีโปรแกรมอรรถประโยชน์ที่ใช้ในการผลิตเอกสาร เช่น nroff/troff ที่ใช้ภายใต้ระบบปฏิบัติการยูนิกซ์ หรือ format scribe ที่ใช้ในเครื่อง IBM ซึ่งใช้ในการผลิตเอกสารภาษาอังกฤษ หรือภาษาทางตะวันตกทั่วไป ไม่สามารถนำมาใช้ในการผลิตเอกสารภาษาไทยได้ เนื่องจากลักษณะที่แตกต่างกันของภาษา โดยเฉพาะในภาษาไทยมีการเขียนหลายระดับ และการแบ่งพยางค์ซึ่งไม่มีรูปแบบที่แน่นอน การใช้ภาษาไทยภายในเครื่องคอมพิวเตอร์ทั่วไป จะต้องทำการแก้ไขหรือดัดแปลงโปรแกรมที่มีอยู่ แต่ในระบบปฏิบัติการยูนิกซ์ไม่สามารถดัดแปลงโปรแกรมให้รับข้อความภาษาไทยได้ เนื่องจากปัญหาการใช้ภาษาไทย ยังไม่ได้รับการแก้ไขให้เป็นมาตรฐานเดียวกัน ขณะนี้ภายในระบบปฏิบัติการยูนิกซ์ ไม่มีชุดอักษรภาษาไทย หากจะต้องทำการเพิ่มเติมเข้าไปด้วยทุกครั้งที่ต้องการใช้ หรือด้วยวิธีการอื่นๆ ซึ่งขึ้นอยู่กับบริษัทผู้ขาย ทำให้การใช้ภาษาไทยภายใต้ระบบปฏิบัติการยูนิกซ์มีหลายรูปแบบ จึงยากแก่การพัฒนาโปรแกรมให้มีมาตรฐานเดียวกันได้ อย่างไรก็ตาม ระบบปฏิบัติการยูนิกซ์มีแนวโน้มว่าจะมีบทบาทในวงการคอมพิวเตอร์ของไทยมากขึ้น ดังนั้นการพัฒนาโปรแกรมที่สามารถรับข้อความภาษาไทยได้จึงเป็นปัญหาเร่งด่วนที่ควรสนใจ เพื่อแก้ไขปัญหานี้เฉพาะหน้าเกี่ยวกับการผลิตเอกสารภายใต้ระบบปฏิบัติการยูนิกซ์ในขณะนี้ จึงได้ทำการสร้างระบบการจัดรูปแบบข้อความภาษาไทยภายใต้ระบบปฏิบัติการยูนิกซ์ขึ้น ซึ่งประกอบด้วยโปรแกรมหลัก 4 โปรแกรม คือ

โปรแกรมบรรณาธิกรข้อความภาษาไทย tvi ใช้สำหรับรับและแสดงผลข้อความภาษาไทยในรูปแบบการเขียนที่ถูกต้องตามลักษณะการเขียนภาษาไทย โดยใช้รหัสภาษาไทยของสมาคมมาตรฐานอุตสาหกรรม (สมอ.) การทำงานของโปรแกรม tvi นี้ใช้รูปแบบเดียวกันกับ vi ซึ่งเป็นโปรแกรมบรรณาธิกรที่ใช้ภายใต้ระบบปฏิบัติการยูนิกซ์

โปรแกรมแบ่งพยางค์ภาษาไทย ใช้สำหรับแบ่งพยางค์ข้อความภาษาไทย เพื่อช่วยในการจัดขอบการพิมพ์เอกสารผลลัพธ์ ในที่นี้ใช้อัลกอริทึมช่วยในการแบ่งพยางค์

โปรแกรมจัดข้อความเอกสารภาษาไทย tnroff ใช้สำหรับจัดรูปแบบเอกสารตามที่กำหนด โดยนำข้อความเอกสารภาษาไทยที่ได้จากโปรแกรม tvi เป็นข้อความนำเข้า และแสดงผลที่ได้จากการจัดรูปแบบแล้วเป็นเอกสารผลลัพธ์

โปรแกรมควบคุมการพิมพ์เอกสารผลลัพธ์ทางอุปกรณ์การพิมพ์ต่างๆ สำหรับใช้ในการพิมพ์ผลลัพธ์ทางเครื่องพิมพ์ใดๆ ได้ ทำให้ไม่ถูกจำกัดด้วยอุปกรณ์การพิมพ์ต่างชนิดกัน

ภาควิชา วิศวกรรมคอมพิวเตอร์
สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์
ปีการศึกษา 2530

ลายมือชื่อนิสิต
ลายมือชื่ออาจารย์ที่ปรึกษา
.....



NITTIDA NUANSRI : A CONSTRUCTION OF THAI TEXT FORMATTING SYSTEM UNDER UNIX OPERATING SYSTEM. THESIS ADVISORS : VIZES NAKORNCHAI, Ph.D., ASST. PROF. SUMET VACHARACHAISURAPOL, 109 pp.

Producing all kinds of repeated publications such as typing letters, memos or articles is time consuming and expensive; therefore, the computer applications take advantage of producing them more quickly and efficiently. At the present, there is a text formatter program such as nroff/troff, format and scribe. However, these programs accept English or European texts as input and can't apply to Thai text due to a few differences between Thai and English characteristics; especially Thai writing style has different position of vowels. Furthermore, the Thai syllable separation is not consistent. Therefore, the system utility programs are not portable with Thai character especially in UNIX. In particular, there are no Thai characters in the original UNIX operating systems. They must be installed into system. In actual use, there are various methods to implement them into UNIX, each one is depended on distributors. As seen here, it is very difficult to develop any compatible softwares. In fact, UNIX becomes so popular that Thai utility development should be urgently considered to solve the problem of Thai processing under UNIX.

This research involved with Thai text formatting system under UNIX and composed of 4 main programs suggested as following.

Tvi, the Thai editor, used to manipulate Thai text by means of using ISA Thai character code. This program is like vi editor; the popular screen editor in UNIX.

Tnroff, the Thai text formatter program, it formats Thai input and sends the result text to standard output or printer.

Thai syllable separation program used to separate Thai text into syllable by using algorithms in order to rearrange the margin of output in program tnroff.

Device driver program used to generate output for various printers without any modifications to the source code.

ภาควิชา วิศวกรรมคอมพิวเตอร์
สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์
ปีการศึกษา 2530

ลายมือชื่อนิสิต
ลายมือชื่ออาจารย์ที่ปรึกษา



กิตติกรรมประกาศ

การทำวิทยานิพนธ์ครั้งนี้สำเร็จได้โดยได้รับความช่วยเหลือจาก ผู้ช่วยศาสตราจารย์ สุ่มเมธ วัชรชัยสุรณ และ ดร.วิเศษ นครชัย อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งท่านได้กรุณาให้ คำสอนและแนะนำเกี่ยวกับการทำวิจัยตลอดมา กราบขอบพระคุณอาจารย์ทั้งสองท่านมา ณ ที่นี้

อนึ่ง การทำวิจัยเรื่องนี้ไม่อาจประสบผลสำเร็จได้ หากมิได้รับความช่วยเหลือจาก Dr. Ken J. McDonell แห่งภาควิชาวิทยาศาสตร์คอมพิวเตอร์ มหาวิทยาลัยโมนาช ประเทศ ออสเตรเลีย ในเรื่องเกี่ยวกับการประมวลผลข้อความภาษาไทยได้ระบบปฏิบัติการยูนิกซ์ และ กรูณาเอื้อเพื่อ โปรแกรมที่ใช้สำหรับการแสดงผลข้อความทางจอภาพของยูนิกซ์ เพื่อตัดแปลงให้รับ และแสดงข้อความภาษาไทยได้ จึงขอขอบคุณ Dr. Ken J. McDonell เป็นอย่างสูง

ขอขอบคุณศูนย์คอมพิวเตอร์ มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่ ที่กรุณา ให้ใช้สถานที่รวมทั้งเครื่องคอมพิวเตอร์ และอุปกรณ์อื่นๆ ในการทำวิทยานิพนธ์ครั้งนี้

นอกจากนี้ ขอขอบคุณ คุณณรงค์ ไชยกุล รองผู้จัดการบริษัทไทยนำคอมพิวเตอร์จำกัด หาดใหญ่ ที่กรุณาให้คำปรึกษาเกี่ยวกับการใช้ภาษาไทยในคอมพิวเตอร์ คุณสุคนธ์ ไชยวงศ์ คุณอารี จันทสุวีร์ และคุณสิงวาล สวัสดิ์ เจ้าหน้าที่พิมพ์ดีด ศูนย์คอมพิวเตอร์ มหาวิทยาลัยสงขลานครินทร์ สำหรับความช่วยเหลือในด้านภารกิจต้นฉบับวิทยานิพนธ์

นิษิณีดา นวลศรี

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
กิตติกรรมประกาศ	ฉ
รายการตารางประกอบ	ญ
รายการรูปประกอบ	ฎ
บทที่	
1 บทนำ	1
1.1 ความเป็นมาของปัญหา	1
1.2 วัตถุประสงค์ของการวิจัย	4
1.3 ขอบเขตการวิจัย	5
1.4 ที่ตั้งของการวิจัย	5
1.5 ประโยชน์ที่ได้รับจากการวิจัย	5
2 ระบบการประมวลผลข้อความในเอกสาร	6
2.1 การจัดรูปแบบคำ	6
2.2 การประมวลผลคำ	7
2.3 การจัดรูปแบบเอกสารภายใต้ระบบปฏิบัติการยูนิกซ์	8
3 โปรแกรมบรรณาธิการ vi	13
3.1 โปรแกรมบรรณาธิการพื้นฐานของระบบปฏิบัติการยูนิกซ์	14
3.2 vi	14
3.3 ลักษณะสำคัญของ vi	14
3.4 ลักษณะงานที่ทำของ vi	16
3.5 ลักษณะเด่นอื่นๆ ของ vi	25
4 โปรแกรมจัดรูปแบบเอกสาร nroff	28
4.1 แนวความคิดเบื้องต้นของ nroff	28
4.2 รูปแบบการใช้ nroff	29

สารบัญ (ต่อ)

๗

	หน้า
4.3 ลักษณะงานที่ทำภายใต้ nroff	29
5 การแบ่งพยางค์ภาษาไทย	39
5.1 คุณลักษณะของอักษรภาษาไทย	39
5.2 การแบ่งพยางค์ภาษาไทย	45
6 โปรแกรมบรรณาธิการภาษาไทย tvi	51
6.1 โครงสร้างข้อมูล	51
6.2 การจัดการบนจอภาพ	57
6.3 การจัดการเกี่ยวกับการค้นหาข้อความ	59
6.4 การแสดงผลข้อความภาษาไทย	61
6.5 คำสั่งของ tvi	63
7 โปรแกรมจัดเอกสารภาษาไทย tnroff	67
7.1 การแบ่งพยางค์ภาษาไทยภายใต้โปรแกรม tnroff	67
7.2 device driver	72
7.3 คำสั่งของ tnroff	73
8 สรุปผลการวิจัยและข้อเสนอแนะ	76
8.1 สรุปผลการวิจัย	76
8.2 ข้อเสนอแนะ	77
เอกสารอ้างอิง	79
ภาคผนวก ก. รหัสมาตรฐานภาษาไทยของสมาคมมาตรฐานอุตสาหกรรม	84
ข. รหัสภาษาไทยบนเครื่องพิมพ์ของบริษัทบางกอกดาต้า เซ็นเตอร์ จำกัด	85
ค. รหัสภาษาไทยบนเครื่องพิมพ์ของบริษัทสหวิทยานินเตอร์เนชั่นแนล คอมพิวเตอร์จำกัด	86
ง. สรุปคำสั่ง tvi	87
จ. สรุปคำสั่ง tnroff	90
ฉ. แสดงข้อความนำเข้าและผลลัพธ์ที่ได้จากการใช้โปรแกรม tnroff	91

สารบัญ (ต่อ)


๕. แสดงผังการทำงานของระบบการจัดรูปแบบข้อความภาษาไทยภายใต้ ระบบปฏิบัติการยูนิกซ์	100
๖. แสดงการทำงานของแต่ละฟังก์ชัน	112
ประวัติผู้เขียน	118



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

รายการตารางประกอบ

ตารางที่		หน้า
7.1	แสดงรูปแบบการแบ่งพยางค์ภาษาไทยโดยใช้สระเป็นตัวกำหนดจุดแบ่ง	69
7.2	แสดงรูปแบบและการทำงานของคำสั่งในโปรแกรม tnroff	73



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

รายการรูปประกอบ

รูปที่	หน้า
6.1 แสดงรูปแบบโครงสร้างข้อมูล	52
6.2 แสดงการเก็บข้อความครั้งแรกในบัฟเฟอร์	53
6.3 แสดงการเก็บข้อความภายในบัฟเฟอร์ก่อนการแทรกข้อความ	54
6.4 แสดงการเก็บข้อความภายในบัฟเฟอร์หลังการแทรกข้อความ	54
6.5 แสดงการเก็บข้อความภายในบัฟเฟอร์ก่อนการลบข้อความ	55
6.6 แสดงการเก็บข้อความภายในบัฟเฟอร์หลังการลบข้อความ	56
6.7 แสดงการเก็บข้อความในแฟ้มข้อมูลแบบข้อความ	56
7.1 แสดงการกำหนดจุดเริ่มต้นสำหรับการหาจุดแบ่งพยางค์ภาษาไทย	68

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย