เอกสารอ้างอิง

1. ยืน ภู่วรวรรณ, ชัยยงค์ วงศ์ชัยสุวัฒน์, "เครื่องมือพื้นฐานสำหรับงานประมวลผลภาษาธรรมชาติ (2532)", เอกสารประกอบการสัมมนาทางวิชาการเรื่อง Natural Language Processing, สมาคมส่งเสริมเทคโนโลยี(ไทย-ญี่ปุ่น)

2. นิตยา กาญจนวรรณ, ภาษาศาสตร์คอมพิวเตอร์ 1, กิ่งจันทร์การพิมพ์, กรุงเทพมหานคร, 2529.

3. มัณฑนา ปราการสมุทร, โครงสร้างข้อมูล, เอกสารประกอบการสอนวิชา Data Structure

4. พิชัย บูรณสมบัติ, โครงสร้างและระบบข้อมูล, สำนักพิมพ์กราฟิคอาร์ต, กรุงเทพมหานคร, 2527.

5. A. V. Aho, R. Sethi, and J.D. Ullman, Compilers Principles, Techniques, and Tools., Reading MA: Addison-Wesley, 1986, ch. 2.

6. A. V. Aho, J. E. Hopcroft, and J. D. Ullman, Data Structures and Algorithms., Reading, MA: Addison-Wesley, 1983.

7. D. E. Knuth, The Art of Computer Programming, Vol I, Fundamental Algorithms, vol III, Sorting and Searching., pp. 295-304, 481-505, Reading, MA: Addison-Wesley, 1973.

8. Y. Poowarawan and C. Wongchaisurawat (1989), "Data Compression Techniques for Electronic Dictionary", Proceeding of The Regional Workshop on Computer Processing of Asian Language (CPAL), Asia Institute of Technology, (Thailand)

9. W. D. Jonge, A. S. Tanenbaum, and R. P. Reit, "Two access methods using compact binary trees", IEEE Trans. Software Eng., Vol. SE-13, pp. 799-810 July 1987.

10. Gonnet, G.H. (Gaston H.), Handbook of Algorithm and Data Structures, pp. 103-109, Addison-Wesley, 1984.

11. Thomas A. Standish, University of California at Irvine, Data Structure Techniques, pp.122-123, Addison-Wesley, 1980.

12. J. Aoe, "An Efficient Digital Search Algorithm by Using a Double-Array_ Structure", IEEE Trans. Software Eng., Vol. 15, pp. 1066-1077, Sept. 1989.

13. J. Aoe, "An efficient implementation of static string pattern matching machines" , IEEE Trans. Software Eng., Vol. 15, pp. 1010-1016, Aug. 1989.

14. วิจินตน์ ภาณุพงศ์, โครงสร้างของภาษาไทย, โรงพิมพ์สำนักพิมพ์ มหาวิทยาลัยรามคำแหง, 2528.

15. Brian W. Kernigham, Dannis M. Ritchie., The C Programming Language, Prentice-Hall of India Private Limited, 1988.

16. Kohn Thomas Berry, The Waite Group, Advanced C Programming, Prentice Hall Press, 1986.

17. Al Stevens, C Data Base Deveplopment, Management Information Source, Inc. . 1987.

18. พรพจน์ สุริยะวงศ์ และ ยืน ภู่วรวรรณ, "dicttool : โปรแกรมสำหรับตัดคำและแก้ไขพจนานุกรมคำโดด", บทความภายในห้องปฏิบัติการวิจัยไมโครคอมพิวเตอร์ ม.เกษตร R-99-01, กุมภาพันธ์, 2534.

ภาคผนวก ก.

ตัวอักษร	ค่าที่ใช้	ตัวอักษร	ค่าที่ใช้
ก	1	ข	2
ค	4	ฆ	6
ง	7	จ	8
ฉ	9	ช	10
ซ	11	ฌ	12
ญ	13	ฎ	14
ฏ	15	ฐ	16
ฑ	17	ฒ	18
ณ	19	ด	20
ต	21	ถ	22
ท	23	ธ	24
น	25	บ	26
ป	27	ผ	28
ฝ	29	พ	30
ฟ	31	ภ	32
ม	33	ย	34
ร	35	ล	37
ว	39	ศ	40
ษ	41	ส	42
ห	43	ฬ	44
อ	45	ฮ	46
ฤ	36	ฦ	38

ตารางที่ 1 ตารางแสดงค่าตัวอักษรที่ใช้ในโครงสร้างข้อมูลสำหรับเก็บพจนานุกรมอิเล็กทรอนิกส์ภาษาไทย

ตัวอักษร	ค่าที่ใช้	ตัวอักษร	ค่าที่ใช้
ะ	48	า	50
ิ	52	ี	53
ึ	54	ื	55
ุ	56	ู	57
เ	64	แ	65
โ	66	ำ	51
่	72	้	73
๊	74	๋	75
็	71	ๆ	70
์	79		

ตารางที่ 1 (ต่อ) ตารางแสดงค่าตัวอักษรที่ใช้ในโครงสร้างข้อมูลสำหรับเก็บพจนานุกรมอิเล็กทรอนิกส์ภาษาไทย

ภาคผนวก ข.

ตัวอย่างคำศัพท์

ฟก, ฟ, ฟอก, ฟอด, ฟอน, ฟัก, ฟัง, ฟัด, ฟัน, ฟาก, ฟาง, ฟาด, ฟิน, ฟุต, ฟุบ, ฟูก, ฟูด, ฟูม, ฟ้อ, ฟ้า,
ฟั่น, ฟื้น, ฟุ้ง, ฟอง, ฟอน, ฟาง, ฟาม, ฟ้อง, ฟ้อน, ฟกช้ำ, ฟันผุ, ฟิล์ม, ฟอกชัก, ฟองน้ำ,
ฟองมัน, ฟักทอง, ฟักฟูม, ฟันดาบ, ฟันปลา, ฟันฝา, ฟันฟอน, ฟันรวน, ฟันโยก, ฟันโหล, ฟากผ้า,
ฟากฟัน, ฟื้นฟู, ฟุตบอล, ฟูมฟัก, ฟูมฟาย, ฟ้อนรำ, ฟ้าผ่า, ฟ้าสาง, ฟ้าแลบ, ฟอกหนัง, ฟองสบู่,
ฟักข้าว, ฟันกราม, ฟันปลอม, ฟื้นดิน, ฟื้นตัว, ฟื้นฝอย, ฟ้ามุ่ย, ฟ้าร้อง, ฟ้าลั่น,
ฟังเทศน์, ฟันน้ำนม, ฟันเฟือง, ฟั่นเฝือ, ฟายน้ำตา, ฟุ้งซ่าน, ฟุ้งเฟ้อ, ฟ้องร้อง,
ฟ้องหย่า, ฟ้องแย้ง, ฟ้าคะนอง, ฟันเลื่อย, ฟั่นเชือก, ฟั่นเทียน, ฟั่นเฟือน, ฟุ้มเฟือย,
ฟ้องหาว่า, ฟ้องอนาถา, ฟากข้างนี้, ฟุ้งเฟื่อง, ฟ้าร้องอิม, ฟอกเสื้อผ้า, ฟากข้างโน้น,
ฟื้นจากหลับ, ฟ้องร้องอุธรณ์, ฟังไม่รู้เรื่อง,
ภพ, ภมร, ภัย, ภาค, ภาพ, ภาย, ภาร, ภูต, ภู่,
ภาคี, ภาระ, ภาวะ, ภาษา, ภาษี, ภูผา, ภูมิ, ภูษา, ภรรยา, ภริยา, ภักดี, ภัณฑ์, ภาชนะ, ภายใน,
ภาวนา, ภาษิต, ภิกษุ, ภูเขา, ภวังค์, ภาพวิว, ภายนอก, ภายใต้, ภารโรง, ภาคผนวก, ภาคภูมิ,
ภาคหลวง, ภาพถ่าย, ภาพพจน์, ภาพร่าง, ภาพหมู่, ภายหน้า, ภายหลัง, ภารธุระ, ภาษามือ,
ภาษาใบ้, ภิกษุณี, ภูมิฐาน, ภูมิภาค, ภูมิภาพ, ภูมิรู้, ภูเขาไฟ, ภยันตราย, ภราดรภาพ, ภาคทัณฑ์,
ภาคเรียน, ภาพยนตร์, ภาพลวงตา, ภาพสีน้ำ, ภาษาเด็ก, ภาษีมรดก, ภาษีอากร, ภิกขาจาร,
ภาษีเสริม, ภูมิลำเนา, ภูมิวิทยา, ภัยอุบาทว์, ภาพนิมพ์สี, ภาษีการค้า, ภูมิประเทศ,
ภูมิศาสตร์, ภาพสีน้ำมัน, ภาษีเงินได้, ภาษาท้องถิ่น, ภาษาปัจจุบัน, ภาษีโภคภัณฑ์,
ภูเขาน้ำแข็ง, ภาษาพื้นเมือง, ภาษาต่างประเทศ, ภาษีสินค้าขาเข้า,
ฤชา, ฤดู, ฤษี, ฤกษ์, ฤษยา, ฤาษี, ฤชากร, ฤดูฝน, ฤทธิ์, ฤดูกาล, ฤดูร้อน, ฤดูหนาว, ฤดูแล้ง,
ฤกษ์เช้า, ฤดูใบไม้ผลิ, ฤดูใบไม้ร่วง, ฤดูประจำเดือน, ฤดูเกี่ยวข้าว,
ศก, ศพ, ศร, ศอ, ศอก, ศาล, ศีล, ศึก, ศุภ, ศารท, ศาลา, ศิขร, ศิลป, ศิลา, ศรีษะ, ศัตรู, ศัพท์,
ศานติ, ศาสดา, ศาสนา, ศิลปี, ศิษย์, ศึกษา, ศุกร์, ศุทธิ, ศูนย์, ศตวรรษ, ศรัทธา, ศฤงคาร,
ศักดิ์, ศักราช, ศัสตรา, ศาสตร์, ศิลปกร, ศิลปิน, ศีลขาด, ศีลสิบ, ศีลห้า, ศพอาบยา, ศักดินา,
ศาลฎีกา, ศาลทหาร, ศาลอาญา, ศาลาทาน, ศาลเจ้า, ศาลแขวง, ศาลแพ่ง, ศิลายอน,
ศิลาลาย, ศิลาแดง, ศิลาแลง, ศีลธรรม, ศุลกากร, ศัพท์แสง, ศัลยกรรม, ศาลอาชญา,
ศาลากลาง, ศาสโนวาท, ศิลปกรรม, ศิษย์โยม, ศุภนิมิต, ศุลกสถาน, ศาสนาพุทธ, ศิลปศึกษา,
ศีลสมาทาน, ศูนย์กลาง, ศัลยศาสตร์, ศาลพระภูมิ, ศาลอุทธรณ์, ศาลโปรีสภา, ศิลปศาสตร์,
ศาลยุติธรรม, ศาลอุทรธรณ์, ศาสตราจารย์, ศาสนาคริสต์, ศาสนูปถัมภก, ศาสนาของชาติ,
ศึกษานิเทศก์, ศาลาการเปรียญ, ศาสนามะหะหมัด, ศึกเสือเหนือใต้,

ภาคผนวก ค.

รูปแบบโครงสร้างข้อมูลที่นำมาใช้ทดสอบประสิทธิภาพการทำงาน

1. ลักษณะโครงสร้างข้อมูล

โครงสร้างข้อมูลที่นำมาใช้สำหรับทดสอบเพื่อเปรียบเทียบประสิทธิภาพการทำงานเป็นโครงสร้างข้อมูลที่ประกอบด้วย 2 ส่วนดังนี้ คือ ส่วนแรกเก็บตำแหน่งที่อยู่ของคำศัพท์ และส่วนที่เก็บคำศัพท์ สำหรับรายละเอียดของเนื้อหาที่มีในโครงสร้างทั้ง 2 ส่วน มีดังนี้คือ

1.1 ส่วนที่เก็บตำแหน่งที่อยู่ของคำศัพท์

โครงสร้างข้อมูลส่วนนี้จะถูกเก็บไว้ในอะเรย์ โดยที่ใช้ตัวพยัญชนะไทยมาเป็นสิ่งที่กำหนดค่าดรรชนีของอะเรย์ เช่น สารสนเทศของตัวพยัญชนะ "ก" จะถูกเก็บในอะเรย์ที่ 1 เป็นต้น สำหรับรายละเอียดสารสนเทศของส่วนนี้ประกอบด้วย ตำแหน่งเริ่มต้นที่เก็บชุดของคำศัพท์ที่ขึ้นต้นด้วยพยัญชนะที่สอดคล้องกับค่าดรรชนีของอะเรย์นั้นๆ พร้อมกับจำนวนของคำศัพท์ที่ขึ้นต้นด้วยพยัญชนะดังกล่าว จำแนกตามความยาวของคำศัพท์ เช่น คำศัพท์ที่ขึ้นต้นด้วยตัว "ก" ความยาว 2 ตัวมี 50 คำ คำศัพท์ที่ขึ้นต้นด้วยด้วย "ก" ความยาว 3 มี 80 คำ เป็นต้น โดยที่รายละเอียดจำนวนคำศัพท์นี้ จะถูกเก็บในอะเรย์เช่นกัน และดรรชนีที่ใช้ค้นหานั้นใช้ความยาวเป็นสิ่งกำหนดค่าของดรรชนีนั่นเอง

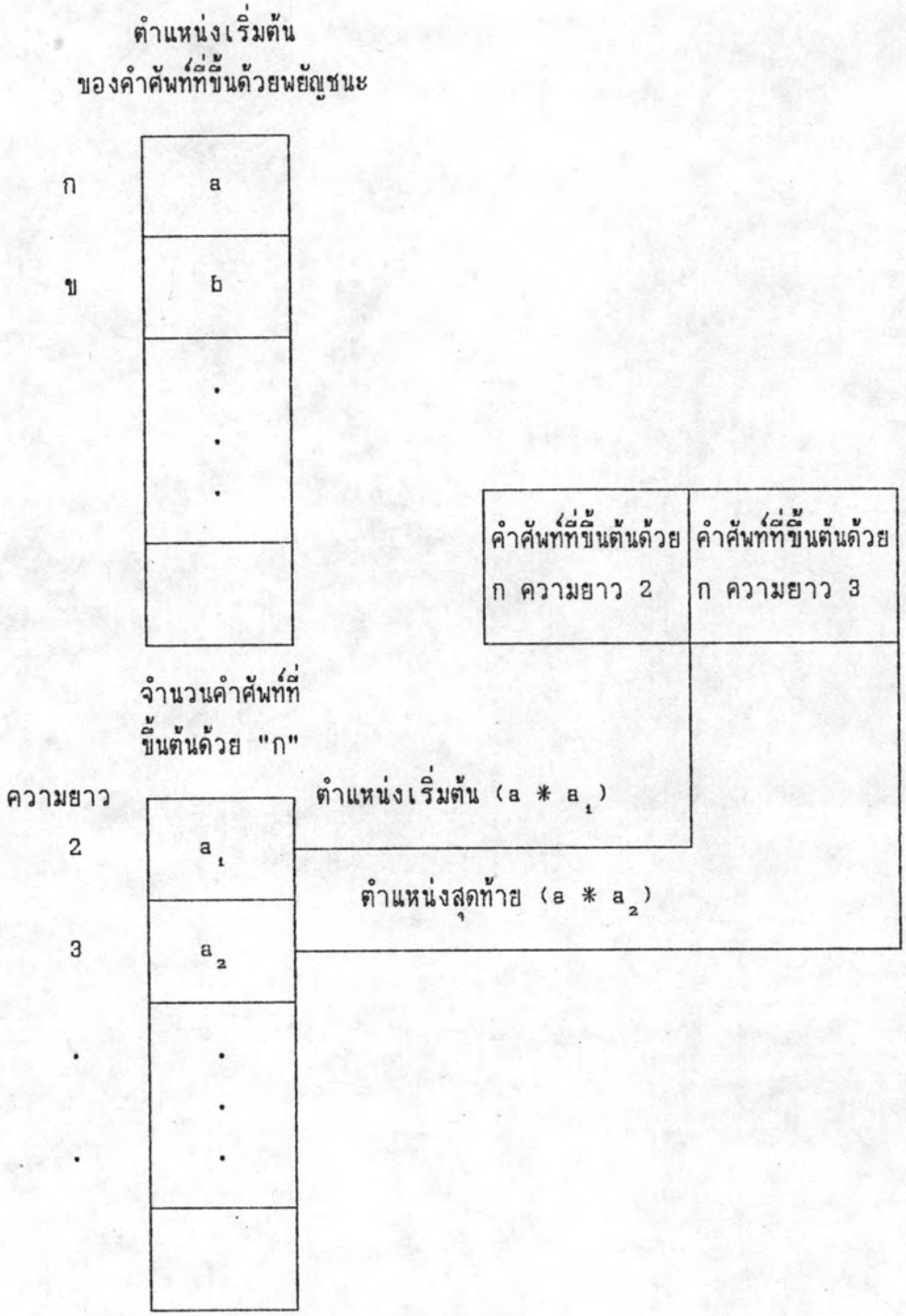
1.2 ส่วนที่เก็บคำศัพท์

ในส่วนนี้จะเก็บคำศัพท์โดยแบ่งออกเป็นกลุ่มๆ โดยใช้ตัวพยัญชนะที่ขึ้นต้นและความยาวมาช่วยในการจัดกลุ่มแบ่งคำศัพท์ เช่น กลุ่มคำศัพท์ที่ขึ้นต้นด้วยพยัญชนะ "ก" ความยาว 2 จัดเป็นกลุ่มหนึ่ง และ กลุ่มคำศัพท์ที่ขึ้นต้นด้วยพยัญชนะ "ก" ความยาว 3 จัดเป็นอีกกลุ่มหนึ่ง ซึ่งตำแหน่งที่เก็บจะคำนวณมาจากส่วนที่เก็บตำแหน่ง และรายละเอียดการสืบค้นคำศัพท์จากโครงสร้างนี้จะกล่าวถึงในหัวข้อต่อไป

2. ขั้นตอนวิธีการสืบค้นคำศัพท์

วิธีการสืบค้นคำศัพท์กระทำโดย พิจารณาว่าคำศัพท์ที่ต้องการสืบค้นขึ้นต้นด้วยพยัญชนะตัวใด และความยาวเท่าใด แล้วนำเอาพยัญชนะที่ขึ้นต้นนั้นไปหาตำแหน่งในอะเรย์ชุดแรก แล้วจากนั้นจึงนำเอาความยาวของคำศัพท์ไปเป็นเกณฑ์สำหรับหาจำนวนคำศัพท์อีกครั้งหนึ่ง เพื่อที่จะนำมาคำนวณหาตำแหน่งเริ่มต้นและตำแหน่งสุดท้ายของคำศัพท์

ชุดนี้เมื่อทราบตำแหน่งแล้วจึงทำการสืบค้นต่อไปโดยใช้การค้นแบบทวิภาค (binary search)



รูปที่ ค.1 รูปแบบโครงสร้างข้อมูลที่นำมาทดสอบประสิทธิภาพ

หมายเหตุ * หมายถึง โอเปอร์เรชัน (OPERATION) ที่ใช้คำนวณตำแหน่งที่เก็บคำศัพท์

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวสมปรารถนา รัทยานนท์ เกิดเมื่อวันที่ 6 พฤษภาคม ที่ กรุงเทพมหานคร สำเร็จการศึกษาระดับปริญญาตรีวิทยาศาสตรบัณฑิต สาขาคณิตศาสตร์ ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ และเข้าศึกษาต่อใน หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิทยาศาสตร์คอมพิวเตอร์ ที่จุฬาลงกรณ์มหาวิทยาลัย เมื่อ พ.ศ. 2531 ปัจจุบันทำงานตำแหน่งผู้วิเคราะห์ ฝ่ายวางแผนและปฏิบัติการคอมพิวเตอร์ ธนาคารแห่งประเทศไทย