

การพัฒนาโปรแกรมสร้างพจนานุกรมเสมือน



นางสาว อรศรี สงวนชาติศรไกร

ศูนย์วิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

ภาควิชาวิศวกรรมคอมพิวเตอร์

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

พ.ศ. 2535

ISBN 974-582-369-4

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

019711 117391520

DEVELOPMENT OF A VIRTUAL DICTIONARY



Miss Orasri Sanguanchartsornkrai

ศูนย์วิจัยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirement
for the Degree of Master of Science

Graduate School

Chulalongkorn University

1992

ISBN 974-582-369-4

หัวข้อวิทยานิพนธ์ การพัฒนาโปรแกรมสร้างพจนานุกรมเสมือน
โดย นางสาว อรศรี สงวนชาติศรไกร
ภาควิชา วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษา อาจารย์ ดร. ยรรยง เต็งอำนวย



บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัย
เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

..... คณบดีบัณฑิตวิทยาลัย
(ศาสตราจารย์ ดร. ถาวร วิษกรภักย์)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(รองศาสตราจารย์ สมชาย ทยานยง)

.....อาจารย์ที่ปรึกษา
(อาจารย์ ดร. ยรรยง เต็งอำนวย)

.....กรรมการ
(รองศาสตราจารย์ มณฑนา ปราการสมุทร)

.....กรรมการ
(อาจารย์ จารุมাত্র ปันทอง)

พิมพ์ต้นฉบับบทคัดย่อวิทยานิพนธ์ภายในกรอบสี่เหลี่ยมนี้เพียงแผ่นเดียว

อรศรี สงวนชาติศรไกร : การพัฒนาโปรแกรมสร้างพจนานุกรมเสมือน
(DEVELOPMENT OF A VIRTUAL DICTIONARY) อ.ที่ปรึกษา :
ดร. ขรรขง เต็งอำนาจ, 95 หน้า. ISBN 974-582-369-4



วิทยานิพนธ์ฉบับนี้เป็นการออกแบบพจนานุกรมเสมือน เป็นโมดูลหนึ่งในโปรแกรมตัดคำภาษาไทย เนื่องจากการนำพจนานุกรมเข้ามาไว้ในหน่วยความจำหลักทั้งหมด เมื่อคำศัพท์จำนวนมากขึ้น จะทำให้หน่วยความจำหลักเต็มไม่สามารถเพิ่มคำศัพท์ลงในพจนานุกรมได้อีก จึงมีการแก้ไขโครงสร้างของพจนานุกรม ให้สามารถแบ่งส่วนได้ เมื่อนำพจนานุกรมเข้าสู่หน่วยความจำหลัก จะนำเข้ามาเพียงบางส่วน นอกจากนี้ยังได้สร้างตารางแสดงความสัมพันธ์ระหว่างตัวอักษรในคำศัพท์กับตำแหน่งของตัวอักษรนั้นๆ

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา
สาขาวิชา
ปีการศึกษา 2535

ลายมือชื่อนิติกร
ลายมือชื่ออาจารย์ที่ปรึกษา
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม

C210703 : MAJOR COMPUTER
KEY WORD: VIRTUAL DICTIONARY

ORASRI SANGUANCHARTSORNKRAI : DELVELOPMENT OF A VIRTUAL
DICTIONARY. THESIS ADVISOR : DR. YONGYUNG THENG-AM-NOY
95 pp. ISBN 974-582-369-4

The propose of this thesis is to present the design of virtual dictionary which is used in the process of word cutting in Thai language. It isn't possible to store large amount of vocabularies in main memory, which highly affects the insertion of new words to the dictionary. Therefore, the dictionary structure organized so that it can be paged. Words from the dictionary can be retrieved by using "The Least Recently Used (LRU)" concept which only some parts of the dictionary loaded into main memory. Besides, the table of relation between characters of words and their position in main memory are also created.



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา ๐๓๓๓ ๐๓๓๓ ๐๓๓๓

สาขาวิชา ๐๓๓๓ ๐๓๓๓ ๐๓๓๓

ปีการศึกษา ๒๕๓๕

ลายมือชื่อนิสิต. ๐๓๓๓ ๐๓๓๓ ๐๓๓๓

ลายมือชื่ออาจารย์ที่ปรึกษา. ๐๓๓๓ ๐๓๓๓ ๐๓๓๓

ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ สำเร็จลุล่วงไปได้ด้วยความช่วยเหลือจากบุคคลหลายท่าน ผู้เขียนขอขอบพระคุณ อาจารย์ ดร. ยรรยง เต็งอำนวย เป็นอย่างสูงสำหรับคำปรึกษา คำแนะนำ และแนวคิดที่เป็นประโยชน์อย่างมากสำหรับวิทยานิพนธ์ฉบับนี้ ขอขอบคุณ คุณสัมพันธ์ ธารินทร์ และเพื่อนๆ พี่น้อง ที่ให้คำแนะนำในหลายๆด้าน จนกระทั่งวิทยานิพนธ์ฉบับนี้สำเร็จสมบูรณ์

ท้ายนี้ ผู้วิจัยใคร่ขอกราบขอบพระคุณ บิดา-มารดา ซึ่งสนับสนุนในด้านการเงิน และให้กำลังใจแก่ผู้วิจัยเสมอมาจนสำเร็จการศึกษา

อรศรี สงวนชาติศรไกร



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ



	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญตาราง.....	ณ
สารบัญภาพ.....	ญ
บทที่	
1. บทนำ	
1.1 ความเป็นมาของปัญหา.....	1
1.1.1 ลักษณะของภาษาไทยและปัญหา.....	1
1.1.2 การแบ่งคำ.....	1
1.1.2.1 การแบ่งคำโดยใช้กฎ.....	1
1.1.2.2 การแบ่งคำโดยใช้พจนานุกรม.....	1
1.2 วัตถุประสงค์ของวิทยานิพนธ์.....	2
1.3 ขอบเขตของวิทยานิพนธ์.....	2
1.4 ประโยชน์ที่คาดว่าจะได้รับจากการทำวิทยานิพนธ์.....	2
1.5 เนื้อหาของวิทยานิพนธ์.....	3
2. หลักการของโครงสร้างพจนานุกรม	
2.1 โครงสร้างข้อมูลแบบทราเย.....	4
2.2 โครงสร้างข้อมูลแบบทราเยในภาษาซี.....	4
2.3 การตัดคำของโครงสร้างข้อมูลแบบทราเย.....	6
3. หลักการของพจนานุกรมเสมือน	
3.1 การสร้างพจนานุกรมเสมือน.....	8
3.2 เทคนิคการทำ Block Mapping.....	12

3.3	วิธีการแทนที่.....	22
4.	การออกแบบพจนานุกรมเสมือน	
4.1	ลักษณะการจัดเก็บพจนานุกรม.....	30
4.1.1	การจัดเก็บพจนานุกรมตามลำดับการเชื่อมโยง ในทางลึก (Depth First Search).....	30
4.1.2	การจัดเก็บพจนานุกรมตามลำดับการเชื่อมโยง ในทางกว้าง (Breath Frist Search).....	33
4.2	โครงสร้างพจนานุกรมเสมือน.....	36
4.3	ฟังก์ชันพื้นฐาน.....	38
4.4	ขั้นตอนการสร้างพจนานุกรมเสมือน.....	39
4.5	ขั้นตอนการนำพจนานุกรมเสมือนมาใช้งาน.....	40
5.	สรุป	
5.1	สรุปผล.....	41
5.2	ข้อเสนอแนะ.....	43
	เอกสารอ้างอิง.....	44
	ภาคผนวก	
ก.	ผังงานการสร้างพจนานุกรมเสมือน.....	45
ข.	ผังงานการตัดคำภาษาไทย.....	46
ค.	โปรแกรมการสร้างพจนานุกรมเสมือน.....	47
ง.	โปรแกรมการตัดคำภาษาไทย.....	81
	ประวัติผู้เขียน.....	95

สารบัญตาราง

ตารางที่

หน้า

- 4.1 โครงสร้างการจัดเก็บพจนานุกรม..... 32
- 5.1 แสดงการเปรียบเทียบประสิทธิภาพการจัดเก็บพจนานุกรม ตามหลัก
การเข้าถึงในทางกว้าง และ การเข้าถึงในทางลึก..... 37



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญภาพ

ภาพที่		หน้า
2.1	แสดงโครงสร้างแบบทราย.....	5
3.1	แสดงความสัมพันธ์ระหว่าง ตำแหน่งเสมือน และ ตำแหน่งจริง.....	8
3.2	แสดงการแปลงตำแหน่งข้อมูลจากตำแหน่งเสมือนเป็นตำแหน่งในหน่วย ความจำหลัก.....	10
3.3	แสดงวิธีการของการแบ่งปล้อง.....	14
3.4	แสดงหลักการของการแบ่งเพจ.....	16
3.5	แสดงวิธีการของการแบ่งเพจ.....	18
3.6	แสดงการแทนที่โดยใช้หลักการของ LRU.....	22
3.7	แสดงการแทนที่โดยใช้หลักการของ aging.....	24
4.1	แสดงขั้นตอนการเชื่อมต่อในทางลึก.....	28
4.2	แสดงขั้นตอนการเชื่อมต่อในทางกว้าง.....	31
4.3	แสดงขั้นตอนในการสร้างพจนานุกรมเสมือน.....	35
4.4	แสดงขั้นตอนในการนำพจนานุกรมเสมือนมาใช้งาน.....	36

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย