

การออกแบบและพัฒนาส่วนเชื่อมโยงสำหรับการตัดคำ  
และการแทรกอักขระแบ่งคำในภาษาไทย



นาย บุญเรือง ธนาสุนทรไพศาล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาคณะหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

ภาควิชาวิศวกรรมคอมพิวเตอร์

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

พ.ศ. 2533

ISBN 974-877-101-5

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

016291

I10306857

DESIGN AND DEVELOPMENT OF INTERFACE FOR THAI WORD SEPARATION  
AND SEPARATOR INSERTION

Mr. Eonrueng Tanasuntanpaisan

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

1990

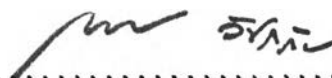
ISBN 974-577-101-5

หัวข้อวิทยานิพนธ์ การออกแบบและพัฒนาล้วนเชื่อมโยงสำหรับการตัดคำ  
และการแทรกอักขระแบ่งคำภาษาไทย

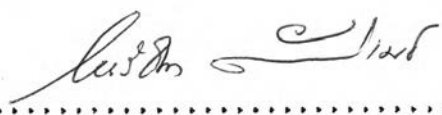
โดย นาย บุญเรือง ชนาสนทรไพศาล  
ภาควิชา วิศวกรรมคอมพิวเตอร์  
อาจารย์ที่ปรึกษา อาจารย์ ดร. ยรรยง เต็งอำนวยการ




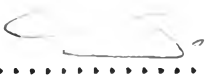
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัยนี้เป็นส่วนหนึ่ง  
ของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

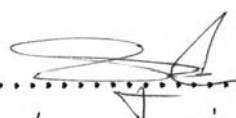
  
..... คณะบดีบัณฑิตวิทยาลัย  
( ศาสตราจารย์ ดร. ทวาร วัชรภักย์ )

คณะกรรมการสอบวิทยานิพนธ์

  
..... ประธานกรรมการ  
( รองศาสตราจารย์ ไกรวิชิต ตันติเมธ )

  
..... อาจารย์ที่ปรึกษา  
( อาจารย์ ดร. ยรรยง เต็งอำนวยการ )

  
..... กรรมการ  
( รองศาสตราจารย์ สมชาย ทยานอง )

  
..... กรรมการ  
( อาจารย์ จารุมাত্র ปันทอง )

พิมพ์ต้นฉบับบทคัดย่อ วิทยานิพนธ์ภายในกรอบสี่เหลี่ยมนี้เพียงแผ่นเดียว



บทคัดย่อ ธานัฐพร ไพศาล : การออกแบบและพัฒนาส่วนเชื่อมต่อโยยสำหรับการตัดคำและการแทรกอักขระแบ่งคำในภาษาไทย (DESIGN AND DEVELOPMENT OF INTERFACE FOR THAI WORD SEPARATION AND SEPARATOR INSERTION)

อ.ที่ปรึกษา : อ. ดร. ยรรยง เต็งคำณวย, 43 หน้า, ISEI 974-537-101-5

วิทยานิพนธ์ฉบับนี้เป็นการออกแบบส่วนเชื่อมต่อโยยสำหรับมอดูลการตัดคำภาษาไทย เพื่อให้สามารถใช้กับโปรแกรมประยุกต์อื่นๆได้ นอกเหนือจากโปรแกรมประมวลผลคำภาษาไทย ตัวเชื่อมต่อโยยที่ออกแบบจะไม่ใช่ขึ้นอยู่กับรูปแบบของขั้นตอนวิธีการตัดคำ และสามารถรองรับรูปแบบการตัดคำทั้งแบบกลุ่มและแบบโต้ตอบ นอกจากนี้ยังได้ออกแบบและพัฒนาขั้นตอนวิธีการแทรกอักขระแบ่งคำลงหน้าคำแต่ละคำ โดยยังสามารถแยกความแตกต่างระหว่างอักขระแบ่งคำกับข้อมูลจริยได้ อักขระแบ่งคำสามารถใช้ตัวอักษรใดก็ได้ตามความเหมาะสม จากตัวเชื่อมต่อโยยมอดูลการตัดคำและขั้นตอนวิธีการแทรกอักขระแบ่งคำข้างต้น ได้ทำการพัฒนาฟังก์ชันบรรทัดประโยชน์ขึ้น เพื่อให้โปรแกรมประยุกต์สามารถเรียกใช้มอดูลการตัดคำ โดยงานตัวเชื่อมต่อโยย มอดูลการตัดคำ ได้โดยสะดวก

ภาควิชา วิศวกรรมคอมพิวเตอร์  
สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์  
ปีการศึกษา 2532

ลายมือชื่อนิสิต [Signature]  
ลายมือชื่ออาจารย์ที่ปรึกษา [Signature]



พิมพ์ต้นฉบับบทคัดย่อวิทยานิพนธ์ภายในกรอบสี่เหลี่ยมนี้เพียงแผ่นเดียว

BOONRUENG TANASUNTONPAISAN : DESIGN AND DEVELOPMENT OF INTERFACE FOR THAI WORD SEPARATION AND SEPARATOR INSERTION.

THESIS ADVISOR : YUNYONG TENGAMNUAY, Ph.D. 43 pp.

The purpose of this thesis is to design the interface for Thai word separator routine so that the routine can be used not only with Thai word processor program but also any application program. The interface is independent from methodology of word separator algorithm and can be used with both batch and interactive word separator applications. The separator insertion algorithm is designed for inserting separator after each word while still can recognize the difference between separator and real data. The separator can be any reasonable character. The utility functions were designed so that application can use it easily. The implementation of utility functions are based on interface of Thai word separator routine and separator insertion algorithm.

ภาควิชา วิศวกรรมคอมพิวเตอร์

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์

ปีการศึกษา 2532

ลายมือชื่อนิสิต 

ลายมือชื่ออาจารย์ที่ปรึกษา Yun Yong Tengamnuay

ลายมือชื่ออาจารย์ที่ปรึกษาช่วย

กิตติกรรมประกาศ



วิทยานิพนธ์ฉบับนี้ สำเร็จลุล่วงไปได้ด้วยความช่วยเหลือจากบุคคลหลายท่าน ผู้เขียนขอขอบคุณ อาจารย์ ดร. ยรรยง เต็งอำนวยการ และ อาจารย์ จารุมাত্র ปิ่นทอง เป็นอย่างสูง สำหรับคำปรึกษา คำแนะนำและ แนวคิดที่เป็นประโยชน์อย่างมากสำหรับวิทยานิพนธ์ฉบับนี้ ขอขอบคุณ คุณพลสมิทธิ์ ปาละกุล ที่เป็นกำลังใจให้ตลอดมา ขอขอบคุณ คุณพ่อคุณแม่สำหรับทุกสิ่งทุกอย่าง ขอขอบคุณ คุณอุรฉัตร จินตนาวงศ์ ที่ช่วยพิมพ์วิทยานิพนธ์อย่างไม่เห็นแก่เห็นดเห็น้อย ขอขอบคุณเพื่อนและพี่น้องอีกหลายท่านซึ่งมิได้เอ่ยนามไว้ ณ ที่นี้ที่ช่วยเหลือและให้คำแนะนำในหลายๆด้านที่ขาดเสียไม่ได้ คือ ขอขอบคุณ คุณบุญเรือง ธนาสุนทรไพศาล สำหรับความตั้งใจและพยายามจนกระทั่งวิทยานิพนธ์ฉบับนี้สำเร็จสมบูรณ์

สารบัญ



หน้า

บทคัดย่อภาษาไทย .....	ง
บทคัดย่อภาษาอังกฤษ .....	จ
กิตติกรรมประกาศ .....	ฉ
สารบัญตาราง .....	ญ
สารบัญภาพ .....	ฎ

บทที่

1. บทนำ .....	1
1.1 ความเป็นมาของปัญหา .....	1
1.1.1 ลักษณะของภาษาไทยและปัญหา .....	1
1.1.2 การตัดคำกับโปรแกรมประมวลผลคำ .....	1
1.2 วัตถุประสงค์ของวิทยานิพนธ์ .....	2
1.3 ขอบเขตของวิทยานิพนธ์ .....	2
1.4 ประโยชน์ที่คาดว่าจะได้รับจากการทำวิทยานิพนธ์ .....	3
1.5 เนื้อหาของวิทยานิพนธ์ .....	3
2. การตัดคำภาษาไทย .....	4
2.1 ประเภทของขั้นตอนวิธีการตัดคำภาษาไทย .....	4
2.2 ประโยชน์ของขั้นตอนวิธีการตัดคำภาษาไทย .....	4
3. ตัวเชื่อมโยงการตัดคำ .....	6
3.1 การออกแบบโครงสร้าง .....	6
3.1.1 การเชื่อมต่อและความยืดหยุ่น .....	6
3.1.2 การปกปิดข่าว .....	8

3.2	ตัวเชื่อมโยง .....	9
3.2.1	ตัวเชื่อมโยงมอดูลการตัดคำ ในโปรแกรมประมวลผลคำ .....	9
3.2.2	รูปแบบของการตัดคำแบบกลุ่มและโต้ตอบ .....	10
3.3	การออกแบบมอดูลการตัดคำ .....	12
3.3.1	โครงสร้างข้อมูลและพฤติกรรมของ ขั้นตอนวิธีการตัดคำ .....	12
3.3.2	ตัวเชื่อมโยง .....	12
3.4	ตัวเชื่อมโยงมอดูลการตัดคำในภาษาซี .....	14
3.5	การใช้วิธีการเชื่อมโยงสำหรับภาษาอื่น .....	15
3.6	การใช้ตัวเชื่อมโยงข้ามภาษา .....	16
3.7	องค์ประกอบข้างเคียงอื่นๆเกี่ยวข้องกับการตัดคำ .....	16
3.8	สรุป .....	17
4.	อักขระแบ่งคำ .....	18
4.1	ความหมายของอักขระแบ่งคำ .....	18
4.2	ชนิดของอักขระแบ่งคำ .....	18
4.3	ปัญหาของการแทรกอักขระแบ่งคำ .....	18
4.4	การเลือกอักขระแบ่งคำ .....	19
5.	ขั้นตอนวิธีการแทรกอักขระแบ่งคำ .....	22
5.1	การสอดไส้บิต .....	22
5.2	การฆ่าความหมาย .....	23
5.3	การออกแบบและพัฒนาขั้นตอนวิธีการแทรกอักขระแบ่งคำ .....	23
5.4	สรุป .....	31
6.	ฟังก์ชันอรรถประโยชน์ .....	33
6.1	การออกแบบฟังก์ชันอรรถประโยชน์ .....	33
6.2	ฟังก์ชันพื้นฐาน .....	34



6.3	ฟังก์ชันอรรถประโยชน์ .....	35
6.4	การใช้ฟังก์ชันอรรถประโยชน์ .....	36
7.	สรุป .....	39
7.1	การประยุกต์ใช้งาน .....	39
7.2	ข้อเสนอแนะ .....	39
	เอกสารอ้างอิง .....	41
	ประวัติผู้เขียน .....	43

## สารบัญตาราง

ตารางที่		หน้า
4.1	แสดงความถี่ของตัวอักษรไทย (เป็น เปอร์เซ็นต์) .....	20

## สารบัญภาพ

ภาพที่		หน้า
3.1	แสดงระดับของการเชื่อมต่อ .....	7
3.2	แสดงระดับของความยืดหยุ่น .....	8
3.3	แสดงความสัมพันธ์ระหว่างมอดูลการตัดคำและ โปรแกรมประมวลผลคำ .....	10
3.4	แสดงปัญหาการใช้มอดูลการตัดคำของ โปรแกรมประมวลผลคำกับ โปรแกรมใช้งานอื่น .....	11
5.1	การทำความเข้าใจและการแทรกอักขระแบ่งคำ .....	22
5.2	การลดความซ้ำซ้อน .....	23
5.3	ความซ้ำซ้อนกับอักขระแบ่งคำตามธรรมชาติ .....	25
5.4	การสืบสวนของอักขระแบ่งคำ .....	26
5.5	ปัญหาของการใช้อักษรปกติเป็นตัวสร้างอักขระแบ่งคำพิเศษ .....	27
5.6	การใช้อักขระแบ่งคำพิเศษ .....	28
5.7	การใช้อักขระแบ่งคำพิเศษกับคำที่ซับซ้อน .....	29
6.1	ความสัมพันธ์ของวัตถุและการกระทำ .....	33
6.2	ขั้นตอนการนำฟังก์ชันอรรถประโยชน์ไปใช้ .....	37