

เครื่องมือพัฒนาพจนานุกรมอิเล็กทรอนิกส์อังกฤษ-ไทย



นางสาววันเพ็ญ ศิริจันทร์

สถาบันวิทยบริการ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2540

ISBN 974-638-282-9

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

๘ ๓.ค. 254๖

ENGLISH - THAI ELECTRONIC DICTIONARY DEVELOPMENT TOOL



Miss Wanpen Sirinirund

สถาบันวิทยบริการ

จุฬาลงกรณ์มหาวิทยาลัย

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

Department of Computer Engineering

Graduate School

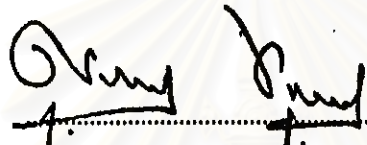
Chulalongkorn University

Academic Year 1997

ISBN 974-638-282-9

หัวข้อวิทยานิพนธ์ เครื่องมือพัฒนาพจนานุกรมอิเล็กทรอนิกส์อังกฤษ-ไทย
โดย นางสาววันเพ็ญ ศิริจันทร์
ภาควิชา วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษา รองศาสตราจารย์ ดร.วันชัย ธีรไพบูลย์


บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้วิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการ
ศึกษาตามหลักสูตรปริญญาโทบัณฑิต

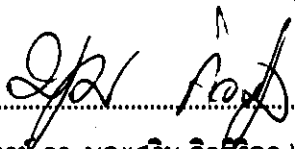

..... คณบดีบัณฑิตวิทยาลัย
(รองศาสตราจารย์ นายแพทย์ สุภวัฒน์ ชูติวงศ์)

คณะกรรมการสอบวิทยานิพนธ์


..... ประธานกรรมการ
(รองศาสตราจารย์ เตือน สินธุพันธ์ประทุม)


..... อาจารย์ที่ปรึกษา
(รองศาสตราจารย์ ดร. วันชัย ธีรไพบูลย์)


..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. สมชาย ประสิทธิ์จตุระกุล)


..... กรรมการ
(อาจารย์ ดร. นุญเสริม กิจศิริกุล)

วันเกิด ศิรินันท์ : เครื่องมือพัฒนาพจนานุกรมอิเล็กทรอนิกส์ อังกฤษ - ไทย (ENGLISH - THAI ELECTRONIC DICTIONARY DEVELOPMENT TOOL) อ. ที่ปรึกษา : รศ.ดร. วันชัย วิชาญชัย, 146 หน้า. ISBN 974-638-282-9

วัตถุประสงค์ของการวิจัยนี้ เพื่อพัฒนาโปรแกรมพจนานุกรมอิเล็กทรอนิกส์ อังกฤษ - ไทย สำหรับจัดเก็บ แก้ไข สืบค้นและแปลคำศัพท์และความหมาย โดยสามารถใช้ร่วมกับจุฬาริก 78 และ โปรแกรมประมวลผลคำอื่น

ขั้นตอนการออกแบบและพัฒนาโปรแกรม ได้ออกแบบและพัฒนาโปรแกรมเชิงวัตถุโดยวิธีการของ โคทและยอร์ดอน (P. Code and Ed. Yourdon) ส่วนการจัดเก็บและสืบค้นข้อมูลได้ใช้โครงสร้างการจัดเก็บข้อมูลแบบดัชนีคีย์ - ทรี ซึ่งในเชิงโปรแกรมจะใช้ดัชนีเปิดอะเรย์ทรี ในการจัดเก็บและสืบค้นข้อมูล

โปรแกรมพจนานุกรมอิเล็กทรอนิกส์ได้กำหนดโครงสร้างข้อมูลในภาคแสดงความหมายมาไว้และเปิดโอกาสให้ผู้ใช้สามารถกำหนดโครงสร้างข้อมูลเองได้ด้วย ซึ่งจากการทดสอบโปรแกรมพบว่าสามารถเพิ่ม แก้ไข สืบค้น และสรข้อมูลในพจนานุกรม ในการสืบค้นคำศัพท์สามารถระบุเฉพาะส่วนต้นของคำศัพท์ และจาก คำเหมือนหรือคำตรงข้ามสามารถใช้เป็นคำศัพท์เพื่อสืบค้นความหมายต่อไปได้ นอกจากนี้ยังสามารถคัดลอกข้อความจากโปรแกรมประยุกต์มาสืบค้นความหมายได้

ในส่วนการเชื่อมต่อกับจุฬาริก 78 จะแทรกเมนูในจุฬาริกและสามารถเรียกพจนานุกรมขึ้นมาทำงานได้

นอกจากนี้ เนื่องจากพจนานุกรมต้องใช้ในการจัดเก็บมาก จึงทำการบีบอัดข้อมูลด้วยวิธี แอลแตรคต์แบบถ่วงเพื่อลดขนาดของข้อมูล ซึ่งจากการทดสอบพบว่าสามารถลดพื้นที่จัดเก็บได้ร้อยละ 54

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา วิศวกรรมคอมพิวเตอร์
สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์
ปีการศึกษา 2540

ลายมือชื่อนิสิต วันชัย วิชาญชัย
ลายมือชื่ออาจารย์ที่ปรึกษา
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม

UC718713 : MAJOR COMPUTER SCIENCES
KEY WORD: ELECTRONICS / DICTIONARY / COMPRESSION

WANPEN SIRINIRUMD : ENGLISH - THAI ELECTRONIC DICTIONARY DEVELOPMENT
TOOL. THESIS ADVISOR : ASSOC. PROF. DR. WANCHAI RIVEPAIBOON, Ph.D.
145 pp. ISBN 974-638-282-9

This research has the objective to develop English-Thai electronics dictionary program to insert, modify, search and delete words and their definition. Which is can use with CU writer for windows version 78 and other word processing application.

For designing and developing process used object oriented design and programming, under P. Code and Ed. Yourdon concept. And used Digital Search Tree for data structures and manipulation. In the programming method used double array trie algorithm to manipulate an electronics dictionary.

The Electronics Dictionary has default structure of definition and allow user to define his own structure. After testing program found that it can insert, modify, search and delete data in dictionary. For searching word can be done by prefix searching, and from synonym or antonym can select word to find the meanings. More over can copy text from definition to insert into other word processing or can copy word from other word processing to search the meaning in dictionary.

Interfacing with CWW78, this system will insert dictionary menu into CWW78 and can execute when user select this menu.

Concerning to dictionary use such of disk space, so decreasing the file size had to be done by LZW for data compression. The result of compression dictionary found that it can save 54 percents of data.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา วิศวกรรมคอมพิวเตอร์
สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์
ปีการศึกษา ๒๕๓๐

ลายมือชื่อนิสิต วันเพ็ญ ภิรมย์ทวี
ลายมือชื่ออาจารย์ที่ปรึกษา
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลืออย่างดียิ่งของ รศ.ดร. วันชัย รั้วไพฑูริย์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ซึ่งท่านได้สละเวลาให้คำแนะนำ และข้อคิดเห็นต่างๆสำหรับงานวิจัยด้วยดีตลอดมา

ขอขอบพระคุณ คุณเฉลิมเกียรติ วิไลรัตนารักษ์ และคุณสมภพ ตาลสอน ที่ได้ให้ความช่วยเหลือและให้คำแนะนำต่างๆที่เป็นประโยชน์ต่อการทำวิทยานิพนธ์

ท้ายนี้ ใคร่ขอขอบพระคุณ บิดา มารดา และน้องๆที่คอยให้กำลังใจตลอดระยะเวลาในการทำงานวิจัย



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	๓
สารบัญตาราง.....	๘
สารบัญภาพ.....	๑๑

บทที่

1. บทนำ

ความเป็นมาและความสำคัญของปัญหา.....	1
วัตถุประสงค์ของวิทยานิพนธ์.....	2
ขอบเขตการวิจัย.....	2
ขั้นตอนการดำเนินการวิจัย.....	2
ประโยชน์ที่จะได้รับการจากการวิจัย.....	3

2. พจนานุกรมอิเล็กทรอนิกส์

ประวัติการพัฒนาพจนานุกรมอิเล็กทรอนิกส์.....	4
ผลงานพจนานุกรมอิเล็กทรอนิกส์ที่มีผู้ผลิตคิดค้นขึ้น.....	4
ประเภทของพจนานุกรมอิเล็กทรอนิกส์.....	9

3. โครงสร้างการจัดเก็บและวิธีการสืบค้นข้อมูล

โครงสร้างการจัดเก็บและสืบค้นข้อมูลวิธีการต่างๆ.....	13
โครงสร้างของพจนานุกรมอิเล็กทรอนิกส์ที่มีผู้พัฒนา.....	17

4. การลดขนาดข้อมูล

ประเภทของวิธีการลดขนาดข้อมูล.....	26
วิธีการลดขนาดข้อมูล.....	28

5. โครงสร้างข้อมูลแบบดี-เฮส ทรีและการลดขนาดข้อมูลโดยวิธีแฮชแชนด์ดับบลิว

โครงสร้างข้อมูลแบบดี-เฮส ทรี.....	48
อัลกอริทึมการสืบค้นข้อมูลแบบดี-เฮส ทรี.....	54
อัลกอริทึมการลดขนาดข้อมูลโดยวิธีแฮชแชนด์ดับบลิว.....	68

6. การออกแบบและพัฒนาพจนานุกรมอิเล็กทรอนิกส์	
การวิเคราะห์และออกแบบเชิงวัตถุ.....	71
ส่วนประกอบของแบบจำลองเชิงวัตถุ โดยวิธีการของโคต.....	73
การเชื่อมต่อระหว่างโปรแกรมพจนานุกรมอิเล็กทรอนิกส์กับจุฬารีก.....	88
รายละเอียดของแต่ละคลาสในพจนานุกรมอิเล็กทรอนิกส์.....	90
7. การทดสอบการทำงานของโปรแกรมพจนานุกรมอิเล็กทรอนิกส์และผลการบีบอัดข้อมูล	
การทดสอบการทำงานของโปรแกรมพจนานุกรม.....	115
ผลการบีบอัดข้อมูล.....	127
8. สรุปผลการวิจัยและข้อเสนอแนะ	
สรุปผลการวิจัย.....	128
ข้อเสนอแนะ.....	129
รายการอ้างอิง.....	130
ภาคผนวก.....	132
ประวัติผู้เขียน.....	145

สารบัญตาราง

		หน้า
ตารางที่ 4-1	แสดงการเรียงความถี่ของตัวอักษรจากมากไปน้อย.....	28
ตารางที่ 4-2	แสดงการแบ่งกลุ่มรหัสครั้งที่ 1.....	29
ตารางที่ 4-3	แสดงการแบ่งกลุ่มรหัสทุกชั้นตอน.....	29
ตารางที่ 4-4	แสดงความน่าจะเป็นของอักษรในข้อความ "BILL GATES".....	30
ตารางที่ 4-5	แสดงช่วงความน่าจะเป็นของอักษรต่างๆ.....	31
ตารางที่ 4-6	แสดงช่วงความน่าจะเป็นของอักษรต่างๆ ภายใต้ช่วง 0.20 - 0.30.....	32
ตารางที่ 4-7	แสดงการถอดรหัสวิธีเชิงคำนวณ.....	33
ตารางที่ 4-8	แสดงกลุ่มข้อมูลเมื่อนำมาเรียงลำดับตามค่าความถี่.....	34
ตารางที่ 4-9	แสดงอักษรที่ได้จากการถอดรหัส.....	36
ตารางที่ 4-10	แสดงการสร้างรหัสโดยวิธี LZ78.....	44
ตารางที่ 4-11	แสดงตารางรหัสที่สร้างโดยวิธี LZ78.....	44
ตารางที่ 4-12	แสดงการสร้างรหัสโดยวิธี LZW.....	45
ตารางที่ 4-13	แสดงการขยายข้อมูลโดยวิธี LZW.....	46
ตารางที่ 6-1	แสดงส่วนประกอบในแบบจำลองเชิงวัตถุของพจนานุกรมอิเล็กทรอนิกส์.....	74

สารบัญภาพ

		หน้า
รูปที่ 2-1	การค้นหาคำศัพท์ของเล็กซิตรอน.....	5
รูปที่ 2-2	แสดงการค้นหาคำความหมายของคำศัพท์ของ ET100 โดยการป้อนเฉพาะ ส่วนต้นของคำศัพท์.....	7
รูปที่ 2-3	แสดง ET100 สามารถเรียกใช้ได้ในขณะที่ใช้โปรแกรมประยุกต์อื่น.....	7
รูปที่ 2-4	แสดงการทำงานของ ET100 ในขณะที่ใช้โปรแกรมไมโครซอฟต์เวิร์ด.....	7
รูปที่ 2-5	แสดงการค้นหาคำศัพท์ของพจนานุกรม สอ เสถบุตร.....	9
รูปที่ 2-6	แสดงการค้นหาคำศัพท์ในรูปแบบหนังสือ.....	9
รูปที่ 3-1	โครงสร้างข้อมูลแบบเทเบิล-เทเบิล โลว์เนี่ยเสิช.....	19
รูปที่ 3-2	โครงสร้างข้อมูลแบบเทเบิล-อินเด็กซ์เสิช.....	20
รูปที่ 3-3	โครงสร้างข้อมูลพจนานุกรมแบบทรี.....	21
รูปที่ 3-4	โครงสร้างแฟ้มข้อมูลของพจนานุกรมภาษาอังกฤษ-ไทย.....	24
รูปที่ 3-5	วิธีการจัดเก็บข้อมูล.....	25
รูปที่ 4-1	การเข้ารหัสโดยวิธีสติดิต.....	26
รูปที่ 4-2	การบีบอัดข้อมูลโดยวิธีอะแดปทีฟ.....	27
รูปที่ 4-3	การขยายข้อมูลโดยวิธีอะแดปทีฟ.....	27
รูปที่ 4-4	แสดงแผนภูมิต้นไม้ฮัฟแมนขั้นตอนแรก.....	35
รูปที่ 4-5	แสดงแผนภูมิต้นไม้ฮัฟแมนที่สมบูรณ์.....	35
รูปที่ 4-6	แสดงการบีบอัดข้อมูลโดยวิธีฮัฟแมน.....	37
รูปที่ 4-7	แสดงการถอดรหัสข้อมูลโดยวิธีฮัฟแมน.....	38
รูปที่ 4-8	แสดงแผนภูมิต้นไม้ฮัฟแมน.....	39
รูปที่ 4-9	แสดงการปรับแผนภูมิต้นไม้เมื่อโหนดถูกปรับค่าความถี่ (กรณีแรก).....	40
รูปที่ 4-10	แสดงการปรับแผนภูมิต้นไม้เมื่อโหนดถูกปรับค่าความถี่ (กรณีที่สอง).....	41
รูปที่ 4-11	แสดงการเปลี่ยนแปลงรูปร่างแผนภูมิต้นไม้.....	42
รูปที่ 4-12	แสดงหน้าต่างข้อความของวิธี LZ77.....	42
รูปที่ 4-13	แสดงหน้าต่างข้อความหลังจากเข้ารหัส 14, 4, " ".....	43
รูปที่ 5-1	โครงสร้างข้อมูลแบบดับเบิลอะเรย์ทรี สำหรับ $g(s, a)$	51
รูปที่ 5-2	แสดงความสัมพันธ์ของดับเบิลอะเรย์และเทล.....	52
รูปที่ 5-3	แสดงดี-เอส ทรี ของ K.....	53

รูปที่ 5-4	แสดงดับเบิลอะเรย์ทรีของ K.....	54
รูปที่ 5-5	แสดงการทำงานของฟังก์ชัน A_Insert.....	57
รูปที่ 5-6	แสดงการทำงานของฟังก์ชัน B_Insert.....	58
รูปที่ 5-7	แสดงดับเบิลอะเรย์สำหรับฟังก์ชัน $g(s_r, b) = s_r$ และฟังก์ชัน $g(s_r, a_h) = s_r$	59
รูปที่ 5-8	แสดงดับเบิลอะเรย์ซึ่งสอดคล้องกับสมการ $BASE[k] + d = r$	59
รูปที่ 5-9	แสดงการทำงานของฟังก์ชัน Modify.....	60
รูปที่ 5-10	แสดงดับเบิลอะเรย์หลังจากเพิ่มคำศัพท์ bachelor# โดยฟังก์ชัน A_Insert(1, bachelor#).....	65
รูปที่ 5-11	แสดงดับเบิลอะเรย์หลังจากเพิ่มคำศัพท์ bcs# โดยฟังก์ชัน B_Insert(3, E, cs#, achelor#).....	65
รูปที่ 5-12	แสดงดับเบิลอะเรย์หลังจากเพิ่มคำศัพท์ badge# โดยฟังก์ชัน B_Insert(2, E, dge#, chelor#).....	65
รูปที่ 5-13	แสดงดับเบิลอะเรย์หลังจากเพิ่มคำศัพท์ baby# โดยฟังก์ชัน A_Insert(2, by#).....	66
รูปที่ 5-14	แสดงดี-เอส ทรี ของ bachelore#.....	66
รูปที่ 5-15	แสดงดี-เอส ทรี ของ bachelor# และ bcs#.....	66
รูปที่ 5-16	แสดงดี-เอส ทรี ของ bachelor# ,bcs# และ badge#.....	66
รูปที่ 5-17	แสดงดี-เอส ทรี ของ bachelor# ,bcs# ,badge# และ baby#.....	67
รูปที่ 6-1	โครงสร้างการจัดเก็บข้อมูลของพจนานุกรมอิเล็กทรอนิกส์.....	70
รูปที่ 6-2	แบบจำลองเชิงวัตถุของพจนานุกรมอิเล็กทรอนิกส์(PD).....	78
รูปที่ 6-3	การบันทึกคำศัพท์และความหมาย(HI).....	79
รูปที่ 6-4	การสืบค้นความหมาย(HI).....	80
รูปที่ 6-5	การจัดการเพิ่มข้อมูลพจนานุกรม(DM).....	80
รูปที่ 6-6	การเชื่อมต่อพจนานุกรมกับจุฬารีก 78 (SI).....	81
รูปที่ 6-7	แสดงการเพิ่มคำศัพท์ลงในดีเอส-ทรี.....	82
รูปที่ 6-8	แสดงการสืบค้นคำศัพท์และความหมาย.....	83
รูปที่ 6-9	การสร้างจอภาพสำหรับการบันทึกคำศัพท์และความหมาย กรณีผู้ใช้กำหนดโครงสร้างเอง.....	84
รูปที่ 6-10	แสดงการเพิ่มคำศัพท์ลงในดีเอส-ทรี กรณีผู้ใช้กำหนดโครงสร้าง ในภาคแสดงความหมายเอง.....	85

รูปที่ 6-11	แสดงการสืบค้นคำศัพท์และความหมายกรณีผู้ใช้กำหนดโครงสร้าง ในภาคแสดงความหมายเอง.....	86
รูปที่ 6-12	แสดงการเชื่อมต่อระหว่างพจนานุกรมกับจุฬารีก.....	87
รูปที่ 7-1	การบันทึกคำศัพท์และความหมายลงในแฟ้มข้อมูล และการเพิ่มคำศัพท์ ลงในดีเอส-ทรี.....	115
รูปที่ 7-2	การสืบค้นความหมายของคำศัพท์ที่ต้องการ.....	116
รูปที่ 7-3	การเลือกคำว่า "desert" ในส่วนของคำเหมือน เพื่อสืบค้นหาความหมายต่อไป	116
รูปที่ 7-4	การแสดงความหมายของคำศัพท์ "desert" ซึ่งเป็นผลจากรูปที่ 7-3.....	117
รูปที่ 7-5	แสดงการสืบค้นโดยใช้ส่วนต้นของคำศัพท์.....	117
รูปที่ 7-6	แสดงผลของการเลือกคำว่า "absolutely" จากรูปที่ 8-5.....	118
รูปที่ 7-7	แสดงการคัดลอกคำว่า "เคลื่อนไหว" เพื่อนำไปแทรกในข้อความของจุฬารีก	118
รูปที่ 7-8	แสดงผลจากรูปที่ 7-7 ซึ่งเป็นการคัดลอกคำว่า "เคลื่อนไหว" มาแทรกใน จุฬารีก.....	119
รูปที่ 7-9	การเลือกคำว่า "ability" เพื่อใช้เป็นคำหลักในการสืบค้นความหมาย ในพจนานุกรม.....	119
รูปที่ 7-10	แสดงผลของการสืบค้นความหมายของคำว่า "ability" จากรูปที่ 7-9.....	120
รูปที่ 7-11	แสดงการแก้ไขข้อความในส่วนของภาคแสดงความหมาย.....	120
รูปที่ 7-12	การลบคำว่า "desert" ออกจากพจนานุกรม.....	121
รูปที่ 7-13	หลังจากที่ลบคำว่า "desert" และทำการสืบค้นใหม่จะไม่พบคำศัพท์นี้ ในพจนานุกรม.....	121
รูปที่ 7-14	แสดงการเลือกชนิดข้อมูลที่ต้องการให้มีในภาคแสดงความหมาย.....	122
รูปที่ 7-15	แสดงช่องสำหรับบันทึกข้อมูลตามชนิดที่เลือกได้.....	123
รูปที่ 7-16	แสดงผลของการสืบค้นคำว่า "one"	124
รูปที่ 7-17	แสดงการสืบค้นคำศัพท์ที่ขึ้นต้นด้วยอักษร "o".....	125
รูปที่ 7-18	แสดงการสืบค้นคำว่า "one" ซึ่งสืบเนื่องจากรูปที่ 7-17.....	125
รูปที่ 7-19	แสดงการแทรกเมนูพจนานุกรมลงในจุฬารีก.....	126
รูปที่ 7-20	แสดงการทำงานของพจนานุกรม เมื่อผู้ใช้งานเลือกเมนูพจนานุกรม ในจุฬารีก.....	127