

บทที่ 4

การกำหนดค่าน้ำหนัก (Weighting) กฎไวยากรณ์ภาษา

1. ความเบื้องต้น

การให้น้ำหนักคำ (Term Weighting) เป็นการกำหนดค่าน้ำหนักของคำหรือวลีเพื่อการเรียนรู้และทดสอบโครงข่ายประสาทเทียม โดยเป็นหลักการให้ความสำคัญกับคำหรือวลีในด้านของคำสถิติศาสตร์และไวยากรณ์ภาษาศาสตร์ คือ

- ค่าสถิติความถี่การเกิดขึ้นของคำ จากหลักไวยากรณ์ภาษาในการใช้คำหรือวลีเดิมซ้ำ ๆ เพื่อเป็นการเน้นย้ำถึงความสำคัญของคำหรือวลีนั้นในเอกสาร
- ค่าความสำคัญทางไวยากรณ์ภาษา ในด้านตำแหน่งการเกิดขึ้นของแต่ละคำหรือวลีในส่วนต่าง ๆ ของประโยค
- ค่าความสำคัญทางไวยากรณ์ภาษา ในด้านตำแหน่งการเกิดขึ้นของแต่ละคำหรือวลีในส่วนต่าง ๆ ของย่อหน้า
- ค่าความสำคัญทางไวยากรณ์ภาษา ในตำแหน่งการเกิดขึ้นของแต่ละคำหรือวลีในส่วนต่าง ๆ ของเอกสาร
- ค่าความสำคัญทางไวยากรณ์ภาษา ในการให้ความสำคัญกับหน้าที่ของแต่ละคำหรือวลีในการเขียน

2. รูปแบบการกำหนดน้ำหนักความสำคัญของคำหรือวลีในเอกสาร

ลักษณะทางไวยากรณ์ของภาษา ในด้านหลักการของการเขียนเอกสารภาษาไทยต่าง ๆ อาทิเช่น การใช้คำ การแต่งประโยค การเขียนย่อหน้า การเขียนเอกสาร เหล่านี้ล้วนมีผลต่อใจการกำหนดใจความสำคัญของเอกสาร ดังนั้น ผู้วิจัยจึงได้นำมาเป็นแนวคิดในการสร้างกฎของการให้น้ำหนักความสำคัญของคำหรือวลีในงานวิจัยนี้ โดยเป็นการกำหนดรูปแบบค่าน้ำหนักของคำหรือวลีภาษาไทยที่สัมพันธ์กับคำสถิติศาสตร์และไวยากรณ์ ดังนี้

2.1 คำห้คาคำความถี่การเกิดขึ้นของคำ (Word & Phrase Frequency)

2.1.1 แนวคิดสู่กฎเกี่ยวกับการให้ว้ความสำคัญของคำซ้ำ

แนวค้ดทางด้า่นไวยากรณ์ภาษาเกี่ยวกับการให้ว้ความสำคัญกับคำความถี่การเกิดขึ้นของคำนั้น มีค้ดงนี้

2.1.1.1 คำหรือวลีที่มีความสำคัญที่สามารถเป็นตัวแทนของเอกสารได้นั้น ค้ดกล่าวใน ล้ดดา ยินดีมาก โกรดี (2534) นั้น ควรจะปรากฏอยู่เป็นจ้านวนมากในเนื้อหาของเอกสารเฉพาะฉบับนั้น และปรากฏอยู่น้อยมากในชุดเอกสารที่เหลือทั้งหมด แต่ถ้าคำนั้นปรากฏเป็นจ้านวนมากในทุก ๆ เอกสาร แสดงว่าคำค้ดกล่าวไม่สามารถเป็นตัวแทนของเอกสารใด ๆ ได้ ค้ดงั้นการให้น้าหนักคำหรือวลีหนึ่ง ๆ ในเอกสารฉบับหนึ่งจะพิจารณาจากความถี่ของการเกิดคำหรือวลีที่ปรากฏในเอกสารนั้น ค้ด การจ้ดน้าหนักให้กับคำ เพื่อเลือกเฉพาะคำหรือวลีที่มีน้าหนักมากหรือความถี่สูงไว้เป็นสำคัญ (keyword & keyphrase) ของเอกสาร โดยในที่นี้ใช้ 2 วิธีรวมกัน ค้ด

2.1.1.1.1 การพิจารณาจากความถี่สัมบูรณ์ของคำ (absolute frequency) โดยค้ดน้างว่าผู้เขียนต้องการเน้นเนื้อหาเกี่ยวกับเรื่องอะไรก็จะเขียนคำ ๆ นั้นบ่อยครั้ง ในกรณีน้าหนักของคำ ค้ดความถี่ของคำนั้น

2.1.1.1.2 การพิจารณาจากวลี (phrase or token) ที่ประกอบเข้าเป็นคำ เนื่องจากคำไทยส่วนใหญ่เป็นคำประสมจากหลาย ๆ คำรวมกันเป็น 1 คำ

2.1.1.2 กฎไวยากรณ์ภาษาศาสตร์เกี่ยวกับการใช้คำซ้ำ ค้ดที่กล่าวเกี่ยวกับการใช้คำซ้ำ¹ ในด้านการใช้คำที่ต้องการเน้นเป็นใจความสำคัญของเอกสารนั้น ๆ ค้ด ใช้คำเดิมซ้ำเพื่อเน้นน้าหนักของคำ

2.1.1.3 การย้าเน้นโดยใช้คำ วลี หรือประโยคซ้ำ ๆ กัน นั้น การใช้คำเดิมซ้ำ ๆ กันในเอกสารเป็นการทำให้ผู้อ่านสามารถเข้าใจจุดมุ่งหมายหรือความคิดสำคัญที่ผู้เขียนต้องการสื่อสาร

2.1.2 การกำหนดน้าหนักคำความถี่การเกิดขึ้นของคำ

ค้ดน้าหนักความถี่ ค้ด การนับจ้านวนคำเดิมที่เกิดขึ้นซ้ำในเอกสาร เพื่อเป็นการวิเคราะห์ที่แสดงให้เห็นถึงการเน้นย้าหรือการใช้ซ้ำในแต่ละเอกสาร

¹ ดูเทคนิคการใช้คำซ้ำ ในบทที่ 2 หน้า 12

2.2 การกำหนดค่าน้ำหนักตำแหน่งของคำหรือวลีในแต่ละประโยค

2.2.1 แนวคิดสู่กฎเกี่ยวกับการให้ความสำคัญของตำแหน่งของคำหรือวลีสำคัญในประโยค

(Position of Sentence)

แนวคิดทางด้านไวยากรณ์ภาษาเกี่ยวกับการให้ความสำคัญกับตำแหน่งการเกิดขึ้นของคำหรือวลีในประโยคนั้น คือ การให้ความสำคัญในการเขียนและจัดวางตำแหน่งของคำหรือวลีในประโยคในภาษาไทย² มีรายละเอียดดังนี้

2.2.1.1 น้ำหนักของคำหรือวลีมักจะอยู่ตอนต้นของประโยค จึงเป็นที่มาของการให้ค่าน้ำหนักมากในตอนต้นของประโยคในงานวิจัย

2.2.1.2 การจัดตำแหน่งของคำในประโยคให้ตรงกับฐานน้ำหนัคนั้น โดยฐานน้ำหนักในประโยค แบ่งออกเป็น 3 ตอน คือ ตอนต้นประโยค กลางประโยค และปลายประโยค จึงเป็นที่มาของการให้ค่าน้ำหนักมากที่สุดในตอนต้นและท้ายประโยค และการให้ค่าน้ำหนักรองลงมาที่ตอนกลางของประโยคในงานวิจัย

2.2.1.3 คำหรือวลีสรุปความ หรือ การใช้คำสรุปความก่อนเน้นคำสำคัญในตอนท้ายประโยค จึงเป็นที่มาของการให้ค่าน้ำหนักมากในตอนท้ายของประโยคในงานวิจัย

2.2.2 การให้ค่าน้ำหนักต่าง ๆ ของคำหรือวลีในประโยค

ความสัมพันธ์ของคำหรือวลีกับตำแหน่งในประโยค ดังหลักไวยากรณ์ที่ได้กล่าวมานั้น ผู้วิจัยได้ทำการกำหนดค่าน้ำหนักระดับความสำคัญของคำในประโยค ดังนี้

- ตำแหน่งส่วนต้นและส่วนท้ายของประโยค ที่เป็นส่วนเกริ่นนำและสรุป มีค่าน้ำหนักใจความสำคัญ คือ 1.0 ดังภาพประกอบ 4-1 และ 4-2

- ตำแหน่งส่วนกลางของประโยค มีค่าน้ำหนักใจความสำคัญ คือ 0.75 ดังภาพประกอบ 4-1 และ 4-2

- ตำแหน่งในส่วนถัดจากส่วนต้นของประโยคและตำแหน่งส่วนก่อนสิ้นสุดประโยค มีค่าน้ำหนักใจความสำคัญ คือ 0.5 ดังภาพประกอบ 4-1 และ 4-2

² ดูเทคนิคการใช้คำในประโยค ในบทที่ 2 หน้า 12-13

2.2.3 การกำหนดค่าน้ำหนักตำแหน่งของคำหรือวลีในแต่ละประโยค

ผู้วิจัยได้ทำการแบ่งประโยคออกเป็น 5 ส่วน เท่า ๆ กันในแนวตั้ง (Vertical) ตั้งแต่เริ่มต้นประโยคจนถึงสุดประโยคจากการตรวจสอบเคาะวรรค 2 เคาะ (space) และทำการกำหนดค่าน้ำหนักแต่ละคำหรือวลีนั้นในรูปแบบของเมตริกซ์น้ำหนัก โดยการอ้างอิงจากหลักการในการวางระดับความสำคัญของแต่ละตำแหน่งในประโยคดังกล่าวในบทที่ 2 โดยค่าน้ำหนักตำแหน่งในประโยคนั้น มีรายละเอียดค่าน้ำหนักดังนี้

$$\text{Weight of Sentence Position} = [1, 0.5, 0.75, 0.5, 1]$$

ภาพประกอบที่ 4-1 รูปแบบการกำหนดค่าน้ำหนักตำแหน่งในประโยค

2.3 การกำหนดค่าน้ำหนักตำแหน่งของคำหรือวลีในแต่ละย่อหน้านั้น

2.3.1 แนวคิดสู่กฎเกี่ยวกับการให้ความสำคัญของตำแหน่งของคำหรือวลีสำคัญในย่อหน้า (Position of Paragraph)

หลักการเขียนย่อหน้า³ เกี่ยวกับลักษณะของย่อหน้าที่ดีและลักษณะงานเขียนที่ดีนั้น ในด้านของการเขียนให้ย่อหน้ามีสารัตถภาพ คือ การย้ำเน้นในที่ที่ควรเน้น โดยมีแนวคิดทางด้านไวยากรณ์ภาษาเกี่ยวกับการให้ความสำคัญกับตำแหน่งการเกิดขึ้นของคำหรือวลีในย่อหน้า ดังนี้

2.3.1.1 การวางตำแหน่งประโยคใจความสำคัญในย่อหน้า ลำดับที่ 1

คือ การให้ค่าน้ำหนักหรือการให้ความสำคัญของตำแหน่งคำหรือวลี โดยการให้ค่าน้ำหนักมากในช่วงบรรทัดตอนต้นและช่วงบรรทัดตอนท้ายของย่อหน้า

2.3.1.3 โดยการวางตำแหน่งประโยคใจความสำคัญลำดับที่ 2

คือ การให้ค่าน้ำหนักมากรองลงมาในช่วงบรรทัดตอนกลางของย่อหน้า

³ ดูเทคนิคการใช้คำหรือวลีในย่อหน้าในบทที่ 2 หน้า 13-19

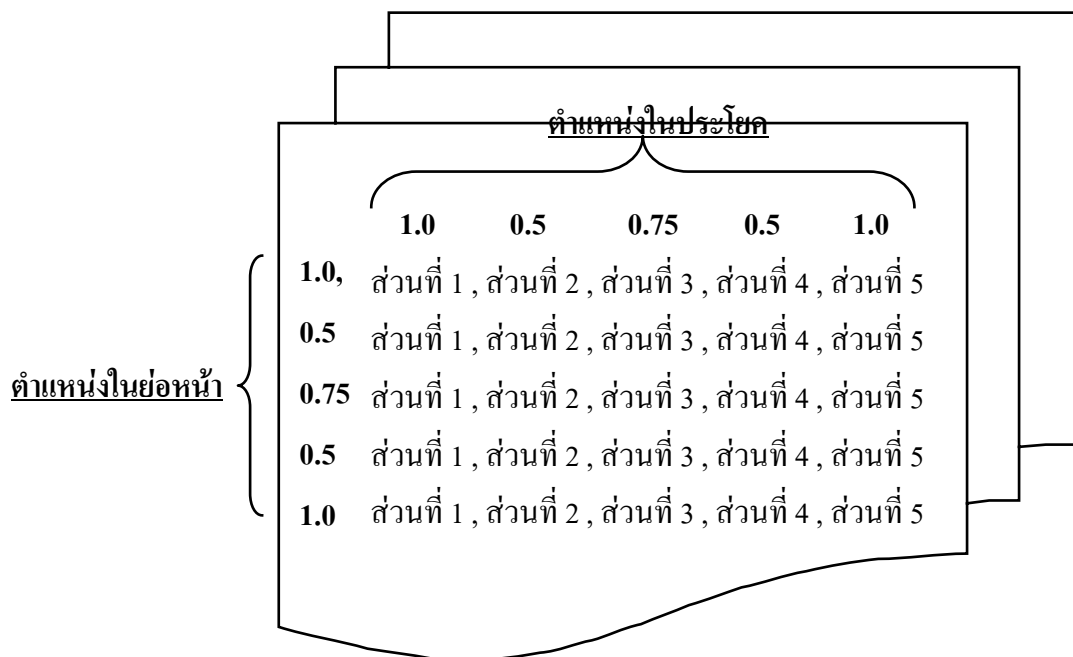
2.3.2 หลักการให้ค่าน้ำหนักในย่อหน้า

ความสัมพันธ์ของคำหรือวลีกับตำแหน่งในย่อหน้า จากหลักไวยากรณ์ดังที่ได้กล่าวมานั้น ผู้วิจัยจึงได้ทำการกำหนดค่าน้ำหนักระดับความสำคัญของคำในย่อหน้า ดังนี้

- ตำแหน่งของคำในช่วงบรรทัดตอนต้นและตำแหน่งของคำในช่วงบรรทัดส่วนท้ายของย่อหน้า มีค่าน้ำหนักใจความสำคัญ คือ 1.0 ซึ่งเป็นค่าน้ำหนักที่แสดงถึงระดับความสำคัญของตำแหน่งการเกิดขึ้นของคำหรือวลีสำคัญในบรรทัดย่อหน้านั้น ๆ ดังภาพประกอบ 4-2

- ตำแหน่งของคำในช่วงบรรทัดตอนกลางของย่อหน้า มีค่าน้ำหนักใจความสำคัญ คือ 0.75 ซึ่งเป็นค่าน้ำหนักที่แสดงถึงระดับความสำคัญระดับที่ 2 ของตำแหน่งการเกิดขึ้นของคำหรือวลีสำคัญในบรรทัดย่อหน้านั้น ๆ ดังภาพประกอบ 4-2

- ตำแหน่งของคำในช่วงบรรทัดในส่วนถัดจากส่วนต้นของย่อหน้า และตำแหน่งในช่วงบรรทัดสิ้นสุดย่อหน้า มีค่าน้ำหนักใจความสำคัญ คือ 0.5 ซึ่งแสดงถึงส่วนขยายของใจความส่วนต้นและส่วนท้ายของย่อหน้า ดังแสดงในภาพประกอบที่ 4-2



ภาพประกอบที่ 4-2 แสดงรูปแบบค่าน้ำหนักของคำในประโยคและในย่อหน้า

2.3.3 การกำหนดค่าน้ำหนักตำแหน่งของคำหรือวลีในย่อหน้า

ผู้วิจัยได้ทำการแบ่งย่อหน้าออกเป็น 5 ส่วน เท่า ๆ กันในแนวนอน (Horizontal) ตั้งแต่เริ่มต้นย่อหน้าด้วยสัญลักษณ์การขึ้นบรรทัดใหม่ (new line) พร้อมกับการย่อหน้า (tap) และจะสิ้นสุดย่อหน้าด้วยการขึ้นบรรทัดใหม่พร้อมการย่อหน้าถัดไป โดยทำการกำหนดค่าน้ำหนักแต่ละคำหรือวลีนั้นในรูปแบบของเมตริกซ์น้ำหนัก โดยการกำหนดระดับความสำคัญช่วงของบรรทัดในแต่ละย่อหน้าแตกต่างกันไป โดยการอ้างอิงจากหลักการในการให้ความสำคัญในแต่ละส่วนของย่อหน้า ดังกล่าวในบทที่ 2 โดยส่วนของย่อหน้าที่เป็นส่วนสำคัญนั้น มักจะเป็นที่วางของตำแหน่งคำหรือวลีสำคัญในย่อหน้า ผู้วิจัยมีการกำหนดรายละเอียดค่าน้ำหนักดังนี้

$$\text{Weight of Paragraph Position} = [1, 0.5, 0.75, 0.5, 1]$$

ภาพประกอบที่ 4-3 รูปแบบการกำหนดค่าน้ำหนักตำแหน่งในย่อหน้า

2.4 การกำหนดค่าน้ำหนักตำแหน่งของคำหรือวลีในแต่ละส่วนของเอกสารนั้น

2.4.1 แนวคิดสู่กฎเกี่ยวกับการให้ความสำคัญของตำแหน่งของคำหรือวลีสำคัญในแต่ละหน้าที่ของคำ (Part of Speech)

การแบ่งหมวดคำเป็นสิ่งสำคัญสำหรับการวิเคราะห์ไวยากรณ์ภาษา เนื่องจากในระบบโครงสร้างภาษา ความสัมพันธ์ระหว่างคำในโครงสร้างไม่ได้เป็นเพียงความสัมพันธ์ระหว่างคำดังกล่าวเท่านั้น แต่ยังแสดงให้เห็นถึงความสัมพันธ์ในด้านความหมายและการใช้ระหว่างหมวดคำของภาษาด้วย โดยมีความสัมพันธ์ของคำหรือวลีในการกำกับหน้าที่คำ ที่มีด้วยกันทั้งสิ้น 47 คำน้ำหนัก ที่เป็นการกำหนดชนิดของหน้าที่คำโดยนักภาษาศาสตร์ การกำหนดค่าน้ำหนักหน้าที่คำนี้ดังจะกล่าวในหัวข้อ 2.4.2

2.4.2 การให้ค่าน้ำหนักหน้าที่คำ

ลำดับที่	หน้าที่คำ	รายละเอียด	ค่าน้ำหนัก
1	NPRP	Proper noun	1.0
2	NCNM	Cardinal number	0.5
3	NONM	Ordinal number	0.5
4	NLBL	Label noun	0.5
5	NCMN	Common noun	1.0
6	NTTL	Title noun	0.5
7	PPRS	Personal pronoun	0.2
8	PDMN	Demonstrative pronoun	0.1
9	PNTR	Interrogative pronoun	0.1
10	PREL	Relative pronoun	0.2
11	VACT	Active verb	0.6
12	VSTA	Stative verb	0.6
13	VATT	Attributive verb	0.5
14	XVBM	Pre-verb auxiliary, before negator “ไม่”	0.3
15	XVAM	Pre-verb auxiliary, after negator “ไม่”	0.3
16	XVMM	Pre-verb, before or after negator “ไม่”	0.3
17	XVBB	Pre-verb auxiliary, in imperative mood	0.5
18	XVAE	Post-verb auxiliary	0.5
19	DDAN	Definite determiner, after noun without classifier in between	0.1
20	DDAC	Definite determiner, allowing classifier in between	0.1
21	DDBQ	Definite determiner, between noun and classifier or preceding quantitative expression	0.1
22	DDAQ	Definite determiner, following quantitative expression	0.1
23	DIAC	Indefinite determiner, following noun; allowing classifier in between	0.1

ลำดับที่	หน้าที่คำ	รายละเอียด	ค่าน้ำหนัก
24	DIBQ	Indefinite determiner, between noun and classifier or preceding quantitative expression	0.2
25	DIAQ	Indefinite determiner, following quantitative expression	0.2
26	DCNM	Determiner, cardinal number expression	0.3
27	DONM	Determiner, ordinal number expression	0.3
28	ADV N	Adverb with normal form	0.3
29	ADV I	Adverb with iterative form	0.3
30	ADV P	Adverb with prefixed form	0.3
31	ADV S	Sentential adverb	0.3
32	CNIT	Unit classifier	0.5
33	CLTV	Collective classifier	0.5
34	CMTR	Measurement classifier	0.5
35	CFQC	Frequency classifier	0.5
36	CVBL	Verbal classifier	0.5
37	JCRG	Coordinating conjunction	0.0
38	JCMP	Comparative conjunction	0.0
39	JSBR	Subordinating conjunction	0.0
40	RPRE	Preposition	0.1
41	INT	Interjection	0.0
42	FIXN	Nominal prefix	0.0
43	FIXV	Adverbial prefix	0.1
44	EAFF	Ending for affirmative sentence	0.0
45	EITT	Ending for interrogative sentence	0.0
46	NEG	Negator	0.1
47	PUNC	Punctuation	0.0

ตารางที่ 4-1 ตารางการกำหนดค่าน้ำหนักหน้าที่คำในงานวิจัย

2.4.3 การกำหนดค่านำหน้าตำแหน่งของคำหรือวลีในตำแหน่งเอกสาร

2.4.3.1 แนวคิดสู่กฎเกี่ยวกับการให้ความสำคัญของตำแหน่งของคำหรือวลีสำคัญในเอกสาร (Position of Document)

การลำดับความนึกคิดในเอกสาร⁴ นั้น เป็นการอธิบายถึงการลำดับการเขียนเรื่องและความสำคัญของแต่ละส่วนในเอกสาร มีด้วยกันดังนี้

2.4.3.1.1 ชื่อเรื่อง คือ การกำหนดรายละเอียดเกี่ยวกับหัวเรื่องหลักของเอกสาร ซึ่งเป็นส่วนที่มีความสำคัญมากที่จะสามารถบอกได้ว่าเรื่องที่จะพูดถึงในเอกสารนั้น จะมีเนื้อหาหรือเกี่ยวกับเรื่องใด โดยผู้วิจัยได้กำหนดคำหรือวลีที่จะเป็นการบ่งบอกถึงส่วนของชื่อเรื่องของเอกสาร โดยประกอบด้วยคำดังนี้ ชื่อเรื่อง และ TITLE

2.4.3.1.2 บทคัดย่อ คือ การกำหนดรายละเอียดเกี่ยวกับเนื้อหาโดยสรุปของเอกสารนั้น ซึ่งเป็นส่วนที่มีความสำคัญและสามารถบอกได้ถึงใจความหลักโดยครอบคลุมใจความสำคัญของเอกสารเกือบทั้งหมด โดยส่วนของบทคัดย่อของเอกสารนั้น ผู้วิจัยได้กำหนดคำหรือวลีที่จะเป็นการบ่งบอกถึงส่วนของบทคัดย่อของเอกสาร โดยประกอบด้วยคำดังนี้ บทคัดย่อ และ ABSTRACT

2.4.3.1.3 คำนำ คือ การกำหนดรายละเอียดเกี่ยวกับส่วนของคำนำของเอกสารนั้น โดยเมื่อผู้อ่านทำการอ่านเอกสารนั้น ส่วนของคำนำก็จะสามารถบอกได้ว่าเรื่องที่ผู้อ่านกำลังอ่านนั้น มีมุมมองของเอกสารในภาพกว้างเป็นอย่างไร โดยผู้วิจัยเห็นว่าส่วนของคำนำนั้น มีความสำคัญพอสมควรในการที่จะสกัดคำที่จะเป็นตัวแทนของเอกสาร ผู้วิจัยได้กำหนดคำหรือวลีที่จะเป็นการบ่งบอกถึงส่วนของคำนำของเอกสาร โดยประกอบด้วยคำดังนี้ คำนำ บทนำ และ INTRODUCTION

2.4.3.1.4 ความเป็นมา โดยการกำหนดรายละเอียดเกี่ยวกับส่วนที่กล่าวถึงความเป็นมาของเอกสารนั้น ๆ โดยผู้วิจัยเห็นว่าส่วนของความเป็นมานั้น มีความสำคัญพอสมควรที่จะบอกถึงที่มาที่ไปของเอกสารนั้น ๆ ในส่วนของความเป็นมาของเอกสารนั้น ผู้วิจัยได้กำหนด คำหรือวลีที่จะเป็นการบ่งบอกถึงส่วนที่มาของการสร้างเอกสารนั้น ๆ ดังนี้ โดยประกอบด้วยคำดังนี้ ความเป็นมา

2.4.3.1.5 บทสรุป โดยการกำหนดรายละเอียดโดยการสรุปเนื้อหาในเอกสารนั้น ๆ โดยในส่วนของบทสรุปนั้น ผู้วิจัยเห็นว่าเป็นส่วนสำคัญมากอีกส่วนหนึ่งของเอกสาร เพราะเป็นส่วนที่จะสรุปรายละเอียดต่าง ๆ ทั้งเอกสารอย่างโดยย่อและครอบคลุมข้อมูลเกือบทั้งหมดของเอกสารนั้น ๆ โดยในส่วนบทสรุปของเอกสารนั้น ผู้วิจัยได้กำหนดคำหรือวลีที่

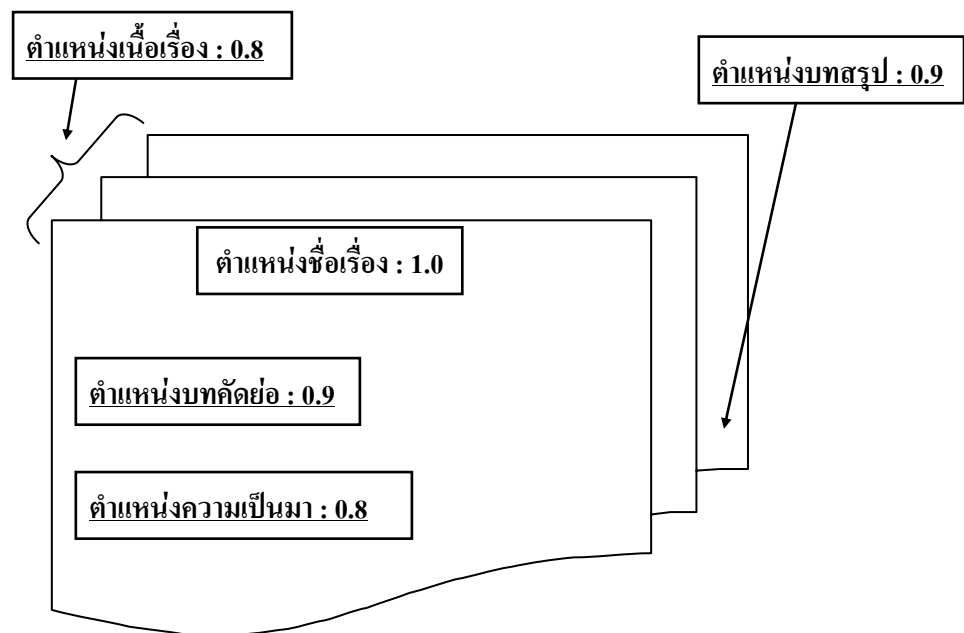
⁴ ดูเทคนิคหลักการเขียนเอกสาร ในบทที่ 2 หน้า 19-20

จะเป็นการบ่งบอกถึงส่วนการสรุปผลของเอกสารนั้น ดังนี้ โดยประกอบด้วยคำดังนี้ สรุปผลบทสรุป SUMMARY และ CONCLUSION

2.4.3.2 หลักการให้ค่าน้ำหนักคำหรือวลีในตำแหน่งเอกสาร

ความสัมพันธ์ของคำหรือวลีกับตำแหน่งในเอกสาร จากหลักไวยากรณ์ดังที่ได้กล่าวมานั้น ผู้วิจัยได้ทำการกำหนดค่าน้ำหนักระดับความสำคัญของคำในแต่ละตำแหน่งของเอกสาร ดังนี้

- ตำแหน่งของคำในส่วนของชื่อเรื่อง ที่มีค่าจะเป็นส่วนบอกถึงความหมายของเรื่องทั้งหมดในเอกสาร มีค่าน้ำหนักใจความสำคัญ คือ 1.0 ดังภาพประกอบ 4-4
- ตำแหน่งของคำในส่วนของบทคัดย่อ มีค่าน้ำหนักใจความสำคัญ คือ 0.9 ดังภาพประกอบ 4-4
- ตำแหน่งของคำในส่วนของความเป็นมา มีค่าน้ำหนักใจความสำคัญ คือ 0.8 ดังภาพประกอบ 4-4
- ตำแหน่งของคำในส่วนของบทนำ มีค่าน้ำหนักใจความสำคัญ คือ 0.8 ดังภาพประกอบ 4-4
- ตำแหน่งของคำในส่วนของบทสรุป มีค่าน้ำหนักใจความสำคัญ คือ 0.9 ดังภาพประกอบ 4-4



ภาพประกอบที่ 4-4 แสดงรูปแบบค่าน้ำหนักของคำในเอกสาร

2.4.3.3 การกำหนดค่าน้ำหนักตำแหน่งของคำหรือวลีในตำแหน่งเอกสาร

ผู้วิจัยได้ทำการแบ่งส่วนต่าง ๆ ของเอกสาร ออกเป็น 5 ส่วน และทำการกำหนดค่าน้ำหนักของแต่ละคำหรือวลีนั้นในรูปแบบของเมตริกซ์น้ำหนัก โดยการกำหนดระดับความสำคัญในแต่ละส่วนของเอกสารแตกต่างกันไป โดยมีกำหนดรายละเอียดค่าน้ำหนัก ดังนี้

Part Title of Document = {"TITLE", "ABSTRACT", "INTRODUCTION", "SUMMARY", "CONCLUSION", "ชื่อเรื่อง", "บทคัดย่อ", "ความเป็นมา", "บทนำ", "สรุปผล"}

Weight of Document Position = [1.0, 0.9, 0.75, 0.8, 0.9, 1.0, 0.9, 0.8, 0.75, 0.9]

ภาพประกอบที่ 4-5 รูปแบบการกำหนดค่าน้ำหนักตำแหน่งในเอกสาร

3. ไวยากรณ์ภาษาไทยเกี่ยวกับการละเว้นคำหยุดในภาษาไทย

จำนวนของคำหยุดไทยมีผลกระทบกับประสิทธิภาพของการประมวลผลในแง่ของความเร็วและเนื้อที่ที่ใช้ในการเก็บหน่วยความจำสำรอง แต่ถ้าการกำหนดคำหยุดไทยที่ไม่เหมาะสมก็ทำให้ดรชนีคำค้นหาหายไปจากระบบ ก็อาจมีผลกระทบกับผู้ใช้เช่นกัน โดยคำหยุดในแต่ละภาษานั้น เป็นกลุ่มคำที่มีการใช้มากเป็นพิเศษและไม่มีประโยชน์ในการสืบค้นเลย โดยคำเหล่านี้เกิดขึ้นเพราะหลักไวยากรณ์ภาษานั้นเอง และไม่มีคุณสมบัติเป็นคำสืบค้นหรือคำสำคัญในเอกสารได้เลย

ตัวอย่างของคำหยุดไทย ยืน ภู่วรรณ (2527) ดังภาคผนวก ค. อาทิเช่น “ที่, ว่า, ใน, จะ,” ซึ่งถ้าดูตามคุณลักษณะของหลักภาษาไทยบางก็อาจไม่มีความขัดแย้งที่จะกำหนดให้เป็นคำหยุด แต่คำดังกล่าวถ้าปรากฏในข้อความ “สัญญาจะซื้อจะขาย” ก็คงเป็นการยากลำบากในการที่จะกำหนดให้เป็นคำหยุด ดังนั้นควรมีการศึกษาหาคำหยุดที่เหมาะสมกับประเภทของเนื้อความด้วย

4. กฎต่าง ๆ ในการให้ค่าน้ำหนักคำหรือวลีในงานวิจัย

การกำหนดกฎให้กับโปรแกรมให้ค่าน้ำหนักนั้น เป็นสิ่งสำคัญในงานวิจัยนี้ โดยการสร้างกฎในการให้ค่าน้ำหนักของคำหรือวลีนั้น ได้อ้างอิงจากแนวคิดและหลักการที่ผู้วิจัยได้กล่าวมาในตอนต้นของบทนี้ โดยมีกฎต่าง ๆ ดังนี้

4.1 ถ้าพบชุดอักขระบรรทัดแรก ถือเป็นตำแหน่งชื่อเรื่อง

ถ้า [(อักขระ = 'n') และ (บรรทัดที่ 1 ของเอกสาร)]

ดังนั้นให้ค่าน้ำหนักตำแหน่งเอกสาร เท่ากับ '1'

4.2 ถ้าพบอักขระที่แสดงถึงย่อหน้า และอักขระก่อนหน้า ถือว่าเป็นการขึ้นต้นย่อหน้าใหม่ในแต่ละย่อหน้า

ถ้า [(อักขระก่อนหน้า = 'n') และ (อักขระปัจจุบัน = 't')]

ดังนั้น ถือเป็นการขึ้นย่อหน้าใหม่

4.3 ถ้าพบอักขระช่องว่าง และอักขระก่อนหน้าคือช่องว่าง ถือว่าเป็นการสิ้นสุดประโยคหรือการเกาะวรรค

ถ้า [(อักขระ = ' ') และ (อักขระ = ' ')]

ดังนั้น ถือเป็นการสิ้นสุดประโยค

4.4 เจอนไขการกำหนดค่าน้ำหนักในแต่ละช่วงของประโยค ทั้ง 5 ช่วง

ถ้าทำการอ่านอักขระจากเริ่มต้นประโยคจนเจอสัญลักษณ์สิ้นสุดประโยค ดังนั้นแบ่งประโยคออกเป็น 5 ช่วง เท่า ๆ กัน

ช่วงที่ 1	ช่วงที่ 2	ช่วงที่ 3	ช่วงที่ 4	ช่วงที่ 5
-----------	-----------	-----------	-----------	-----------

ดังนั้น

กลุ่มคำหรือวลีในช่วงที่ 1 และ 5 กำหนดค่าน้ำหนัก เท่ากับ "1"

กลุ่มคำหรือวลีในช่วงที่ 2 และ 4 กำหนดค่าน้ำหนัก เท่ากับ "0.5"

กลุ่มคำหรือวลีในช่วงที่ 3 กำหนดค่าน้ำหนัก เท่ากับ "0.75"

4.5 เงื่อนไขการกำหนดค่าน้ำหนักในแต่ละช่วงของย่อหน้า ทั้ง 5 ช่วง

ถ้าทำการอ่านอักขระจากเริ่มต้นย่อหน้าจนเจอสัญลักษณ์สิ้นสุดย่อหน้า ดังนั้นแบ่งย่อหน้าออกเป็น 5 ช่วง เท่า ๆ กัน

ช่วงที่ 1
ช่วงที่ 2
ช่วงที่ 3
ช่วงที่ 4
ช่วงที่ 5

กลุ่มบรรทัดในช่วงที่ 1 และ 5 กำหนดค่าน้ำหนัก เท่ากับ “1”

กลุ่มบรรทัดในช่วงที่ 2 และ 4 กำหนดค่าน้ำหนัก เท่ากับ “0.5”

กลุ่มบรรทัดในช่วงที่ 3 กำหนดค่าน้ำหนัก เท่ากับ “0.75”

4.6 ถ้าพบคำ “ TITLE ” หรือ “ ชื่อเรื่อง ” ถือเป็นตำแหน่งสำคัญที่สุดของตำแหน่งเอกสาร

ถ้า [(คำหรือวลี = ‘TITLE’) หรือ (คำหรือวลี = ‘ชื่อเรื่อง’)]

ดังนั้นให้ค่าน้ำหนักตำแหน่งเอกสาร เท่ากับ ‘1’

4.7 ถ้าพบคำ “ ABSTRACT ” หรือ “ บทคัดย่อ ” ถือเป็นใจความสำคัญของตำแหน่งในเอกสาร

ถ้า [(คำหรือวลี = ‘ABSTRACT’) หรือ (คำหรือวลี = ‘บทคัดย่อ’)]

ดังนั้นให้ค่าน้ำหนักตำแหน่งเอกสาร เท่ากับ ‘0.9’

4.8 ถ้าพบคำ “ SUMMARY ” หรือ “ CONCLUSION ” หรือ “ สรุปผล ” ถือเป็นใจความสำคัญของตำแหน่งในเอกสาร

ถ้า [(คำหรือวลี = ‘SUMMARY’) หรือ (คำหรือวลี = ‘CONCLUSION’) หรือ (คำหรือวลี = ‘สรุปผล’)]

ดังนั้นให้ค่าน้ำหนักตำแหน่งเอกสาร เท่ากับ ‘0.9’

4.9 ถ้าพบคำ “INTRODUCTION ” หรือ “บทนำ” ถือเป็นใจความสำคัญของตำแหน่งในเอกสาร

ถ้า [(คำหรือวลี = ‘SUMMARY’) หรือ (คำหรือวลี = ‘CONCLUSION’) หรือ (คำหรือวลี = ‘สรุปผล’)]

ดังนั้นให้ค่าน้ำหนักตำแหน่งเอกสาร เท่ากับ ‘0.75’

4.10 ถ้าประโยคที่อ่านเข้ามานั้นเป็นประโยคสั้น และมีจำนวนคำหรือวลีไม่เกิน 5 คำน้ำหนักที่กำหนดให้ คือค่ากลางของน้ำหนักของประโยค

ถ้า (จำนวนของคำหรือวลีในประโยคน้อยกว่า 5 จำนวน < 5)

ดังนั้น ค่าน้ำหนักตำแหน่งประโยคเท่ากับ 0.75

4.11 ถ้าจำนวนของบรรทัดในแต่ละย่อหน้ามีจำนวนไม่เกิน 5 บรรทัด ค่าน้ำหนักที่กำหนดให้คือค่ากลางของน้ำหนักในย่อหน้า

ถ้า (จำนวนบรรทัดในย่อหน้าน้อยกว่า 5 จำนวน < 5)

ดังนั้น ค่าน้ำหนักตำแหน่งย่อหน้าเท่ากับ 0.75