

## รายการอ้างอิง

### ภาษาไทย

ดวงแก้ว สวามิภักดิ์. 2533. การสร้างซอฟต์แวร์วิเคราะห์ไวยากรณ์ไทยภายใต้ระบบยูนิกซ์. กรุงเทพฯ : สำนักพิมพ์มหาวิทยาลัยธรรมศาสตร์.

บุญเรือง ธนาสุนทรไพศาล. 2533. การออกแบบและพัฒนาส่วนเชื่อมโยงสำหรับการตัดคำและการแทรกอักขระแบ่งคำภาษาไทย. วิทยานิพนธ์ ปริญญาโทบริหารศึกษาศาสตร์ สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

พิสิทธิ์ พรมจันทร์. 2540. การวิเคราะห์แนวทางการเปรียบเทียบสมรรถนะของโปรแกรมแยกคำภาษาไทย. วิทยานิพนธ์ ปริญญาโทบริหารศึกษาศาสตร์ สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

ยีน ภู่วรรณ และ วิวรรณ อิมฮารมณ. 2529. การแบ่งแยกพยางค์ไทยด้วยตึกขันทวี. รายงานการประชุมวิชาการวิศวกรรมไฟฟ้า ครั้งที่ 9.

รัตติกร วรากุลศิริพันธ์, จงกล งามวิวิทย์, สมศักดิ์ จันทร์, สุรชาติพิทย์ จิรัชยากุล และ ศักดิ์ชัย ทิพย์จักรภูรัตน์. 2538. การตัดคำจากประโยคภาษาไทยด้วยวิธีการเทียบคำที่ยาวที่สุด. Papers on Natural Language Processing, Compiled by Virach Sornlertlamvanich.

วิรัช ศรีเลิศล้ำวานิช. 2536. การตัดคำภาษาไทยในระบบแปลภาษา. การแปลภาษาด้วยคอมพิวเตอร์. หน้า 50-55. กรุงเทพฯ : ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ.

สมปรารถนา รัตนานนท์. 2535. โครงสร้างข้อมูลสำหรับพจนานุกรมอิเล็กทรอนิกส์ภาษาไทย. วิทยานิพนธ์ ปริญญาโทบริหารศึกษาศาสตร์ สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

สัมพันธ์ ระวีรัมย์. 2534. การแบ่งคำไทยด้วยพจนานุกรม. โครงการวิศวกรรม ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

ภาษาอังกฤษ

- Allen, J. 1995. Natural Language Understanding. 2nd ed. Redwood City, California : Benjamin/Cummings.
- Aoe, J. 1989. An Efficient Digital Search Algorithm by Using a Double-Array Structure. IEEE Trans. Software Eng. Vol. 15, pp. 1066-1077.
- Blum, A. 1997. Empirical Support for Winnow and Weighted-Majority Algorithm: Results on a Calendar Scheduling Domain, Machine Learning, 26: 5-23.
- Chamiak, E. 1996. Statistical Language Learning. Cambridge : MIT Press.
- Chamyapompong, S. 1983. A Thai Syllable Separation Algorithm. Master Thesis. Asian Institute of Technology.
- Charoenpomsawat, P., Kijirikul, B. and Meknavin, S. 1998. Feature-based Thai Unknown Word Boundary Identification Using Winnow. In Proceedings of the 1998 IEEE Asia-Pacific Conference on Circuits and Systems (APCCAS'98).
- Charoenpornsawat, P., Kijirikul, B. and Meknavin, S. 1998. Feature-Based Proper Name Identification in Thai. In Proceeding of the National Computer Science and Engineering Conference'98 (NCSEC'98).
- Cohen, W., W. 1995. Fast Effective Rule Induction. In Proceedings of Twelfth International Conference on Machine Learning. Morgan Kaufmann.
- Corman, T., H., Leiserson, C., E. and Rivest, R., L. 1990. Introduction to Algorithms. Mit Press.
- Frakes, W., B. and Baeza-Yates, R. 1992. Introduction to Data Structures and Algorithms Related to Information Retrieval. New Jersey : Prentice Hall. pp.13-27.
- Golding, A., R. and Roth, D. 1996. Applying Winnow to Context-Sensitive Spelling Correction. In Proceedings of the Thirteenth International Conference on Machine Learning.

- Johnson, S. C. 1975. YACC-Yet another compiler-compiler. NJ. Comput. Sci. Tech. Rep.32: 1-34.
- Kanlayanawat, W., Prasitjutrakul, S. 1997. Automatic Indexing for Thai Text with Unknown Words using Trie Structure. In Proceedings of the Natural Language Processing Pacific Rim Symposium 1997(NLPRS'97).
- Kawtrakul, A., Kumtanode, S., Jamjanya, T. and Jewriyavech C. 1995. A Lexicon Model for Writing Production Assistant System. In Proceedings of the Symposium on Natural Language Processing in Thailand'95.
- Kawtrakul, A., Thumkanon, C., Poovorawan, Y., Varasrai, P. and Suktarachan, M. 1997. Automatic Thai Unknown Word Recognition. In Proceedings of the Natural Language Processing Pacific Rim Symposium 1997(NLPRS'97).
- Kijsirikul, B., Sinthupinyo, S. and Supanwansa, A. 1998. Thai Printed Character Recognition by Combining Inductive Logic Programming with Backpropagation Neural Network. In Proceedings of the 1998 IEEE Asia-Pacific Conference on Circuits and Systems (APCCAS'98).
- Littlestone, N. 1998. Learning Quickly when Irrelevant Attributes Abound: A New Linear-Threshold Algorithm. Machine Learning 2.
- Meknavin, S., Charoenpomsawat, P. and Kijsirikul, B. 1997. Feature-based Thai Word Segmentation. In Proceedings of the Natural Language Processing Pacific Rim Symposium 1997(NLPRS'97).
- Meknavin, S., Kijsirikul, B., Chotimongkol, A. and Nuttee, C. 1998. Progress of Combining Trigram and Winnow in Thai OCR Error Correction. In Proceedings of the 1998 IEEE Asia-Pacific Conference on Circuits and Systems (APCCAS'98).

Thairatananond, Y. 1981. Towards the design of a Thai text syllable analyzer. Master Thesis. Asian Institute of Technology.

Somlertlamvanich, V., Charoenpom, T. and Isahara, H. 1997. ORCHID: Thai Part-Of-Speech Tagged Corpus. In Technical Report Orchid Corpus. Bangkok : NECTEC.



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

## ภาคผนวก ก

## กฎที่ใช้ในการตัดพยางค์ ของดวงแก้ว สวามิภักดิ์

กฎที่ใช้ในการตัดพยางค์มีอยู่ทั้งหมด 43 กฎ โดยสัญลักษณ์ที่ใช้มีดังต่อไปนี้

c	::= พยัญชนะปรกติ (Consonant)
v	::= สระ (Vowel)
t	::= วรรณยุกต์ (Tonal Mark)
s	::= ตัวสะกด (speller)
[...]?	::= ทางเลือก กล่าวคือ อาจจะมีหรือไม่มีก็ได้
[a1 a2 ... an]	::= เลือกตัวอักษรใดตัวอักษรหนึ่งระหว่าง a1 ... an

กฎที่ได้ 43 กฎมีดังต่อไปนี้

1. [c][t]?[ะ ำ า]
2. [c][ ี ื , ุ ]?[ะ ำ า]
3. [c][ ี ื , ุ ][t]?
4. [c][t]?[s]
5. [c] [t]?[s]
6. [เ เน โ ไ ] [c][t]?
7. [เ เน ] [c] ี [s]
8. [เ เน โ ] [c][t]?ะ
9. [เ เน โ ] [กขคตทปพฟจชศส]ร[t]?ะ
10. [เ เน โ ] [กขคปพฝล]ล[t]?ะ
11. [เ เน โ ] [กขคตทปพฟจชศส]ร[ ี ื ุ ] [t]?
12. [เ เน โ ] [กขคปพฝล]ล[ ี ื ุ ] [t]?
13. [เ เน โ ] [กขค]ว[t]?ะ
14. [c] [t]?ย
15. [c][t]?าะ
16. [c][t]?[าะ]
17. [c][t]?
18. [เ ไ ] ห[งญนมยรลว][t]?
19. [c] [t]?อ
20. [c] [t]?อ
21. [c]



## ภาคผนวก ข

## ตารางแสดงหน้าที่ของคำในภาษาไทย จากคลังข้อความอรรถคดี

ประเภทของคำ	รายละเอียด	ตัวอย่าง
NPRP	Proper noun	วินโดวส์ 95, โคอโรน่า, โค้ก, พระอาทิตย์
NCNM	Cardinal number	หนึ่ง, สอง, สาม, 1, 2, 3
NONM	Ordinal number	ที่หนึ่ง, ที่สอง, ที่สาม, ที่ 1, ที่ 2, ที่ 3
NLBL	Label noun	1, 2, 3, 4, ก, ข, ค, ง
NCMN	Common noun	หนังสือ, อาหาร, อาจารย์, คน
NTTL	Title noun	ดร., พลเอก
PPRS	Personal pronoun	คุณ, เขา, ฉัน
PDMN	Demonstrative pronoun	นี้, นั่น, ที่นั่น, ที่นี่
PNTR	Interrogative pronoun	ใคร, อะไร, อย่างไร
PREL	Relative pronoun	ที่, ซึ่ง, อัน, ผู้
VACT	Active verb	ทำงาน, ร้องเพลง, กิน
VSTA	Stative verb	เห็น, รู้, คือ
VATT	Attribute verb	อ้วน, คี, สวย
XVBM	Pre-verb auxiliary, before negator “ไม่”	เกิด, เกือบ, กำลัง
XVAM	Pre-verb auxiliary, after negator “ไม่”	ค่อย, น่า, ได้
XVMM	Pre-verb, before or after negator “ไม่”	ควร, เคย, ต้อง
XVBB	Pre-verb auxiliary, in imperative mood	กรุณา, จง, เชิญ, อย่า, ห้าม
XVAE	Post-verb auxiliary	ไป, มา, ขึ้น
DDAN	Definite determiner, after noun without classifier in between	นี้, นั่น, โน่น, ทั้งหมด



ประเภทของคำ	รายละเอียด	ตัวอย่าง
DDAC	Definite determiner, allowing classifier in between	นี้, นั้น, โน้น, นู้น
DDBQ	Definite determiner, between noun and classifier or preceding Quantitative expression	ทั้ง, อีก, เพียง
DDAQ	Definite determiner, following quantitative expression	พอดี, ถ้วน
DIAC	Indefinite determiner, following noun; allowing classifier in between	ไหน, อื่น, ต่างๆ
DIBQ	Indefinite determiner, Between noun and classifier or preceding quantitative expression	บาง, ประมาณ, เกือบ
DIAQ	Indefinite determiner, following quantitative expression	กว่า, เศษ
DCNM	Determiner, cardinal number expression	หนึ่งคน, สอง 2 ตัว
DONM	Determiner, ordinal number expression	ที่หนึ่ง, ที่สอง, ที่สุดท้าย
ADVN	Adverb with noun form	เก่ง, เร็ว, ช้า, สม่่าเสมอ
ADVI	Adverb with iterative form	เร็วๆ, เสมอๆ, ช้าๆ
ADVP	Adverb with prefixed form	โดยเร็ว
ADVS	Sentential adverb	โดยปกติ, ธรรมดา
CNIT	Unit classifier	ตัว, คน, เล่ม
CLTV	Collective classifier	คู่, กลุ่ม, ผุ่่ง, เซิง, ทาง, ด้าน, แบบ, รุ่น
CMTR	Measurement classifier	กิโลกรัม, แก้ว, ชั่วโมง

ประเภทของคำ	รายละเอียด	ตัวอย่าง
CFQC	Frequency classifier	ครั้ง, เทียว
CVBL	Verbal classifier	ม้วน, มัด
JCRG	Coordinating conjunction	และ, หรือ, แต่
JSBR	Comparative conjunction	กว่า, เหมือนกับ, เท่ากับ
RPRE	Preposition	จาก, ละ, ของ, ได้, บน
INT	Interjunction	โธ้ย, โธ้, เออ, เอ้, อ้อ
FIXN	Nominal prefix	การทำงาน, ความสนุกสนาน
FIXV	Adverbial prefix	อย่างรวดเร็ว
EAFF	Ending for affirmative sentence	จ๊ะ, จ๊ะ, ค่ะ, ครับ, นะ, นำ, เอะ
EITT	Ending for interrogative sentence	หรือ, เหรอ, ไหม, มั้ย
NEG	Negator	ไม่, มิได้, ไม่ได้, มิ
PUNC	Punctuation	(, ), “, ,, ๑

สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

## ภาคผนวก ค

## ความถี่ของข้อความที่กำกวมจากคลังข้อความออร์คิด

## ความถี่ของข้อความกำกวมที่แบบไม่ต้องใช้บริบท 50 อันดับแรก

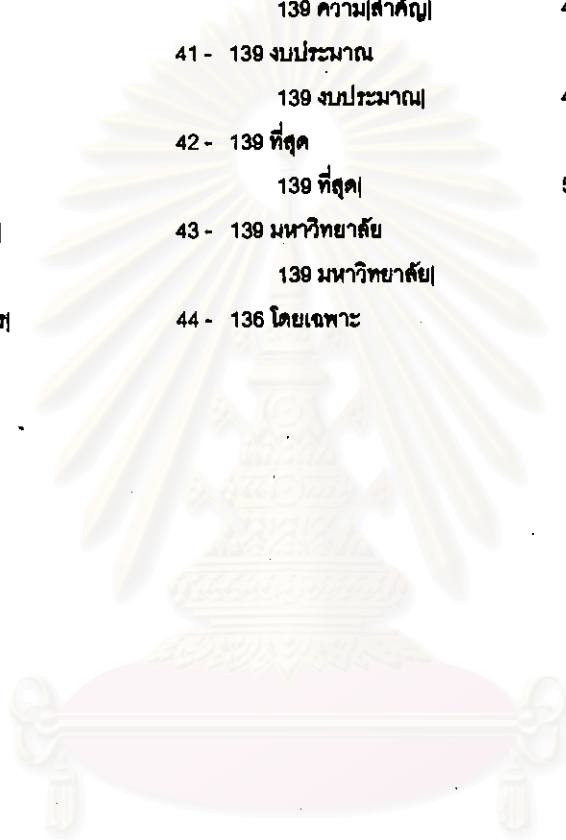
1 - 743 ที่จะ	12 - 211 เพื่อให้	23 - 148 จัดการ
407 ที่จะ	207 เพื่อให้	130 จัดการ
336 ที่จะ	4 เพื่อให้	18 จัดการ
2 - 465 ทำให้	13 - 194 ความต้องการ	24 - 143 ดังกล่าว
346 ทำให้	6 ความต้องการ	141 ดังกล่าว
119 ทำให้	188 ความต้องการ	2 ดังกล่าว
3 - 384 ผิดชอบ	14 - 186 วิธีการ	25 - 142 หรือไม่
226 ผิดชอบ	164 วิธีการ	92 หรือไม่
158 ผิดชอบ	22 วิธีการ	50 หรือไม่
4 - 345 ไม่ได้	15 - 184 การที่	26 - 136 เข้ามา
54 ไม่ได้	174 การที่	129 เข้ามา
291 ไม่ได้	10 การที่	7 เข้ามา
5 - 330 ทำงาน	16 - 180 อาจจะเป็น	27 - 134 ทางด้าน
312 ทำงาน	170 อาจจะเป็น	69 ทางด้าน
18 ทำงาน	10 อาจจะเป็น	65 ทางด้าน
6 - 323 ความรู้	17 - 170 ความสามารถ	28 - 134 มากกว่า
1 ความรู้	1 ความสามารถ	39 มากกว่า
322 ความรู้	169 ความสามารถ	95 มากกว่า
7 - 267 ได้รับ	18 - 168 ในด้าน	29 - 130 ไม่ใช่
253 ได้รับ	86 ในด้าน	50 ไม่ใช่
14 ได้รับ	82 ในด้าน	80 ไม่ใช่
8 - 247 มากขึ้น	19 - 164 เหล่านี้	30 - 129 เท่านั้น
152 มากขึ้น	163 เหล่านี้	127 เท่านั้น
95 มากขึ้น	1 เหล่านี้	2 เท่านั้น
9 - 242 ควรจะ	20 - 158 ปฏิบัติงาน	31 - 122 ตำแหน่งงาน
232 ควรจะ	120 ปฏิบัติงาน	114 ตำแหน่งงาน
10 ควรจะ	38 ปฏิบัติงาน	8 ตำแหน่งงาน
10 - 240 แต่ละ	21 - 157 ต่อไป	32 - 119 เป็นไป
239 แต่ละ	156 ต่อไป	111 เป็นไป
1 แต่ละ	1 ต่อไป	8 เป็นไป
11 - 232 ดำเนินการ	22 - 153 ประเทศไทย	33 - 116 เพิ่มขึ้น
223 ดำเนินการ	142 ประเทศไทย	82 เพิ่มขึ้น
9 ดำเนินการ	11 ประเทศไทย	34 เพิ่มขึ้น

34 - 114 ทางเศรษฐกิจ	107 ความสัมพันธ์	68 การเงิน
82 ทางเศรษฐกิจ	40 - 108 ค่าการ	31 การเงิน
32 ทางเศรษฐกิจ	2 ค่าการ	46 - 98 การเมือง
35 - 113 คงจะ	106 ค่าการ	86 การเมือง
112 คงจะ	41 - 105 ทำได้	12 การเมือง
1 คงจะ	26 ทำได้	47 - 95 ความจำเป็น
36 - 112 เกิดขึ้น	79 ทำได้	4 ความจำเป็น
58 เกิดขึ้น	42 - 103 ในทาง	91 ความจำเป็น
54 เกิดขึ้น	42 ในทาง	48 - 91 มีประสิทธิภาพ
37 - 112 มีที่จะ	61 ในทาง	22 มีประสิทธิภาพ
110 มีที่จะ	43 - 102 วางแผน	69 มีประสิทธิภาพ
2 มีที่จะ	101 วางแผน	49 - 91 ทุกคน
38 - 111 ในประเทศ	1 วางแผน	76 ทุกคน
33 ในประเทศ	44 - 100 แล้วก็	15 ทุกคน
78 ในประเทศ	29 แล้วก็	50 - 90 ก็ได้
39 - 109 ความสัมพันธ์	71 แล้วก็	47 ก็ได้
2 ความสัมพันธ์	45 - 99 การเงิน	43 ก็ได้

#### ความถี่ของความถี่แบบที่ไม่ต้องใช้บริบท 50 อันดับแรก

1 - 812 สามารถ	253 จำเป็น	20 - 204 การบริหาร
812 สามารถ	11 - 234 ตนเอง	204 การบริหาร
2 - 464 ชำรภาพการ	234 ตนเอง	21 - 203 ปรับปรุง
464 ชำรภาพการ	12 - 229 รวมทั้ง	203 ปรับปรุง
3 - 388 ระดับ	229 รวมทั้ง	22 - 203 หน่วยงาน
388 ระดับ	13 - 229 หน้าที่	203 หน่วยงาน
4 - 353 ดึงดูด	229 หน้าที่	23 - 201 หัวหน้า
353 ดึงดูด	14 - 228 คุณภาพ	201 หัวหน้า
5 - 310 ข้อมูล	228 คุณภาพ	24 - 200 ส่งเสริม
310 ข้อมูล	15 - 226 รายได้	200 ส่งเสริม
6 - 279 การศึกษา	226 รายได้	25 - 197 อย่างไร
279 การศึกษา	16 - 224 สินค้า	197 อย่างไร
7 - 271 เกี่ยวกับ	224 สินค้า	26 - 192 เอกชน
271 เกี่ยวกับ	17 - 218 ระหว่าง	192 เอกชน
8 - 262 ประชาชน	218 ระหว่าง	27 - 187 เหมาะสม
262 ประชาชน	18 - 216 เปลี่ยนแปลง	187 เหมาะสม
9 - 258 อบรม	216 เปลี่ยนแปลง	28 - 169 กระทำ
258 อบรม	19 - 206 ลงทุน	169 กระทำ
10 - 253 จำเป็น	206 ลงทุน	29 - 167 ราคา

167 ราคา	37 - 147 ขนบท	136 โดยเฉพาะ
30 - 167 ปกครอง	147 ขนบท	45 - 136 กำสั่งคน
167 ปกครอง	38 - 145 โครงสร้าง	136 กำสั่งคน
31 - 162 เข้าใจ	145 โครงสร้าง	46 - 133 เกี่ยวข้อง
162 เข้าใจ	39 - 140 ป้องกัน	133 เกี่ยวข้อง
32 - 160 พื้นฐาน	140 ป้องกัน	47 - 131 องค์การ
160 พื้นฐาน	40 - 139 ความสำคัญ	131 องค์การ
33 - 155 กำสั่ง	139 ความสำคัญ	48 - 130 บทบาท
155 กำสั่ง	41 - 139 งบประมาณ	130 บทบาท
34 - 153 แก๊โซ	139 งบประมาณ	49 - 128 กฎหมาย
153 แก๊โซ	42 - 139 ที่สุด	128 กฎหมาย
35 - 150 ควบคุม	139 ที่สุด	50 - 125 กิจกรรม
150 ควบคุม	43 - 139 มหาวิทยาลัย	125 กิจกรรม
36 - 148 โครงการ	139 มหาวิทยาลัย	
148 โครงการ	44 - 136 โดยเฉพาะ	



สถาบันวิทยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย

## ประวัติผู้เขียน

นายไพศาล เจริญพรสวัสดิ์ เกิดวันที่ 16 กันยายน พ.ศ. 2517 กรุงเทพมหานคร สำเร็จการศึกษาปริญญาตรีวิศวกรรมศาสตร์ สาขาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปี 2539 และเข้าศึกษาต่อในหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์ ที่จุฬาลงกรณ์มหาวิทยาลัย เมื่อ พ.ศ. 2540 ปัจจุบันทำงานตำแหน่งผู้ช่วยนักวิจัย ห้องปฏิบัติการวิจัยและพัฒนาภาษาและซอฟต์แวร์ ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ



สถาบันวิจัยบริการ  
จุฬาลงกรณ์มหาวิทยาลัย