

บทที่ 3

การสรุปเอกสารภาษาไทยโดยใช้แฟร็กทัล

การสรุปใจความสำคัญจากเอกสารเป็นการสรุปประเด็นหรือเนื้อหาที่สำคัญของเอกสาร เพื่อให้ผู้อ่านได้รับทราบประเด็นหรือเนื้อหาในส่วนที่เป็นใจความสำคัญโดยไม่จำเป็นต้องอ่านเนื้อหาทั้งหมดจากเอกสารฉบับนั้น ทำให้ผู้อ่านสามารถรับทราบข้อมูลข่าวสารต่าง ๆ ได้อย่างครบถ้วนและรวดเร็ว

แนวทางการสรุปใจความสำคัญจากเอกสารแบ่งออกเป็น 3 แนวทาง คือ การสรุปใจความสำคัญโดยพิจารณาจากความถี่และตำแหน่งของคำต่าง ๆ ในเอกสาร การสรุปใจความสำคัญโดยใช้เทคนิคหรือวิธีการต่าง ๆ เช่น Singular Value Decomposition[10], Fuzzy Clustering Algorithm[11] เป็นต้น และการสรุปใจความสำคัญโดยพิจารณาจากความเข้ากันได้และความสอดคล้องกันของประโยคต่าง ๆ ในเอกสาร

วิธีการสรุปใจความสำคัญจากเอกสารแบ่งออกเป็น 2 วิธี คือ การคัดเลือกประโยคหรือย่อหน้าจากเอกสารต้นฉบับมาจัดทำเป็นข้อสรุป และการจัดทำข้อสรุปออกมาในรูปของบทคัดย่อ[2] โดยที่เทคนิคการสรุปแฟร็กทัลเป็นเทคนิคในการสรุปใจความสำคัญจากเอกสารต้นฉบับที่นำวิธีการทางสถิติมาใช้ในการคัดเลือกประโยคที่เป็นใจความสำคัญของเอกสารต้นฉบับเพื่อนำมาจัดทำเป็นข้อสรุปโดยพิจารณาจากโครงสร้างลำดับชั้น คำสำคัญ ชื่อเรื่อง ชื่อหัวเรื่อง และชื่อหัวเรื่องย่อยของเอกสารต้นฉบับ แต่เนื่องจากภาษาไทยเป็นภาษาที่มีโครงสร้างของประโยคที่ค่อนข้างซับซ้อนและไม่มีเครื่องหมายหรือสัญลักษณ์ที่ใช้แสดงถึงจุดสิ้นสุดของประโยคที่ชัดเจนเหมือนกับภาษาอื่น ๆ ทำให้การหาขอบเขตของประโยคภาษาไทยทำได้ยาก ดังนั้นผู้วิจัยจึงได้นำขั้นตอนวิธีวินนาวมาช่วยในการหาขอบเขตของประโยคภาษาไทยให้มีความถูกต้องมากยิ่งขึ้นโดยพิจารณาจากค่าน้ำหนักของคำต่าง ๆ ในชุดข้อมูลที่จัดทำขึ้นจากคลังข้อมูลออร์คิด และทำการจัดเตรียมเอกสารต้นฉบับให้อยู่ในรูปของ LaTeX เพื่อให้ง่ายต่อการวิเคราะห์โครงสร้างลำดับชั้นของเอกสารต้นฉบับแล้วนำเอกสารนี้ไปทำการตัดคำด้วยโปรแกรม Swath เพื่อหาคำสำคัญของเอกสารต้นฉบับ จากนั้นนำเอกสารต้นฉบับที่ถูกจัดทำให้อยู่ในรูปของ LaTeX ที่ผ่านการตัดคำเรียบร้อยแล้วและค่าน้ำหนักของคำต่าง ๆ ในชุดข้อมูลที่จัดทำขึ้นจากคลังข้อมูลออร์คิดมาใช้เป็นข้อมูลในการหาขอบเขตของประโยคภาษาไทยเพื่อนำมาใช้ในการจัดทำข้อสรุปของเอกสารต้นฉบับด้วยเทคนิคการสรุปแฟร็กทัลต่อไป

3.1 การจัดเตรียมข้อมูลก่อนที่จะทำการสรุปใจความสำคัญโดยใช้แฟร็กทัล

3.1.1 การจัดทำชุดข้อมูลฝึกหัดจากคลังข้อมูลออร์คิด

การจัดทำชุดข้อมูลฝึกหัดจากคลังข้อมูลออร์คิดเป็นการจัดทำชุดข้อมูลในการเรียนรู้เพื่อหาค่าน้ำหนักของคำแต่ละคำในคลังข้อมูลออร์คิดซึ่งจะถูกนำมาใช้เป็นข้อมูลในการหาขอบเขตของประโยคภาษาไทยในเอกสารต้นฉบับ โดยที่ผู้วิจัยจะทำการจัดทำชุดข้อมูลขึ้นมา 2 ประเภท คือ ชุดข้อมูลประเภทช่องว่างแบ่งประโยคและชุดข้อมูลประเภทช่องว่างไม่แบ่งประโยค ซึ่งในแต่ละชุดข้อมูลจะประกอบไปด้วยคำและชนิดของคำทั้งหมด 11 คำ ดังตัวอย่างต่อไปนี้

ตัวอย่างชุดข้อมูลประเภทช่องว่างแบ่งประโยค

#4

และสามารถนำไปใช้ประโยชน์มีมูลค่าเพิ่ม//

และ/JCRG

สามารถ/XVAM

นำไปใช้/VACT

ประโยชน์/NCMN

มี/VSTA

มูลค่าเพิ่ม/NCMN

//

#5

ส่งผลต่อการพัฒนาเศรษฐกิจโดยตรง//

ส่งผล/VACT

ต่อ/RPRE

การ/FIXN

พัฒนา/VACT

เศรษฐกิจ/NCMN

โดย/FIXV

ตรง/VATT

สามารถ@XVAM

นำไปใช้@VACT

ประโยชน์@NCMN

มี@VSTA

มูลค่าเพิ่ม@NCMN

<space>@SBS

ส่งผล@VACT

ต่อ@RPRE

การ@FIXN

พัฒนา@VACT

เศรษฐกิจ@NCMN

ตัวอย่างชุดข้อมูลประเภทช่องว่างไม่แบ่งประโยค

#4

เพื่อให้งานวิจัยและพัฒนาของท่านมีคุณค่า \\\

และคุณประโยชน์ต่อประเทศชาติยิ่งขึ้นไป//

เพื่อให้/JSBR

งานวิจัยและพัฒนา/NCMN

ของ/RPRE

ท่าน/PPRS

มี/VSTA

คุณค่า/NCMN

<space>/PUNC

และ/JCRG

คุณประโยชน์/NCMN

ต่อ/RPRE

ประเทศชาติ/NCMN

ยิ่งขึ้น/ADV

ไป/XVAE

งานวิจัยและพัฒนา@NCMN

ของ@RPRE

ท่าน@PPRS

มี@VSTA

คุณค่า@NCMN

<space>@NSBS

และ@JCRG

คุณประโยชน์@NCMN

ต่อ@RPRE

ประเทศชาติ@NCMN

ยิ่งขึ้น@ADV

3.1.2 การจัดทำเอกสารต้นฉบับให้อยู่ในรูปของ LaTeX

การจัดทำเอกสารต้นฉบับให้อยู่ในรูปของ LaTeX มีวัตถุประสงค์เพื่อช่วยให้การวิเคราะห์โครงสร้างลำดับชั้นของเอกสารต้นฉบับทำได้ง่ายขึ้น เนื่องจากเทคนิคการสรุปแฟร็กทัลเป็นเทคนิคที่นำโครงสร้างลำดับชั้นของเอกสารต้นฉบับมาใช้เป็นส่วนหนึ่งในการสรุปใจความสำคัญจากเอกสารต้นฉบับด้วย ดังตัวอย่างต่อไปนี้

ความร่วมมือกับองค์กรทั้งในและต่างประเทศ

เนคเทคได้ดำเนินความร่วมมือในการวิจัยและพัฒนาทางเทคโนโลยีกับต่างประเทศทั้งในระดับทวิภาคีและพหุภาคีสืบเนื่องมาโดยตลอด ทั้งนี้ โดยแบ่งเป็นกิจกรรมต่างๆ คือการดำเนินโครงการวิจัยร่วมกัน การแลกเปลี่ยนและถ่ายทอดเทคโนโลยีระหว่างกัน รวมทั้งการเข้าร่วมแสดงข้อคิดเห็นและสนับสนุนข้อมูลที่เกี่ยวข้องที่จะเป็นประโยชน์ต่อการวิจัยและพัฒนาทางเทคโนโลยีของไทย



\section{\thai ความร่วมมือกับองค์กรทั้งในและต่างประเทศ}

{\thai เนคเทคได้ดำเนินความร่วมมือในการวิจัยและพัฒนาทางเทคโนโลยีกับต่างประเทศทั้งในระดับทวิภาคีและพหุภาคีสืบเนื่องมาโดยตลอด ทั้งนี้ โดยแบ่งเป็นกิจกรรมต่างๆ คือการดำเนินโครงการวิจัยร่วมกัน การแลกเปลี่ยนและถ่ายทอดเทคโนโลยีระหว่างกัน รวมทั้งการเข้าร่วมแสดงข้อคิดเห็นและสนับสนุนข้อมูลที่เกี่ยวข้องที่จะเป็นประโยชน์ต่อการวิจัยและพัฒนาทางเทคโนโลยีของไทย}

3.1.3 การตัดคำจากเอกสารต้นฉบับที่ถูกจัดทำให้อยู่ในรูปของ LaTeX ด้วยโปรแกรม Swath

การตัดคำจากเอกสารต้นฉบับที่ถูกจัดทำให้อยู่ในรูปของ LaTeX ด้วยโปรแกรม Swath มีวัตถุประสงค์เพื่อหาคำสำคัญของเอกสารต้นฉบับ เนื่องจากเทคนิคการสรุปแฟร็กทัลเป็นเทคนิคที่นำคำสำคัญของเอกสารต้นฉบับมาใช้เป็นส่วนหนึ่งในการสรุปใจความสำคัญของเอกสารต้นฉบับด้วย ดังตัวอย่างต่อไปนี้

\section{\thai ความร่วมมือกับองค์กรทั้งในและต่างประเทศ}

{\thai เนกเทคได้ดำเนินความร่วมมือในการวิจัยและพัฒนาทางเทคโนโลยีกับต่างประเทศทั้งในระดับทวิภาคีและพหุภาคีสืบเนื่องมาโดยตลอด ทั้งนี้ โดยแบ่งเป็นกิจกรรมต่างๆ คือการดำเนินโครงการวิจัยร่วมกัน การแลกเปลี่ยนและถ่ายทอดเทคโนโลยีระหว่างกัน รวมทั้งการเข้าร่วมแสดงข้อคิดเห็นและสนับสนุนข้อมูลที่เกี่ยวข้องที่จะเป็นประโยชน์ต่อการวิจัยและพัฒนาทางเทคโนโลยีของไทย}



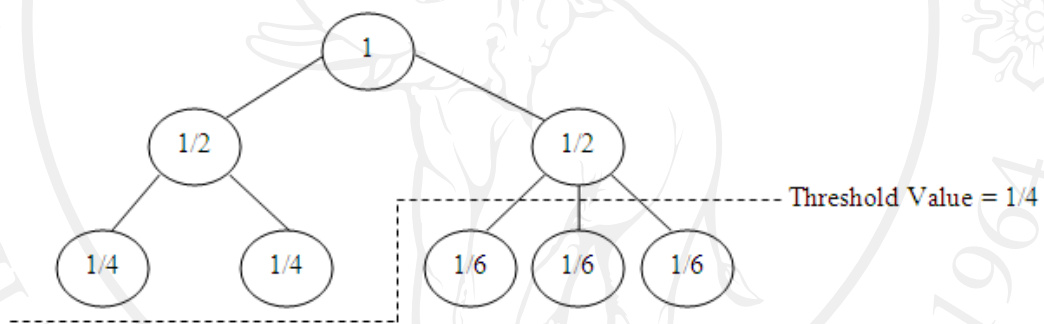
\section{\thai ความร่วมมือ@FIXNร่วมมือ@VACTกับ@RPREองค์กร@NCMN ทั้ง@DDBQใน@RPREและ@JCRGต่าง@PDMNประเทศ@NCMN}

{\thai เนก@NPRPเท@VACTค@NPRPได้@XVAEดำเนิน@VACT ความ@FIXNร่วม@ADVNMือ@NCMNใน@RPREการ@FIXNวิจัย@VACT และ@JCRGพัฒนา@VACTทาง@RPREเทคโนโลยี@NCMNกับ@RPRE ต่าง@PDMNประเทศ@NCMNทั้ง@DDBQใน@RPREระดับ@NCMN ท@NCMNวิภา@NPRPคิ@NPRPและ@JCRGพหุ@JCRGภาคี@NCMN สืบเนื่อง@VSTามา@XVAEโดย@JSBRตลอด@ADVNM ทั้ง@DDBQนี้@DDAC โดย@FIXVแบ่ง@VACTเป็น@VSTAกิจกรรม@NCMNต่างๆ@NPRP คือ@VSTA การ@FIXNดำเนิน@VACTโครงการวิจัย@NCMNร่วม@ADVNgัน@ADVNM การ@FIXNแลกเปลี่ยน@VACTและ@JCRGถ่ายทอด@VACTเทคโนโลยี@NCMN ระหว่าง@RPREกัน@ADVNM รวม@VACTทั้ง@DDBQการ@FIXNเข้า@VACT ร่วม@ADVNMแสดง@VACTข้อคิดเห็น@NCMNและ@JCRGสนับสนุน@VACT ข้อมูล@NCMNที่@PRELเกี่ยวข้อง@VSTAที่@PRELจะ@XVBMเป็น@VSTA ประโยชน์@NCMNต่อ@RPREการ@FIXNวิจัย@VACTและ@JCRG พัฒนา@VACTทาง@RPREเทคโนโลยี@NCMNของ@RPREไทย@NPRP}

3.2 เทคนิคการสรุปแฟร็กทัล

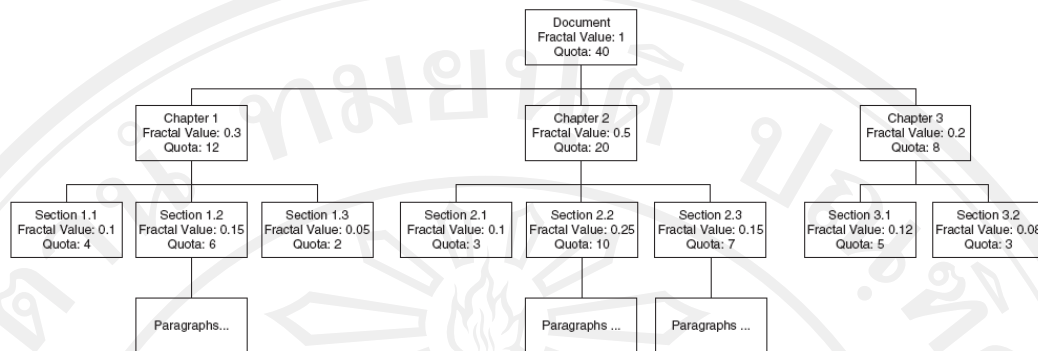
เทคนิคการสรุปแฟร็กทัลเป็นเทคนิคในการสรุปใจความสำคัญจากเอกสารต้นฉบับที่นำวิธีการทางสถิติมาใช้ในการคัดเลือกข้อมูลที่เป็นใจความสำคัญของเอกสารต้นฉบับเพื่อนำมาจัดทำเป็นข้อสรุปโดยพิจารณาจากโครงสร้างลำดับชั้น คำสำคัญ ชื่อเรื่อง ชื่อหัวเรื่อง และชื่อหัวเรื่องย่อย

ของเอกสารต้นฉบับ โดยที่ แฟร็กทัล คือ Object ที่ใช้แทนชื่อเรื่อง ชื่อหัวเรื่อง และชื่อหัวเรื่องย่อยของเอกสารต้นฉบับซึ่งมีระดับความสำคัญที่แตกต่างกันโดยพิจารณาจากโครงสร้างลำดับชั้นของเอกสารต้นฉบับ ซึ่งระดับความสำคัญของ Object แต่ละตัวจะถูกแทนด้วยค่าแฟร็กทัล (Fractal Value) โดยค่าแฟร็กทัลของ Object ตัวบนสุดจะมีค่าเท่ากับ 1 และค่าแฟร็กทัลของ Object ตัวอื่น ๆ จะมีค่าเท่ากับค่าที่คำนวณได้จากสมการ $\text{Fractal Value} = A / B$ โดยที่ A คือ ค่าแฟร็กทัลของ Object ที่เป็นโหนดแม่ (Parent Node) ของ Object ตัวนั้น และ B คือ จำนวน Object ที่อยู่ภายใต้ Object ที่เป็นโหนดแม่ของ Object ตัวนั้น โดยมีค่าขีดแบ่ง (Threshold Value) เป็นตัวแปรที่ใช้ในการควบคุมจำนวนข้อมูลที่จะถูกคัดเลือกออกมาจัดทำเป็นข้อสรุป ซึ่งถ้า Object ใดมีค่าแฟร็กทัลน้อยกว่าค่าขีดแบ่ง Object นั้นก็จะไม่ถูกคัดเลือกออกมาจัดทำเป็นข้อสรุป ดังรูป 3.1[13]



รูป 3.1 ตัวอย่างของแฟร็กทัลซึ่งมีค่าแฟร็กทัลที่แตกต่างกันตามโครงสร้างลำดับชั้น โดยมีค่าขีดแบ่งเป็นตัวแปรที่ใช้ในการควบคุมจำนวนข้อมูลที่จะถูกคัดเลือกออกมาจัดทำเป็นข้อสรุป

สำหรับการสรุปใจความสำคัญด้วยเทคนิคการสรุปแฟร็กทัลนั้น เอกสารต้นฉบับจะถูกแบ่งออกเป็นระดับต่าง ๆ ในลักษณะของโครงสร้างลำดับชั้นเริ่มตั้งแต่ระดับ Document จนถึงระดับ Paragraph โดยในแต่ละระดับจะมีค่าแฟร็กทัลที่แตกต่างกันออกไปซึ่งจะส่งผลต่อจำนวนประโยคที่จะถูกคัดเลือกออกมาจัดทำเป็นข้อสรุปของแต่ละระดับที่คำนวณได้จากสมการ $\text{Sentence} = [(A / B) \times C]$ โดยที่ A คือ ค่าแฟร็กทัลของระดับนั้น B คือ ค่าแฟร็กทัลของระดับที่เป็นโหนดแม่ของระดับนั้น และ C คือ จำนวนประโยคที่จะถูกคัดเลือกออกมาจัดทำเป็นข้อสรุปของระดับที่เป็นโหนดแม่ของระดับนั้น ดังรูป 3.2[13]



รูป 3.2 ตัวอย่างโครงสร้างลำดับชั้นของเอกสารต้นฉบับ[13]

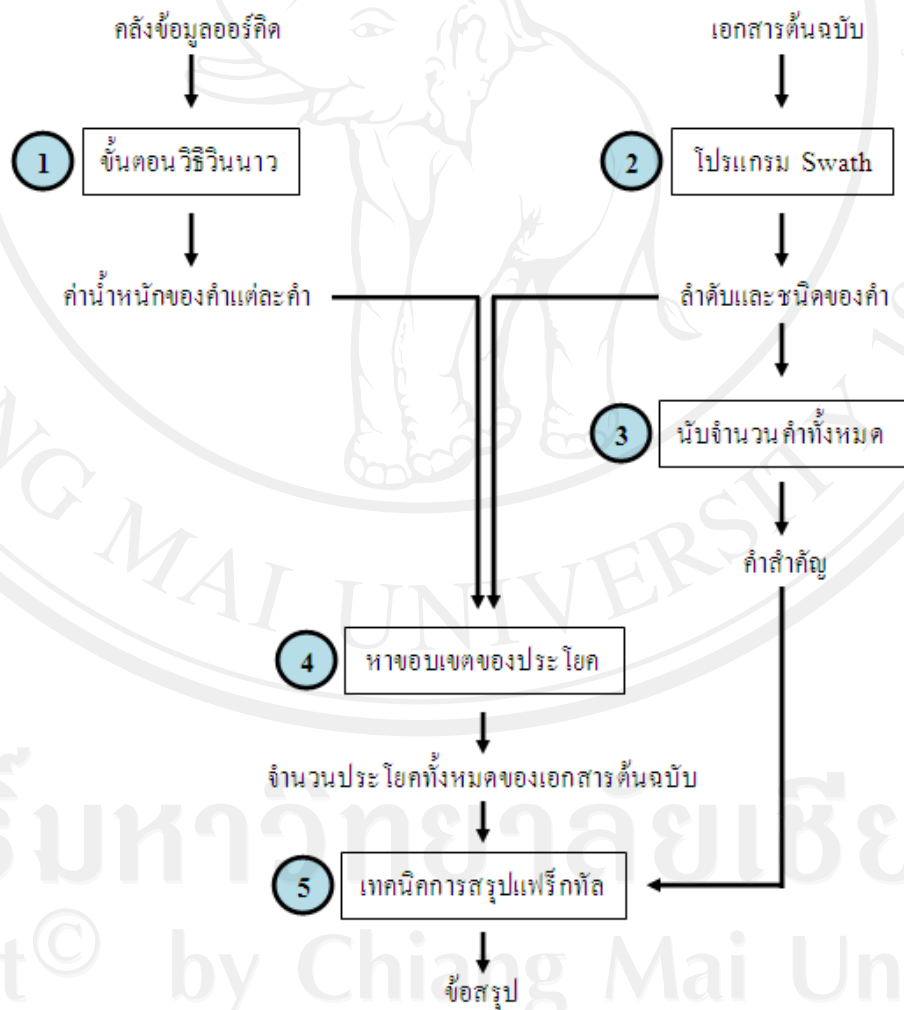
การสรุปใจความสำคัญด้วยเทคนิคการสรุปแฟร็กทัลมีขั้นตอนดังต่อไปนี้

- (1) กำหนดอัตราส่วนการบีบอัด (Compression Ratio) เพื่อใช้ในการควบคุมจำนวนประโยคทั้งหมดของข้อสรุป
- (2) กำหนดค่าขีดแบ่งเพื่อใช้ในการควบคุมจำนวนประโยคที่จะถูกคัดเลือกออกมาจัดทำเป็นข้อสรุป
- (3) คำนวณจำนวนประโยคทั้งหมดของข้อสรุปโดยนำจำนวนประโยคทั้งหมดของเอกสารต้นฉบับคูณด้วยอัตราส่วนการบีบอัด
- (4) กำหนดให้คำว่า “โนนด” ในที่นี้หมายถึง ชื่อเรื่อง ชื่อหัวเรื่อง หรือชื่อหัวเรื่องย่อยของเอกสารต้นฉบับ
- (5) คำนวณค่าคะแนน TF-IDF ให้กับประโยคที่มีคำสำคัญของเอกสารต้นฉบับปรากฏอยู่จากสมการ $TF-IDF = A \times \log_2[(B \times C) / D]$ โดยที่ A คือ ความถี่ของคำสำคัญที่ปรากฏอยู่ในประโยค B คือ จำนวนโนนดทั้งหมดของเอกสารต้นฉบับ C คือ ความยาวของคำสำคัญ และ D คือ จำนวนโนนดที่มีคำสำคัญปรากฏอยู่
- (6) คำนวณค่าคะแนน Location ให้กับประโยคทุกประโยคโดยพิจารณาจากตำแหน่งของประโยคนั้นในเอกสารต้นฉบับจากสมการ $Location = [1 / ((\min(A, B)) \times C)]$ โดยที่ A คือ ระยะห่างระหว่างโนนดที่ประโยคนั้นปรากฏอยู่กับโนนดที่อยู่ด้านซ้ายสุด และโนนดทั้งสองอยู่ภายใต้โนนดแม่ตัวเดียวกัน B คือ ระยะห่างระหว่างโนนดที่ประโยคนั้นปรากฏอยู่กับโนนดที่อยู่ด้านขวาสุด และโนนดทั้งสองอยู่ภายใต้โนนดแม่ตัวเดียวกัน $\min(A, B)$ คือ ค่า A ถ้า $A < B$ หรือ ค่า B ถ้า $B < A$ และ C คือ จำนวนประโยคทั้งหมดของโนนดที่ประโยคนั้นปรากฏอยู่

- (7) คำนวณค่าคะแนน Heading ให้กับประโยคที่มีคำที่ปรากฏในชื่อเรื่อง ชื่อหัวเรื่อง หรือชื่อหัวเรื่องย่อยของเอกสารต้นฉบับปรากฏอยู่ โดยนำค่าคะแนน TF-IDF ของ ประโยคนั้นหารด้วยจำนวนโหนดทั้งหมดที่อยู่ในระดับเดียวกับโหนดที่ประโยคนั้น ปรากฏอยู่
- (8) คำนวณค่าคะแนน Sentence ให้กับประโยคทุกประโยคในเอกสารต้นฉบับโดยนำค่า คะแนน TF-IDF ค่าคะแนน Location และค่าคะแนน Heading ของประโยคนั้นมา บวกกัน
- (9) ออกแบบโครงสร้างลำดับชั้นของเอกสารต้นฉบับตามขั้นตอนดังต่อไปนี้
- (9.1) กำหนดให้ชื่อเรื่องของเอกสารต้นฉบับเป็นโหนดรากซึ่งมีค่าแฟร็กทัล เท่ากับ 1 และมีจำนวนประโยคที่จะถูกคัดเลือกออกมาจัดทำเป็นข้อสรุป เท่ากับจำนวนประโยคทั้งหมดของข้อสรุปที่คำนวณได้จากข้อ (3)
- (9.2) กำหนดให้ชื่อหัวเรื่องของเอกสารต้นฉบับเป็นโหนดลูกของโหนดราก และ คำนวณค่าแฟร็กทัลของโหนดลูกแต่ละโหนดจากสมการ $FV = [A \times (B / C)]$ โดยที่ A คือ ค่าแฟร็กทัลของโหนดราก B คือ ค่าคะแนน Sentence ของประโยคทั้งหมดของโหนดลูก และ C คือ ค่าคะแนน Sentence ของ ประโยคทั้งหมดของโหนดลูกทุกโหนดที่อยู่ภายใต้โหนดรากตัวเดียวกัน
- (9.3) คำนวณจำนวนประโยคที่จะถูกคัดเลือกออกมาจัดทำเป็นข้อสรุปของโหนด ลูกแต่ละโหนดจากสมการ $Sentence = [(A / B) \times C]$ โดยที่ A คือ ค่า แฟร็กทัลของโหนดลูก B คือ ค่าแฟร็กทัลของโหนดราก และ C คือ จำนวน ประโยคที่จะถูกคัดเลือกออกมาจัดทำเป็นข้อสรุปของโหนดราก
- (9.4) ตรวจสอบจำนวนประโยคที่จะถูกคัดเลือกออกมาจัดทำเป็นข้อสรุปของ โหนดลูกแต่ละโหนดที่คำนวณได้ตามเงื่อนไขต่อไปนี้
- (9.4.1) ถ้าจำนวนประโยคที่จะถูกคัดเลือกออกมาจัดทำเป็นข้อสรุปของ โหนดลูกมีค่าน้อยกว่าหรือเท่ากับค่าขีดแบ่ง ให้เลือกประโยคที่มี ค่าคะแนน Sentence จากมากไปหาน้อยตามจำนวนประโยคที่ จะถูกคัดเลือกออกมาจัดทำเป็นข้อสรุปของโหนดนั้นมาจัดทำ เป็นข้อสรุป

(9.4.2) ถ้าจำนวนประโยคที่จะถูกคัดเลือกออกมาจัดทำเป็นข้อสรุปของ โหนดลูกมีค่ามากกว่าค่าขีดแบ่ง จะกำหนดให้โหนดลูกตัวนี้เป็น โหนดรากที่มีค่าแฟร็กทัลเท่ากับค่าที่คำนวณได้จากข้อ (9.2) และมีจำนวนประโยคที่จะถูกคัดเลือกออกมาจัดทำเป็นข้อสรุป เท่ากับค่าที่คำนวณได้จากข้อ (9.3) จากนั้นให้ทำซ้ำข้อ (9.2) – (9.4) โดยกำหนดให้ชื่อหัวเรื่องย่อยของเอกสารต้นฉบับเป็น โหนดลูกของโหนดราก

3.3 การออกแบบวิธีการสรุปเอกสารภาษาไทยที่มีการจัดโครงสร้างโดยใช้เทคนิคการสรุปแฟร็กทัล



รูป 3.3 วิธีการสรุปเอกสารภาษาไทยที่มีการจัดโครงสร้างโดยใช้เทคนิคการสรุปแฟร็กทัล

3.3.1 ขั้นตอนที่ 1 การหาค่าน้ำหนักของค่าแต่ละค่าในคลังข้อมูลออร์คิด

- (1) ตรวจสอบค่าในชุดข้อมูลฝึกหัดแต่ละชุดทีละค่าตามเงื่อนไขต่อไปนี้
 - (1.1) ถ้าเคยตรวจสอบค่านี้แล้ว ค่าน้ำหนักของค่านี้จะมีค่าเท่ากับค่าเดิม
 - (1.2) ถ้าไม่เคยตรวจสอบค่านี้มาก่อน ค่าน้ำหนักเริ่มต้นของค่านี้จะเป็นไปตามเงื่อนไขต่อไปนี้
 - (1.2.1) ถ้าค่านี้อยู่ในชุดข้อมูลประเภทช่องว่างแบ่งประ โยค ค่าน้ำหนักเริ่มต้นของค่านี้จะมีค่าเท่ากับ 1
 - (1.2.2) ถ้าค่านี้อยู่ในชุดข้อมูลประเภทช่องว่างไม่แบ่งประ โยค ค่าน้ำหนักเริ่มต้นของค่านี้จะมีค่าเท่ากับ -1
- (2) เมื่อตรวจสอบค่าในชุดข้อมูลฝึกหัดแต่ละชุดทีละค่าจนครบแล้ว ให้ทำการรวมค่าน้ำหนักของค่าแต่ละค่าโดยผลรวมของค่าน้ำหนักที่ได้จะนำมาทำนายค่าตอบตามเงื่อนไขต่อไปนี้
 - (2.1) ถ้าผลรวมของค่าน้ำหนักมีค่ามากกว่าหรือเท่ากับจำนวนค่าทั้งหมดภายในชุดข้อมูล ชุดข้อมูลนั้นจะถูกทำนายว่าเป็นชุดข้อมูลประเภทช่องว่างแบ่งประ โยค
 - (2.2) ถ้าผลรวมของค่าน้ำหนักมีค่าน้อยกว่าจำนวนค่าทั้งหมดภายในชุดข้อมูล ชุดข้อมูลนั้นจะถูกทำนายว่าเป็นชุดข้อมูลประเภทช่องว่างไม่แบ่งประ โยค
- (3) คำตอบที่ทำนายออกมาจะส่งผลต่อค่าน้ำหนักของค่าแต่ละค่าในชุดข้อมูลนั้นตามเงื่อนไขต่อไปนี้
 - (3.1) ในกรณีที่ทำนายคำตอบผิด ให้ทำการปรับค่าน้ำหนักของค่าแต่ละค่าตามเงื่อนไขต่อไปนี้
 - (3.1.1) ถ้าค่านี้อยู่ในชุดข้อมูลประเภทเดียวกับคำตอบที่ทำนายออกมา ค่าน้ำหนักของค่านี้นั้นจะลดลงครึ่งหนึ่ง
 - (3.1.2) ถ้าค่านี้อยู่ในชุดข้อมูลประเภทเดียวกับคำตอบที่ทำนายออกมา ค่าน้ำหนักของค่านี้นั้นจะเพิ่มขึ้นเป็น 1.5 เท่าของค่าเดิม
 - (3.2) ในกรณีที่ทำนายคำตอบถูก ให้ทำการปรับค่าน้ำหนักของค่าแต่ละค่าตามเงื่อนไขต่อไปนี้
 - (3.2.1) ถ้าค่านี้อยู่ในชุดข้อมูลประเภทเดียวกับคำตอบที่ทำนายออกมา ค่าน้ำหนักของค่านี้นั้นจะมีค่าเท่าเดิม

(3.2.2) ถ้าค่านั้น ไม่ได้อยู่ในชุดข้อมูลประเภทเดียวกับคำตอบที่ทำนาย
ออกมา ค่าน้ำหนักของค่านั้นจะลดลงครึ่งหนึ่ง

(4) ทำซ้ำข้อ (1) – (3) จนกระทั่งได้ตรวจสอบชุดข้อมูลทั้งหมดเรียบร้อยแล้ว

3.3.2 ขั้นตอนที่ 2 การหาลำดับและชนิดของค่าแต่ละค่าในเอกสารต้นฉบับ

นำเอกสารต้นฉบับที่ถูกจัดทำขึ้นให้อยู่ในรูปของ LaTeX มาทำการตัดคำด้วย
โปรแกรม Swath

3.3.3 ขั้นตอนที่ 3 การหาคำสำคัญของเอกสารต้นฉบับ

นับจำนวนคำทั้งหมดจากเอกสารต้นฉบับที่ถูกตัดคำเรียบร้อยแล้ว

3.3.4 ขั้นตอนที่ 4 การหาจำนวนประโยคทั้งหมดของเอกสารต้นฉบับ

(1) หาขอบเขตของประโยคภาษาไทยในเอกสารต้นฉบับด้วยการตรวจสอบช่องว่างที่ละ
ช่องว่างว่าเป็นช่องว่างแบ่งประโยคหรือช่องว่างไม่แบ่งประโยคจากการรวมค่าน้ำหนัก
ของคำทุกคำที่อยู่รอบ ๆ ช่องว่างนั้น โดยผลรวมของค่าน้ำหนักที่ได้จะนำมาพิจารณา
ตามเงื่อนไขต่อไปนี้

(1.1) ถ้าผลรวมของค่าน้ำหนักมีค่ามากกว่าหรือเท่ากับจำนวนคำที่อยู่รอบ ๆ
ช่องว่าง ช่องว่างนั้นจะถูกกำหนดให้เป็นช่องว่างแบ่งประโยค

(1.2) ถ้าผลรวมของค่าน้ำหนักมีค่าน้อยกว่าจำนวนคำที่อยู่รอบ ๆ ช่องว่าง ช่องว่าง
นั้นจะถูกกำหนดให้เป็นช่องว่างไม่แบ่งประโยค

(2) ทำซ้ำข้อ (1) จนกระทั่งได้ตรวจสอบช่องว่างทั้งหมดเรียบร้อยแล้ว

(3) นับจำนวนประโยคทั้งหมดของเอกสารต้นฉบับ

3.3.5 ขั้นตอนที่ 5 การจัดทำข้อสรุป

(1) แบ่งการทดลองออกเป็น 4 การทดลองเพื่อเปรียบเทียบความเที่ยงของข้อสรุปในแต่ละ
การทดลองซึ่งมีอัตราส่วนการบีบอัดและค่าขีดแบ่งที่แตกต่างกันดังต่อไปนี้

(1.1) การทดลอง A มีอัตราส่วนการบีบอัด = 20% และค่าขีดแบ่ง = 3

(1.2) การทดลอง B มีอัตราส่วนการบีบอัด = 20% และค่าขีดแบ่ง = 5

(1.3) การทดลอง C มีอัตราส่วนการบีบอัด = 25% และค่าขีดแบ่ง = 3

(1.4) การทดลอง D มีอัตราส่วนการบีบอัด = 25% และค่าขีดแบ่ง = 5

(2) คำนวณจำนวนประโยคทั้งหมดของข้อสรุปในแต่ละการทดลองโดยนำจำนวนประโยค
ทั้งหมดของเอกสารต้นฉบับคูณด้วยอัตราส่วนการบีบอัด

- (3) คำนวณค่าคะแนน TF-IDF ให้กับประโยคที่มีคำสำคัญของเอกสารต้นฉบับปรากฏอยู่ จากสมการ $TF-IDF = A \times \log_2[(B \times C) / D]$ โดยที่ A คือ ความถี่ของคำสำคัญที่ปรากฏอยู่ในประโยค B คือ จำนวนโหนดทั้งหมดของเอกสารต้นฉบับ C คือ ความยาวของคำสำคัญ และ D คือ จำนวนโหนดที่มีคำสำคัญปรากฏอยู่
- (4) คำนวณค่าคะแนน Location ให้กับประโยคทุกประโยคโดยพิจารณาจากตำแหน่งของประโยคนั้นในเอกสารต้นฉบับจากสมการ $Location = [1 / ((\min(A, B)) \times C)]$ โดยที่ A คือ ระยะห่างระหว่างโหนดที่ประโยคนั้นปรากฏอยู่กับโหนดที่อยู่ด้านซ้ายสุด และโหนดทั้งสองอยู่ภายใต้โหนดแม่ตัวเดียวกัน B คือ ระยะห่างระหว่างโหนดที่ประโยคนั้นปรากฏอยู่กับโหนดที่อยู่ด้านขวาสุด และโหนดทั้งสองอยู่ภายใต้โหนดแม่ตัวเดียวกัน $\min(A, B)$ คือ ค่า A ถ้า $A < B$ หรือ ค่า B ถ้า $B < A$ และ C คือ จำนวนประโยคทั้งหมดของโหนดที่ประโยคนั้นปรากฏอยู่
- (5) คำนวณค่าคะแนน Heading ให้กับประโยคที่มีคำที่ปรากฏในชื่อเรื่อง ชื่อหัวเรื่อง หรือชื่อหัวเรื่องย่อยของเอกสารต้นฉบับปรากฏอยู่ โดยนำค่าคะแนน TF-IDF ของประโยคนั้นหารด้วยจำนวนโหนดทั้งหมดที่อยู่ในระดับเดียวกับโหนดที่ประโยคนั้นปรากฏอยู่
- (6) คำนวณค่าคะแนน Sentence ให้กับประโยคทุกประโยคในเอกสารต้นฉบับโดยนำค่าคะแนน TF-IDF ค่าคะแนน Location และค่าคะแนน Heading ของประโยคนั้นมาบวกกัน
- (7) ออกแบบโครงสร้างลำดับชั้นของเอกสารต้นฉบับตามขั้นตอนดังต่อไปนี้
- (7.1) กำหนดให้ชื่อเรื่องของเอกสารต้นฉบับเป็นโหนดรากซึ่งมีค่าแฟร็กทัลเท่ากับ 1 และมีจำนวนประโยคที่จะถูกคัดเลือกออกมาจัดทำเป็นข้อสรุปเท่ากับจำนวนประโยคทั้งหมดของข้อสรุป
- (7.2) กำหนดให้ชื่อหัวเรื่องของเอกสารต้นฉบับเป็นโหนดลูกของโหนดราก และคำนวณค่าแฟร็กทัลของโหนดลูกแต่ละโหนดจากสมการ $FV = [A \times (B / C)]$ โดยที่ A คือ ค่าแฟร็กทัลของโหนดราก B คือ ค่าคะแนน Sentence ของประโยคทั้งหมดของโหนดลูก และ C คือ ค่าคะแนน Sentence ของประโยคทั้งหมดของโหนดลูกทุกโหนดที่อยู่ภายใต้โหนดรากตัวเดียวกัน

(7.3) คำนวณจำนวนประโยคที่จะถูกคัดเลือกออกมาจัดทำเป็นข้อสรุปของโหนดลูกแต่ละโหนดจากสมการ $\text{Sentence} = [(A / B) \times C]$ โดยที่ A คือ ค่าแฟร็กทัลของโหนดลูก B คือ ค่าแฟร็กทัลของโหนดราก และ C คือ จำนวนประโยคที่จะถูกคัดเลือกออกมาจัดทำเป็นข้อสรุปของโหนดราก

(7.4) ตรวจสอบจำนวนประโยคที่จะถูกคัดเลือกออกมาจัดทำเป็นข้อสรุปของโหนดลูกแต่ละโหนดที่คำนวณได้ตามเงื่อนไขต่อไปนี้

(7.4.1) ถ้าจำนวนประโยคที่จะถูกคัดเลือกออกมาจัดทำเป็นข้อสรุปของโหนดลูกมีค่าน้อยกว่าหรือเท่ากับค่าขีดแบ่ง ให้เลือกประโยคที่มีค่าคะแนน Sentence จากมากไปหาน้อยตามจำนวนประโยคที่จะถูกคัดเลือกออกมาจัดทำเป็นข้อสรุปของโหนดนั้นมาจัดทำเป็นข้อสรุป

(7.4.2) ถ้าจำนวนประโยคที่จะถูกคัดเลือกออกมาจัดทำเป็นข้อสรุปของโหนดลูกมีค่ามากกว่าค่าขีดแบ่ง จะกำหนดให้โหนดลูกตัวนี้เป็นโหนดรากที่มีค่าแฟร็กทัลเท่ากับค่าที่คำนวณได้จากข้อ (7.2) และมีจำนวนประโยคที่จะถูกคัดเลือกออกมาจัดทำเป็นข้อสรุปเท่ากับค่าที่คำนวณได้จากข้อ (7.3) จากนั้นให้ทำซ้ำข้อ (7.2) – (7.4) โดยกำหนดให้ชื่อหัวเรื่องย่อยของเอกสารต้นฉบับเป็นโหนดลูกของโหนดราก

3.4 ตัวอย่างการสรุปเอกสารภาษาไทยที่มีการจัดโครงสร้างโดยใช้เทคนิคการสรุปแฟร็กทัล

3.4.1 การจัดทำเอกสารต้นฉบับให้อยู่ในรูปของ LaTeX

เนคเทคได้ดำเนินความร่วมมือในการวิจัยและพัฒนาทางเทคโนโลยีกับต่างประเทศทั้งในระดับทวิภาคีและพหุภาคีสืบเนื่องมาโดยตลอด ทั้งนี้ โดยแบ่งเป็นกิจกรรมต่างๆ คือการดำเนินโครงการวิจัยร่วมกัน การแลกเปลี่ยนและถ่ายทอดเทคโนโลยีระหว่างกัน รวมทั้งการเข้าร่วมแสดงข้อคิดเห็นและสนับสนุนข้อมูลที่เกี่ยวข้องที่จะเป็นประโยชน์ต่อการวิจัยและพัฒนาทางเทคโนโลยีของไทย



`{\thai เนคเทคได้ดำเนินความร่วมมือในการวิจัยและพัฒนาทางเทคโนโลยีกับต่างประเทศทั้งในระดับทวิภาคีและพหุภาคีสืบเนื่องมาโดยตลอด ทั้งนี้ โดยแบ่งเป็นกิจกรรมต่างๆ คือการดำเนินโครงการวิจัยร่วมกัน การแลกเปลี่ยนและถ่ายทอดเทคโนโลยีระหว่างกัน รวมทั้งการเข้าร่วมแสดงข้อคิดเห็นและสนับสนุนข้อมูลที่เกี่ยวข้องที่จะเป็นประโยชน์ต่อการวิจัยและพัฒนาทางเทคโนโลยีของไทย}`

3.4.2 การตัดคำจากเอกสารต้นฉบับที่ถูกจัดทำให้อยู่ในรูปของ LaTeX ด้วยโปรแกรม Swath

{\thai เนกเทคได้ดำเนินความร่วมมือในการวิจัยและพัฒนาทางเทคโนโลยีกับต่างประเทศทั้งในระดับทวิภาคีและพหุภาคีสืบเนื่องมาโดยตลอด ทั้งนี้ โดยแบ่งเป็นกิจกรรมต่างๆ คือการดำเนินโครงการวิจัยร่วมกัน การแลกเปลี่ยนและถ่ายทอดเทคโนโลยีระหว่างกัน รวมทั้งการเข้าร่วมแสดงข้อคิดเห็นและสนับสนุนข้อมูลที่เกี่ยวข้องที่จะเป็นประโยชน์ต่อการวิจัยและพัฒนาทางเทคโนโลยีของไทย}



```
{\thai เนก@NPRPเท@VACTค@NPRPได้@XVAEดำเนิน@VACT
ความ@FIXNร่วม@ADVNMือ@NCMNใน@RPREการ@FIXNวิจัย@VACT
และ@JCRGพัฒนา@VACTทาง@RPREเทคโนโลยี@NCMNกับ@RPRE
ต่าง@PDMNประเทศ@NCMNทั้ง@DDBQใน@RPREระดับ@NCMNท@NCMN
วิชา@NPRPที่@NPRPและ@JCRGพหุ@JCRGภาคี@NCMNสืบเนื่อง@VSTA
มา@XVAEโดย@JSBRตลอด@ADVN ทั้ง@DDBQนี้@DDAC โดย@FIXV
แบ่ง@VACTเป็น@VSTAกิจกรรม@NCMNต่างๆ@NPRP คือ@VSTAการ@FIXN
ดำเนิน@VACTโครงการวิจัย@NCMNร่วม@ADVNgัน@ADVN การ@FIXN
แลกเปลี่ยน@VACTและ@JCRGถ่ายทอด@VACTเทคโนโลยี@NCMN
ระหว่าง@RPREกัน@ADVN รวม@VACTทั้ง@DDBQการ@FIXNเข้า@VACT
ร่วม@ADVNแสดง@VACTข้อคิดเห็น@NCMNและ@JCRGสนับสนุน@VACT
ข้อมูล@NCMNที่@PRELเกี่ยวข้อง@VSTAที่@PRELจะ@XVBMเป็น@VSTA
ประโยชน์@NCMNต่อ@RPREการ@FIXNวิจัย@VACTและ@JCRG
พัฒนา@VACTทาง@RPREเทคโนโลยี@NCMNของ@RPREไทย@NPRP}
```

3.4.3 การหาขอบเขตของประโยคภาษาไทย

เนค@NPRPเท@VACTค@NPRPได้@XVAEดำเนิน@VACTความ@FIXN
 ร่วม@ADVNmื่อ@NCMNใน@RPREการ@FIXNวิจัย@VACTและ@JCRG
 พัฒนา@VACTทาง@RPREเทคโนโลยี@NCMNกับ@RPREต่าง@PDMN
 ประเทศ@NCMNทั้ง@DDBQใน@RPREระดับ@NCMNท@NCMNวิภา@NPRP
 กิจ@NPRPและ@JCRGพหุ@JCRGภาคี@NCMNสืบเนื่อง@VSTAมา@XVAE
 โดย@JSBRตลอด@ADVN <1> ทั้ง@DDBQนี้@DDAC <2> โดย@FIXV
 แบ่ง@VACTเป็น@VSTAกิจกรรม@NCMNต่างๆ@NPRP <3> คือ@VSTA
 การ@FIXNดำเนิน@VACTโครงการวิจัย@NCMNร่วม@ADVNgัน@ADVN <4>
 การ@FIXNแลกเปลี่ยน@VACTและ@JCRGถ่ายทอด@VACTเทคโนโลยี@NCMN
 ระหว่าง@RPREกัน@ADVN <5> รวม@VACTทั้ง@DDBQการ@FIXN
 เข้า@VACTร่วม@ADVNแสดง@VACTข้อคิดเห็น@NCMNและ@JCRG
 สนับสนุน@VACTข้อมูล@NCMNที่@PRELเกี่ยวข้อง@VSTAที่@PRELจะ@XVBM
 เป็น@VSTAประโยชน์@NCMNต่อ@RPREการ@FIXNวิจัย@VACTและ@JCRG
 พัฒนา@VACTทาง@RPREเทคโนโลยี@NCMNของ@RPREไทย@NPRP



เนค@NPRPเท@VACTค@NPRPได้@XVAEดำเนิน@VACTความ@FIXN
 ร่วม@ADVNmื่อ@NCMNใน@RPREการ@FIXNวิจัย@VACTและ@JCRG
 พัฒนา@VACTทาง@RPREเทคโนโลยี@NCMNกับ@RPREต่าง@PDMN
 ประเทศ@NCMNทั้ง@DDBQใน@RPREระดับ@NCMNท@NCMNวิภา@NPRP
 กิจ@NPRPและ@JCRGพหุ@JCRGภาคี@NCMNสืบเนื่อง@VSTAมา@XVAE
 โดย@JSBRตลอด@ADVN <space> ทั้ง@DDBQนี้@DDAC โดย@FIXV
 แบ่ง@VACTเป็น@VSTAกิจกรรม@NCMNต่างๆ@NPRP คือ@VSTAการ@FIXN
 ดำเนิน@VACTโครงการวิจัย@NCMNร่วม@ADVNgัน@ADVN การ@FIXN
 แลกเปลี่ยน@VACTและ@JCRGถ่ายทอด@VACTเทคโนโลยี@NCMN
 ระหว่าง@RPREกัน@ADVN รวม@VACTทั้ง@DDBQการ@FIXNเข้า@VACT
 ร่วม@ADVNแสดง@VACTข้อคิดเห็น@NCMNและ@JCRGสนับสนุน@VACT
 ข้อมูล@NCMNที่@PRELเกี่ยวข้อง@VSTAที่@PRELจะ@XVBMเป็น@VSTA
 ประโยชน์@NCMNต่อ@RPREการ@FIXNวิจัย@VACTและ@JCRGพัฒนา@VACT
 ทาง@RPREเทคโนโลยี@NCMNของ@RPREไทย@NPRP

จากตัวอย่างพบว่าหัวข้อเรื่องนี้มีจำนวนประโยคที่ผ่านการหาขอบเขตของประโยคด้วยขั้นตอนวิธีวินนาวอยู่ทั้งหมด 2 ประโยค คือ

- (1) **ประโยคที่ 1** เนค@NPRPเท@VACTค@NPRPได้@XVAEดำเนิน@VACT ความ@FIXNร่วม@ADVNmื่อ@NCMNใน@RPREการ@FIXNวิจัย@VACT และ@JCRGพัฒนา@VACTทาง@RPREเทคโนโลยี@NCMNกับ@RPRE ต่าง@PDMNประเทศ@NCMNทั้ง@DDBQใน@RPREระดับ@NCMN ท@NCMNวิชา@NPRPที่@NPRPและ@JCRGพหุ@JCRGภาที่@NCMN สืบเนื่อง@VSTAMA@XVAE โดย@JSBRตลอด@ADVn
- (2) **ประโยคที่ 2** ทั้ง@DDBQนี้@DDAC โดย@FIXVแบ่ง@VACTเป็น@VSTA กิจกรรม@NCMNต่างๆ@NPRP คือ@VSTAการ@FIXNดำเนิน@VACT โครงการวิจัย@NCMNร่วม@ADVnกัน@ADVn การ@FIXN แลกเปลี่ยน@VACTและ@JCRGถ่ายทอด@VACTเทคโนโลยี@NCMN ระหว่าง@RPREกัน@ADVn รวม@VACTทั้ง@DDBQการ@FIXN เข้า@VACTร่วม@ADVnแสดง@VACTข้อคิดเห็น@NCMNและ@JCRG สนับสนุน@VACTข้อมูล@NCMNที่@PRELเกี่ยวข้อง@VSTAที่@PREL จะ@XVBMเป็น@VSTAประโยชน์@NCMNต่อ@RPREการ@FIXN วิจัย@VACTและ@JCRGพัฒนา@VACTทาง@RPREเทคโนโลยี@NCMN ของ@RPREไทย@NPRP

3.4.4 การจัดทำข้อสรุปโดยใช้เทคนิคการสรุปแฟร็กทัล

- (1) สมมติให้คำสำคัญของเอกสารฉบับนี้ คือ คำว่า “งาน@NCMN” และ “ระบบ@NCMN”
- (2) สมมติให้คำ Heading ที่ปรากฏอยู่ในชื่อเรื่องของเอกสารฉบับนี้ คือ คำว่า “รายงาน@NCMN” “ผล@NCMN” “ประจำปี@NCMN” “อิเล็กทรอนิกส์@NCMN” “คอมพิวเตอร์@NCMN” และ “แห่งชาติ@NCMN”
- (3) สมมติให้คำ Heading ที่ปรากฏอยู่ในชื่อหัวเรื่องของเอกสารฉบับนี้ คือ คำว่า “องค์กร@NCMN”
- (4) คำนำน้หนักของประโยคต่าง ๆ คำนวณได้จากสมการ ค่าคะแนน Sentence = Thematic Weight + Location Weight + Heading Weight

(5) Thematic Weight คือ คำนวณน้ำหนักของประโยคที่มีคำสำคัญของเอกสารปรากฏอยู่ ซึ่งคำนวณได้จากสมการ $\text{Thematic Weight} = A \times \log_2[(B \times C) / D]$ โดยที่ A คือ จำนวนครั้งที่คำสำคัญคำนี้ปรากฏอยู่ในประโยค B คือ จำนวนโหนดทั้งหมดของเอกสารฉบับนี้ C คือ ความยาวของคำสำคัญคำนี้ และ D คือ จำนวนโหนดที่สำคัญคำนี้ปรากฏอยู่

(6) Location Weight คือ คำนวณน้ำหนักของประโยคโดยพิจารณาจากตำแหน่งของประโยคนั้นในเอกสารต้นฉบับ ซึ่งคำนวณได้จากสมการ $\text{Location} = [1 / ((\min(A, B)) \times C)]$ โดยที่ A คือ ระยะห่างระหว่างโหนดที่ประโยคนั้นปรากฏอยู่กับโหนดที่อยู่ด้านซ้ายสุด และโหนดทั้งสองอยู่ภายใต้โหนดแม่ตัวเดียวกัน B คือ ระยะห่างระหว่างโหนดที่ประโยคนั้นปรากฏอยู่กับโหนดที่อยู่ด้านขวาสุด และโหนดทั้งสองอยู่ภายใต้โหนดแม่ตัวเดียวกัน $\min(A, B)$ คือ ค่า A ถ้า $A < B$ หรือ ค่า B ถ้า $B < A$ และ C คือ จำนวนประโยคทั้งหมดของโหนดที่ประโยคนั้นปรากฏอยู่

(7) Heading Weight คือ คำนวณน้ำหนักของประโยคที่มีคำ Heading ที่ปรากฏอยู่ในชื่อเรื่อง ชื่อหัวเรื่อง หรือชื่อหัวเรื่องย่อยของเอกสารปรากฏอยู่ ซึ่งคำนวณได้จากสมการ $\text{Heading Weight} = [A / B]$ โดยที่ A คือ ค่าคะแนน TF-IDF ของประโยคนี้ และ B คือ จำนวนโหนดทั้งหมดที่อยู่ในระดับเดียวกับโหนดที่ประโยคนี้ปรากฏอยู่

(8) คำนวณหาคำนวณน้ำหนักของประโยคที่ 1 จากสมการในข้อ (4)

(8.1) เนื่องจากในประโยคนี้ไม่มีคำสำคัญของเอกสารปรากฏอยู่ ดังนั้น

Thematic Weight ของประโยคนี้จึงมีค่าเท่ากับ 0

(8.2) สมมติให้ประโยคนี้ปรากฏอยู่ในโหนดที่อยู่ด้านซ้ายสุด ดังนั้น Location

Weight ของประโยคนี้ $= [1 / (1 \times 2)] = 1/2$

(8.3) เนื่องจากในประโยคนี้ไม่มีคำ Heading ที่ปรากฏอยู่ทั้งในชื่อเรื่องและชื่อหัวเรื่องของเอกสารปรากฏอยู่ ดังนั้น Heading Weight ของประโยคนี้จึงมีค่าเท่ากับ 0

(8.4) ดังนั้นค่าน้ำหนักของประโยคที่ 1 $= 0 + 1/2 + 0 = 1/2$

(9) คำนวณหาคำนวณน้ำหนักของประโยคที่ 2 จากสมการในข้อ (5)

(9.1) เนื่องจากในประโยคนี้ไม่มีคำสำคัญของเอกสารปรากฏอยู่ ดังนั้น

Thematic Weight ของประโยคนี้จึงมีค่าเท่ากับ 0

(9.2) สมมติให้ประโยคนี้ปรากฏอยู่ในโหนดที่อยู่ด้านซ้ายสุด ดังนั้น Location

Weight ของประโยคนี้ $= [1 / (1 \times 2)] = 1/2$

(9.3) เนื่องจากในประโยคนี้ไม่มีคำ Heading ที่ปรากฏอยู่ทั้งในชื่อเรื่องและชื่อหัวเรื่องของเอกสารปรากฏอยู่ ดังนั้น Heading Weight ของประโยคนี้จึงมีค่าเท่ากับ 0

(9.4) ดังนั้นค่าน้ำหนักของประโยคที่ 2 = $0 + 1/2 + 0 = 1/2$

(10) สมมติให้หัวเรื่องนี้มีจำนวนประโยคที่จะถูกคัดเลือกออกมาจัดทำเป็นข้อสรุปเท่ากับ 2 ประโยคซึ่งมีค่าน้อยกว่าค่าขีดแบ่งที่กำหนดเอาไว้เท่ากับ 3 ดังนั้นประโยคทั้งสองจึงถูกคัดเลือกออกมาจัดทำเป็นข้อสรุป