

รายการอ้างอิง

- [1] S. Charnyapompong, "A Thai Syllable Separation Algorithm," M.Eng. Thesis, Asian Institute of Technology, Aug. 1983.
- [2] C. Faloutsos and D.W. Oard, "A survey of Information Retrieval and Filtering Methods," Technical Report CS-TR-3514, University of Maryland, College Park, August 1995.
- [3] W. B. Frakes and R. Baeza-Yates eds., Information Retrieval : Data Structures and Algorithms, Englewood Cliffs, N.J. : Prentice-Hall.
- [4] G. Gonnet, "Unstructured Data Bases or Very Efficient Text Searching," ACM PODS, vol. 2, pp. 117-124, 1983
- [5] G. Gonnet, R. Baeza-Yates, and T. Snider "New Indices for Text: PAT Trees and PAT Arrays," in Information Retrieval : Data Structures and Algorithms, ed., W. B. Frakes and R. Baeza-Yates, Englewood Cliffs, N.J. : Prentice-Hall
- [6] P. Jindavimonlert, "A Thai Text Retrieval System using the PAT tree," M.Sc. Thesis, Department of Computer Engineering Chulalongkorn University, 1996.
- [7] A. Kawtrakul, C. Thumkanon, and S. Seriburi, "A Statistical Approach to Thai Word Filtering," Proc. Of the second Symposium on Natural Language Processing, pp. 398-406, 1995.
- [8] U. Manber and G. Myers, "Suffix Arrays: A New Method for On-line String Searches," First ACM-SIAM Symp. On Discrete Algorithms, pp. 319-327, San Francisco, 1990.

- [9] T.H. Merrett and H. Shang, "Trie Methods for Representing Text," *Proc Fourth Int'l Conf., FODO'93*, LNCS 730, pp. 130-145, Chicago: Springer-Verlag, Oct. 1993.
- [10] G. Salton and M.J. McGill, *Introduction to Modern Information Retrieval*. McGraw-Hill, 1983.
- [11] H. Shang, "Trie Methods for Text and Spatical Data on Secondary Storage," Ph.D. Dissertation, School of Computer Science, McGill University, Nov. 1994.
- [12] H. Shang and T.H. Merrett, "Tries for Approximate String Matching," *IEEE Trans. on Knowledge and Data Eng.*, Vol. 8, No. 4, Aug. 1996, pp. 540-547.
- [13] D. Sintupunpratum and C. Bandhitanont, "การประมวลผลคำไทย (Thai Word Processing) - in Thai," Proc. Of the second Symposium on Natural Language Processing in Thailand, pp. 322-376, March 1993.
- [14] I.H. Witten, A. Moffat, and T.C. Bell, *Managing Gigabytes : Compressing and Indexing Documents and Images*, N.Y., Van Nostrand Reinhold.
- [15] M.J. Folk, B. Zoellick, *File Structures*, Addison-Wesley Publishing Company, Inc.
- [16] Witoon Kanlayanawat and Somchai Prasitjutrakul, Automatic Indexing In Thai Text Using Trie Structure, NLPRS'97
- [17] ดวงแก้ว สวามิภักดิ์, "การสร้างซอฟต์แวร์วิเคราะห์ไวยากรณ์ไทยภายใต้ระบบยูนิกซ์." สถาบันคตศึกษา ธรรมศาสตร์, 2533.
- [18] ยืน ภู่วรรณ, วิวรรณ อิมอรณ, "การแบ่งพยางค์ไทยด้วยดิक्ชันนารี," รายงานการประชุมวิชาการทางไฟฟ้า ครั้งที่ 9, 2529.

- [19] วิรัช ศรีเลิศล้ำวานิชย์, “การตัดคำไทยในระบบแปลภาษา.” การแปลภาษาด้วยคอมพิวเตอร์, กรุงเทพมหานคร : บริษัทซีเอ็ดยูเคชั่น, 2536, หน้า 50-55.
- [20] ประกาศิต ชาติบุรุษ, อาทิตย์ จิตต์จุฬานนท์, “โครงสร้างข้อมูล และอัลกอริทึม”. กรุงเทพมหานคร : บริษัทซีเอ็ดยูเคชั่น จำกัด.
- [21] กำชัย ทองหล่อ, “หลักภาษาไทย”, กรุงเทพมหานคร : บริษัทรวมสาส์น (1977) จำกัด



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ก.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

การแบ่งคำภาษาไทยในโปรแกรมซียูไรท์เตอร์

ลักษณะของตัวอักษรภาษาไทย

จากการวิเคราะห์สถิติการใช้ตัวอักษรภาษาไทย และหลักทางภาษาศาสตร์แล้วสามารถแบ่งตัวอักษรได้เป็น 5 กลุ่มย่อยๆ ดังนี้

1. พยัญชนะ ปัจจุบันมีใช้อยู่ 42 ตัว (ไม่นับ ช ค) สามารถแบ่งพยัญชนะออกเป็น 5 กลุ่มย่อยๆ ดังนี้

1.1 พยัญชนะที่จะเป็นพยัญชนะต้นเสมอ ได้แก่

จ ผ ฝ ฮ

1.2 พยัญชนะที่ปกติจะเป็นพยัญชนะต้น ได้แก่

ห ภ ข ฟ ช ท

1.3 พยัญชนะที่เป็นได้ทั้งพยัญชนะและสระ ได้แก่

อ ว ร (ร ใช้ในรูปรร)

1.4 พยัญชนะที่ปกติจะเป็นตัวสะกด ได้แก่

ศ ณ ญ ษ ร ฎ ฏ ฒ พ ฌ

1.5 พยัญชนะที่เป็นได้ทั้งตัวสะกดและพยัญชนะต้น ได้แก่

ก ข ค ง จ ช ด ต ถ ท ธ

น บ ป พ ม ย ล ส

2. สระ ที่ใช้อยู่ในปัจจุบันมี 17 ตัว ซึ่งสามารถแบ่งออกเป็นกลุ่มย่อย ได้ 5 กลุ่ม ดังนี้

2.1 สระ ที่ปกติจะเป็นตัวอักษรแรกของคำ ได้แก่

แ ไ โ

2.2 สระ ที่ปกติจะเป็นตัวอักษรตัวสุดท้ายของคำ ได้แก่
ะ ำ

2.3 สระ ที่ปกติจะต้องการตัวสะกด ได้แก่
ุ ู ึ ู

2.4 สระ ที่มี หรือไม่มีตัวสะกดก็ได้ ได้แก่
า ิ ี ึ ุ

2.5 สระพิเศษ ที่ใช้เฉพาะในคำบางคำ ได้แก่
ฤ

3. วรรณยุกต์ มี 4 ตัว ได้แก่ ˊ ˋ ˊ ˋ

4. สัญลักษณ์พิเศษ มี 21 ตัว ได้แก่

ฯ ๗ () [] - { } " !
: ; | \$ _ . ? / blank

5. ตัวเลข มี 10 ตัว ได้แก่ ตัวเลข 0-9

โครงสร้างโดยทั่วไปของคำในภาษาไทย

ตามหลักไวยากรณ์ของภาษาไทย สามารถแบ่งรูปแบบของคำได้ 7 รูปแบบดังนี้

[<๑>]

1) <พ>[<พ>][<ต>[<ต>]][<ก>] : จะ ถลา ถ้ำ ถ้ำ บาน มฤต กานต์

[<๑>]

<ส>

2) <พ><พ><ต><ต><ก> : กิน หมั่น ที่ ฉันท คลี มือ

<ว>

3) <พ><พ><ต><ต><ก> : จ อุ คุ่ม บรม

<ส>

<ว>

4) <ส><พ><พ><ต><ต><ก> : โพร แม่ โยม โทธิ์

<ว>

<ส>

5) <ส><พ><พ><ต><ต><ก> : เณ เพลิน เป็น เป็น เซ็นต์

<ว>

<ส>

6) <ส><พ><พ><ว><ต><ต><ก> : เกือบ เกลียด เพลี้ย

<ว>

7) <ส><พ><พ><ว><ก> : เสาร์ เกล้า เช้า เกาะ เธอ

การสร้างกฎเกณฑ์ในการแบ่งคำ

จากการวิเคราะห์รูปแบบของคำ จึงได้สร้างกฎเกณฑ์ของการแบ่งคำเป็น ข้อๆ โดยยึดหลักการทางภาษาศาสตร์และข้อมูลทางสถิติ แต่เนื่องจากภาษาไทยประกอบด้วยคำที่มีรูปแบบแตกต่างกันมากมาย ดังนั้นกฎที่ใช้ทุกกฎจึงต้องมีค่ายกเว้นของกฎนั้นๆ

กฎเกณฑ์การแบ่งคำไทยที่สร้างขึ้นนี้ ได้รวบรวมเป็นหมวดหมู่ไว้แล้ว โดยจะยึดหลักการให้เป็นกฎที่มีความแน่นอน เพื่อจะใช้กับคอมพิวเตอร์ได้ โดยเฉพาะจะมีการกำหนดข้อยกเว้นต่างๆ ไว้ด้วย เพื่อให้มีความสมบูรณ์ของกฎเกณฑ์ ซึ่งจะมีหลักการดังนี้คือ

กฎข้อที่ 1 เครื่องหมายพิเศษ (Special Characer) สามารถใช้แบ่งคำได้ โดยจะแบ่งเครื่องหมายพิเศษนี้ ออกเป็น 2 กลุ่ม ดังนี้ คือ

1. ประเภทที่เป็นวงเล็บเปิดและเครื่องหมายพิเศษบางตัวจะแบ่งคำหน้าตัวอักษรเหล่านั้น ได้แก่ ([{ | /
 เช่น คำว่า "คำนาม (เอกพจน์) เป็นนามนับได้" จะตัดคำได้เป็น "คำนาม" และ "(เอกพจน์) เป็นนามนับได้"
2. ประเภทที่เป็นวงเล็บปิดและเครื่องหมายพิเศษอื่นๆ จะแบ่งคำตรงตำแหน่งตัวอักษรนั้นๆ ได้แก่)] } ! \$ % - : ; ?
 เช่น คำว่า "(บรรทัด) ตรงกัน" จะตัดคำเป็น "(บรรทัด)" และ "ตรงกัน"

กฎข้อที่ 2 ตัวกรันต์ (ˆ) มักจะใช้เป็นตัวสุดท้ายของคำ เช่น ศิลป์ สันต์ เป็นต้น แต่ก็มีคำยกเว้นอยู่หลายคำ โดยมากมักจะเป็นคำที่มาจากภาษาอังกฤษ เช่น บอร์ด फिल्म เป็นต้น ซึ่งคำเหล่านี้ มักจะนำหน้ากรันต์ด้วยตัว ร และ ล เสมอ

กฎข้อที่ 3 ตัวอักษรสระอะ (ะ) และสระอำ (ำ) มักจะใช้เป็นตัวสุดท้ายของคำซึ่งก็มีข้อยกเว้น กรณีสระอะมีตัวอักษร ्ह ลงท้าย เช่น เควาะห์ เป็นต้น หรือกรณีสระอำ มีวรรณยุกต์ตามมวารรณยุกต์ก็จะเป็นตัวสุดท้ายแทน เช่น ช้ำ เป็นต้น

กฎข้อที่ 4 ตัวสระไม้ม้วน (เ) จะใช้นำหน้าพยัญชนะเสมอ จึงจะเป็นตัวอักษรแรกสุดของคำ (คำในภาษาไทยที่ใช้ เ มี 20 คำ เท่านั้น)

กฎข้อที่ 5 ตัวสระไม้หันอากาศ (ั) สระอิ (ิ) สระอี (ึ) และสระอิ (ึ) โดยปกติมักจะต้องการตัวสะกด 1 ตัว เช่น กิน ตัด ยึด แต่มีคำยกเว้นหลายคำ เช่น นัยน์ เกิน เรื่อง เป็นต้น

กฎข้อที่ 6 ตัวสระเอ (เ) สระแอ (แ) สระโอ (โ) และสระไม้มลาย (ไ) ปกติจะใช้นำหน้าพยัญชนะ แต่ก็จะมีคำยกเว้นอยู่หลายคำ เช่น มเหสี สแลง อโหสิ สไป เป็นต้น

กฎข้อที่ 7 ตัวพยัญชนะ จ ผ ฝ ฮ จะใช้เป็นพยัญชนะต้นนำหน้าเสมอ เช่น จกรรจ ฝไท ฝรัง ฮ้อ เป็นต้น แต่อาจจะมีตัวสระในกฎข้อที่ 4, 6 นำหน้าได้ เช่น เจลียง ผย ไซ้ เป็นต้น

กฎข้อที่ 8 ตัวสระอู (,) และสระอุ (,) มักจะใช้ไว้ในได้พยัญชนะตัวแรก หรือ พยัญชนะตัวที่สองของคำ เช่น คุณ กุล ปลุก สุก ชูด มุก กรุด อดู เป็นต้น แต่อาจจะมี คำที่สระไปอยู่ได้พยัญชนะตัวอื่นๆ (ไม่ใช่พยัญชนะตัวแรกหรือตัวที่สอง) เช่น เหตุ ธาตุ เรณู เมณู ไศสุรย์ เป็นต้น

กฎข้อที่ 9 ตัวสระอา (า) โดยปกติจะต้องมีพยัญชนะนำหน้าอย่างน้อย 1 ตัวเสมอ โดยจะพิจารณาเป็น 2 กรณี คือกรณีที่ในคำไม่มีการใช้วรรณยุกต์เลย เช่น ปากกา นาน สบาย สตางค์ และกรณีที่ในคำมีการใช้วรรณยุกต์ร่วมอยู่ด้วย เช่น กร๊าฟ ฝ่าย หม้าย เป็นต้น โดยในแต่ละกรณีจะแบ่งคำหน้าหรือหลังสระก็ได้ แล้วแต่ความเหมาะสมของรูปแบบของคำนั้นๆ

กฎข้อที่ 10 ตัวอักษร อ ที่ใช้ร่วมกับวรรณยุกต์ จะพิจารณาจากตัวอักษรที่อยู่หน้าและ หลังตัวอักษร อ โดยจะดูว่าพยัญชนะต้นแต่ละตัวนั้น เมื่อใช้กับ อ แล้ว จะมีตัวสะกดเป็นตัว ไตได้บ้าง เช่น พยัญชนะต้นเป็นตัว ก สำหรับวรรณยุกต์ (') แล้ว จะมีตัวสะกดเป็นตัวไตได้ บ้าง เช่น พยัญชนะต้นเป็นตัว ก สำหรับวรรณยุกต์แล้ว จะมีตัวสะกดเพียงตัวเดียวคือ น (ก่อน) ส่วนวรรณยุกต์ (") จะมีตัวสะกด คือ น ง ย (ก่อน, ก้อง, ก้อย) เป็นต้น

กฎข้อที่ 11 ตัวอักษรไม้ไตคู่ (") จะเป็นตัวที่เปลี่ยนรูปมาจากสระ -ะ -เะ -เะ -ะ ที่มีตัวสะกด ดังนั้นรูปแบบที่ใช้จะเป็น -อ- -เ- -แ- ทุกรูปแบบจะมีตัวสะกด 1 ตัว ยกเว้น ถ้ามีตัวการ์นต์

กฎข้อที่ 12 สระผสม -เย และ -เอ ที่จะใช้ร่วมกับวรรณยุกต์ต่างๆ โดยจะพิจารณา ว่าแต่ละรูปแบบจะมีตัวสะกดหรือไม่และถ้ามีตัวสะกด ก็จะใช้ตัวไตได้บ้าง เพื่อจะพิจารณา ตำแหน่งแบ่งคำก่อนหน้าหรือหลังสระนั้น

กฎข้อที่ 13 ตัวอักษร ฤ โดยปกติจะให้อยู่ถัดไป จากพยัญชนะต้นนำหน้า เช่น กฤษ ณะ ฤทัย คฤหัส พฤษ เป็นต้น แต่ก็จะมีข้อยกเว้นที่ใช้ตัวอักษร ฤ เป็นพยัญชนะต้นนำหน้า ได้ เช่น ฤดู ฤดี ฤชา ฤษ ฤทธิ ฤทัย ฤษี เป็นต้น

กฎข้อที่ 14 ตัวอักษร ห มักจะใช้เป็นพยัญชนะนำหน้าเสมอ แต่ก็มีคำยกเว้น เช่น สห มหา คหบดี มหกรรม มหรสพ มหศี มหิ พรหม เควาะห์ เป็นต้น นอกจากนั้นจะเป็นคำที่มาจากต่างประเทศ เช่น จอห์น โอห์ม เป็นต้น หรือมีคำนำหน้าเป็นสระ เช่น เหง แห่ง เป็นต้น

กฎข้อที่ 15 ตัวอักษร ว จะต้องมีตัวสะกดอย่างน้อย 1 ตัว เมื่อใช้อยู่ต่อจากวรรณยุกต์ เช่น ม้วน ล้วน ม่วง เป็นต้น และอาจจะใช้ในรูปของสระอว (-ว) เช่น ตัว มัว ชัว เป็นต้น

กฎข้อที่ 16 ตัวอักษร ร โดยปกติจะใช้ในรูปแบบของสระ -รร เช่น วรรณ จรรยา บรรจง เป็นต้น

กฎข้อที่ 17 เป็นกฎของสระลดรูป ซึ่งจะมีตัวสะกดเป็นตัวพยัญชนะต่อไปนี้ คือ ก ง ค น ม บ เช่น คน ชก กค เป็นต้น

กฎข้อที่ 18 ตัวอักษร ศ ณ ญ ษ รุ ฎ ฏ ฒ ฬ ฌ มักจะใช้เป็นตัวสะกดเสมอ เช่น กีฬา คณิกา อุษา อัศจรรย์ หลึง ขญา ปฎัก แต่ก็มีข้อยกเว้นที่จะใช้เป็นพยัญชนะต้นนำหน้าคำได้ เช่น ศุภี ฐาน ฎีกา ญวร เป็นต้น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ข.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างข้อมูลประเภทโคลงกลอน

นิพนธ์ไว้เพื่อกู้	วิญญาณ
กลางคลื่นกระแสนกาล	เชียวกล้า
ชีวิตมีนิรันด	เปลื้องเปล่า
ใจเปล่งแวทพิพท์ห้า	ชั่วฟ้าดินสลาย

จิตการารกรุ่นใหม่	โถมไป ก็ดี
ภาพยร่าหอมแรงใจ	ไปแล้ว
จุดที่ภพไหน	ภพนั้น
ขวัญท่อมพิพท์รุ่งแก้ว	ร่วงน้ำมณีสมัย

ลายสือไหววิเวกให้	เหตุหรรษ์
ฝนท่าแก้วจากสวรรค์	ดับร้อน
ใจปลิวลิวไปฝัน	โลกอื่น
หอมภพนี้สะท้อน	ภพหน้ามาหอม

ซ้ายอมสละทอดทิ้ง	ชีวิต
หวังสิ่งสินนฤมิต	ใหม่แพรว
วิชากรวิจุงค์ศักดิ์สิทธิ์	สูงสุด
ขลังตั้งบุหงาป่าแก้ว	ร่วงฟ้ามาหอม

ตัวอย่างข้อมูลประเภทข่าว

รัฐมนตรีกลาโหมอิสราเอลเตือนอาหรับ

รัฐมนตรีกลาโหมของอิสราเอลเตือนว่า สถานการณ์ในตะวันออกกลาง จะยังคงตึงเครียดต่อไป จนกว่าประธานาธิบดีปาเลสไตน์ จะยุติความรุนแรงในอิสราเอล

นาย "ยิตซ์ค มอร์ติชาย" รัฐมนตรีว่าการกระทรวงกลาโหม ของอิสราเอล กล่าวก่อนเข้าหารือกับนาย "วิลเลียม โคเฮน" รัฐมนตรีกลาโหมของสหรัฐฯ ว่า อิสราเอลจะไม่ยุติการสร้างบ้านเรือน ในพื้นที่ทางตะวันออกของ กรุงเยรูซาเล็มและเขตเวสต์แบงก์ อย่างเด็ดขาด จนกว่าผู้นำปาเลสไตน์ จะยุติความรุนแรง ภายในอิสราเอล

โดยเฉพาะอย่างยิ่ง ในตอนเริ่มต้นนั้น ประธานาธิบดี "อาราฟัต" และพรรคพวก ต้องต่อสู้กับกลุ่มผู้ก่อการร้าย พร้อมระบุว่า จะไม่ให้มีการก่อการร้าย หรือเหตุรุนแรงใดๆ เกิดขึ้นในพื้นที่ยึดครอง ดังนั้น จึงเป็นสิ่งจำเป็นสำหรับอิสราเอล ที่จะดำเนินการทุกอย่าง ด้วยวิธีสันติภาพ

ส่วน นายกรัฐมนตรี "เบนจามิน เนทันยาฮู" แห่งอิสราเอล ก็มีกำหนดจะเดินทาง ไปยังสหรัฐอเมริกา ในสัปดาห์หน้า เพื่อหารือปัญหา สันติภาพตะวันออกกลาง กับประธานาธิบดี "บิล คลินตัน" แห่งสหรัฐฯ

ตัวอย่างข้อมูลประเภทเนื้อเพลง

เพลง ความรัก (อโศกบาน)

รักคือคำ คำนี้

รักคือความอดทนทุกอย่าง จริงใจให้กัน

รักคือความเข้าใจ

รักคือยอมอมภัย ทุกอย่าง ภัยให้กัน

ตั้งดวงตะวัน ที่ยังยั้งยืนคู่ฟ้า

ความรักจึงบังเกิดมา ให้เป็นภาษาทางใจ

ให้ไว้เพื่อช่วยนำทาง คูใจของเรา

รักคือเธอและฉัน รักคือความผูกพันยิ่งใหญ่

จากใจของเรา เพราะเราคู่กัน

ตั้งดวงตะวัน ที่ยังยั้งยืนคู่ฟ้า

ความรักจึงบังเกิดมา ให้เป็นภาษาทางใจ

ให้ไว้เพื่อช่วยนำทาง คูใจของเรา

รักคือความอ่อนโยน
 รักนำทางสู่ความสำเร็จ เป็นจริงสมใจ
 รักคือเธอและฉัน รักคือความผูกพันยิ่งใหญ่
 จากใจของเรา เพราะเรารักกัน

ตัวอย่างข้อมูลประเภทข้อสอบเข้ามหาวิทยาลัย

11. ข้อใดมีคำสมาสที่มีการสนธิทั้งหมด

- 1) มโนมัย สัมมาทฐิ โยธาธิการ
- 2) คุณูปการ นามาทิโรย ขวภาพ
- 3) ทศนูปกรณ์ ชโลทร นิตรารมณ
- 4) มหาวิทยาลัย รัตนติ รัชนีกร

ตอบข้อ 3

ทศนูปกรณ์ = ทศน + อุปกรณ์
 ชโลทร = ชล + อุทร
 นิตรารมณ = นित्र + อารม

สถาบันวิทยบริการ
 จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ค.



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Sistring

Gonnet [5] กำหนด semi-infinite string หรือ sistring คือส่วนท้ายของข้อความที่ เริ่มต้นจากตำแหน่งต่าง ๆ รูปที่ 2.8 แสดง 4 sistring แรกของข้อความ ข้อความที่มีความ ยาว n จะมี sistring มากที่สุดได้ n sistring

ข้อความ : ขอบคุณคุณที่มอบของขวัญ
sistring: ขอบคุณคุณที่มอบของขวัญ
 ขอบคุณคุณที่มอบของขวัญ
 บคุณคุณที่มอบของขวัญ
 คุณคุณที่มอบของขวัญ

...

รูปที่ ค.1 แสดงตัวอย่าง sistrings

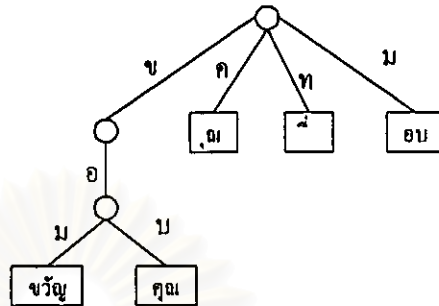
จากรูปที่ ค.1 sistring แรก เริ่มที่ตัวอักษร ข คือ "ขอบคุณคุณที่มอบของขวัญ" sistring ตัวที่สอง เริ่มที่ตัวอักษร อ คือ "อบคุณคุณที่มอบของขวัญ" sistring ตัวที่สาม เริ่มที่ ตัวอักษร บ คือ "บคุณคุณที่มอบของขวัญ" sistring ตัวที่สี่ เริ่มที่ตัวอักษร ค คือ "คุณคุณที่ มอบของขวัญ"

Tries

โครงสร้างข้อมูลแบบทรี (Tries) (Briandais, 1959) เป็นโครงสร้างข้อมูลที่ประกอบด้วย โหนดต่างๆ เช่นเดียวกับโครงสร้างข้อมูลแบบทรี แต่วิธีการเก็บข้อมูลในโครงสร้างข้อมูล แตกต่างกันคือ โครงสร้างข้อมูลแบบทรีใช้ตัวอักษรของคำในการดำเนินการส่วนโครงสร้างข้อมูลแบบทรีใช้คำทั้งคำในการเปรียบเทียบและสร้างโหนดต่างๆ

โครงสร้างข้อมูลแบบทรีประกอบด้วยโหนดต่างๆ ซึ่งสร้างจากตัวอักษรกับตัวชี้ที่ชี้ไปยัง โหนดที่เป็นโหนดลูก ดังรูปที่ ค.2

ข้อความ : ขอบคณที่มอบของขวัญ
 คำที่แบ่งได้ : ขอบคณ
 กณ
 ที่
 มอบ
 ของขวัญ



รูปที่ ค.2 ลักษณะโครงสร้างข้อมูลแบบทรี

เมื่อเปรียบเทียบโครงสร้างแบบทรีกับโครงสร้างแบบ sorted array และบี-ทรี โครงสร้างแบบทรีสิ้นเปลืองเนื้อมก และในการสร้างจะซับซ้อนกว่า ข้อดีคือความเร็วในการค้นหา คำจะขึ้นอยู่กับความยาวของคำ ไม่ขึ้นอยู่กับจำนวนของคำที่เก็บเหมือนโครงสร้างข้อมูลแบบ sorted array และบี-ทรี

สถาบันวิทยบริการ
 จุฬาลงกรณ์มหาวิทยาลัย

ประวัติผู้เขียนวิทยานิพนธ์



นายวิฑูรย์ กัลยาณวัฒน์ เกิดวันที่ 19 สิงหาคม พ.ศ. 2516 ที่ กรุงเทพมหานคร สำเร็จการศึกษาระดับปริญญาตรีวิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง และเข้าศึกษาต่อใน หลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์ ที่จุฬาลงกรณ์มหาวิทยาลัย เมื่อ พ.ศ. 2538 เคยได้รับรางวัลชมเชย ในการสอบแข่งขันคณิตศาสตร์ระดับชั้นมัธยมศึกษาตอนต้น ประจำปี พ.ศ. 2530 ของสมาคมคณิตศาสตร์แห่งประเทศไทย ในพระบรมราชูปถัมภ์ รางวัลชมเชย ในการสอบแข่งขันคณิตศาสตร์ระดับชั้นมัธยมศึกษาตอนปลาย ประจำปี พ.ศ. 2532 ของสมาคมคณิตศาสตร์แห่งประเทศไทย ในพระบรมราชูปถัมภ์ และในปีเดียวกันได้รับคัดเลือกเข้า โครงการส่งเยาวชนไทยไปแข่งขัน คณิตศาสตร์ วิทยาศาสตร์ โอลิมปิกระหว่างประเทศ ได้ผ่านการเข้าค่ายฝึกอบรมแบบเข้ม วิชาคณิตศาสตร์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย