

บทที่ 7

สรุปผลการวิจัยและข้อเสนอแนะ

1. ความเบื้องต้น

งานวิจัยฉบับนี้เป็นการเสนอและทดสอบแนวคิดของการตัดแยกคำหรือวลีที่เหมาะสม และการสกัดคำหรือวลีสำคัญในเอกสารภาษาไทย สำหรับเป็นตัวแทนของเอกสารที่เหมาะสมในระบบการค้นคืนสารสนเทศ โดยอาศัยโครงข่ายประสาทเทียมร่วมกับ เทคนิคอิเวริสติก กฎทางไวยากรณ์ภาษาศาสตร์ รวมทั้งใช้สถิติความถี่และตำแหน่งการเกิดขึ้นของคำหรือวลี โดยในบทนี้จะเป็นการกล่าวถึงบทสรุปเกี่ยวกับงานวิจัย ปัญหาและอุปสรรค และบทวิจารณ์ รวมถึงข้อเสนอแนะสำหรับการนำไปศึกษาและพัฒนาต่อไปได้

2. สรุปเทคนิคที่ใช้ในการทดสอบ และพัฒนาระบบในงานวิจัย

งานวิจัยการประยุกต์ใช้เทคโนโลยีโครงข่ายประสาทเทียมกับการสกัดวลีสำคัญภาษาไทย โดยอาศัยหลักไวยากรณ์ภาษาศาสตร์สามารถนำมาเป็นแนวทางในการค้นหาใจความสำคัญของเอกสารในรูปแบบของคำหรือวลีที่สามารถนำมาเป็นตัวแทนเอกสารได้ ดังนี้

2.1 เทคนิคของการตัดคำ

เทคนิคการตัดคำที่อยู่ในปัจจุบันนี้มีอยู่ด้วยกันหลายเทคนิค โดยสำหรับในงานวิจัยนี้เลือกใช้การตัดคำแบบไบแกรม ที่เป็นการพิจารณาจากประเภทของคำต่าง ๆ ที่มีการเรียงคำตามหลักไวยากรณ์ภาษา โดยเป็นการตัดคำพร้อมการกำกับหน้าที่ของคำหรือความหมายของคำ เพื่อสามารถที่จะนำไปใช้ในการประมวลผลที่มีประสิทธิภาพมากยิ่งขึ้นในด้านความหมายและไวยากรณ์ เพราะคำแต่ละคำนั้นจะมีหน้าที่และความสำคัญในไวยากรณ์ภาษาที่แตกต่างกัน และหน้าที่คำนั้นเป็นปัจจัยหนึ่งในการพิจารณาความสำคัญของคำหรือวลี โดยใน

งานวิจัยนี้ผู้วิจัยเลือกใช้หน้าที่ของคำเป็นพารามิเตอร์หนึ่งของคำนำหน้าของคำสำหรับเป็นอินพุตเข้าสู่โครงข่ายในการเรียนรู้และทดสอบ ดังนั้นการตัดคำแบบไบแกรม จึงมีความเหมาะสมต่อการพิจารณาคำหรือวลีที่อ้างอิงกับหลักไวยากรณ์ภาษาศาสตร์ในกระบวนการสกัดคำหรือวลีสำคัญและการตัดคำแบบไบแกรม นั้นยังสามารถช่วยแก้ไขความคลุมเครือที่เกิดขึ้นในการตัดคำจากหลักการของการตัดคำที่อาศัยพจนานุกรมช่วยในการจัดเก็บคำศัพท์ รวมถึงความรวดเร็วในการตัดคำจากการใช้โครงสร้างของพจนานุกรมแบบทรี

2.2 เทคนิคโครงข่ายประสาทเทียม

โครงข่ายประสาทเทียม เป็นเทคนิคการเรียนรู้ของเครื่องแบบหนึ่งซึ่งนิยมใช้กันอย่างแพร่หลาย เนื่องจากโครงข่ายประสาทเทียมสามารถเรียนรู้จากค่าความผิดพลาดเพื่อปรับค่าน้ำหนักของเส้นเชื่อมภายในโครงข่าย ซึ่งเป็นการกำหนดความสำคัญให้แก่เส้นเชื่อมต่าง ๆ ดังนั้นการนำลักษณะหรือคุณสมบัติสำคัญมาป้อนเป็นอินพุตให้แก่โครงข่ายประสาทเทียมและผ่านกระบวนการเรียนรู้แล้วนั้น โครงข่ายประสาทเทียมจะสามารถเรียนรู้และจดจำลักษณะของข้อมูลที่มีความสำคัญได้จากข้อมูลที่ได้เรียนรู้และมีการเรียนรู้ที่มากพอ และจะทำการทดสอบโดยการสกัดคำหรือวลีสำคัญที่เป็นตัวแทนเอกสารได้ โดยในการทดลองครั้งนี้ได้ใช้เทคนิคแบ็คพรอพาเกชันหรือเทคนิคการแพร่กลับของข้อมูลสำหรับทำการเรียนรู้ จากการกำหนดคำนำหน้าซึ่งเป็นการให้ความสำคัญในแต่ละคุณสมบัติของคำหรือวลี และสถาปัตยกรรมของโครงข่ายที่เลือกใช้ในงานวิจัยนี้ คือ โครงข่ายแบบ เพอร์เซพตรอนหลายชั้นหรือแบบแพร่กลับ โดยมีลักษณะโครงข่ายดังนี้

2.2.1 สถาปัตยกรรมโครงข่ายแบบ เพอร์เซพตรอนหลายชั้น หรือแบบแพร่กลับ

สถาปัตยกรรมโครงข่ายแบบ เพอร์เซพตรอนหลายชั้นหรือแบบแพร่กลับนั้น มีคุณสมบัติที่ดีและเหมาะสมหลายประการ ได้แก่

2.2.1.1 โครงข่ายมีความเหมาะสมสำหรับความต้องการในการดึงโครงสร้างออกจากข้อมูลที่มีอยู่ โดยในงานวิจัยนี้ผู้วิจัยต้องการค้นหาใจความสำคัญหรือความสัมพันธ์ของข้อมูลในเอกสาร ในด้านของค่าความถี่การเกิดขึ้น หน้าที่คำ และตำแหน่งการเกิดขึ้นของคำหรือวลีสำคัญในแต่ละส่วนของเอกสาร เช่น ในประโยคหรือในย่อหน้า โดยผ่านการเรียนรู้และทดสอบของโครงข่าย

2.2.1.2 โครงข่ายสามารถสร้างแบบจำลองสำหรับงานที่มีความซับซ้อนได้ เช่น แบบจำลองแสดงความสัมพันธ์หลายตัวแปร โดยการประเมินความสำคัญของวลีในงานวิจัยนี้ พิจารณาจากราคาความสัมพันธ์ของตัวแปรหลายตัว คือ ตัวแปรค่าน้ำหนักความถี่, ตัวแปรค่าน้ำหนักตำแหน่งในประโยค, ตัวแปรค่าน้ำหนักตำแหน่งในย่อหน้า, ตัวแปรค่าน้ำหนักหน้าที่ของคำที่อ้างอิงหลักไวยากรณ์ภาษา และตัวแปรค่าน้ำหนักตำแหน่งของเอกสาร

2.2.1.3 โครงข่ายสามารถทำงานกับข้อมูลที่มีความผิดพลาด หรือข้อมูลที่ไม่มีโครงสร้างได้ ซึ่งเป็นลักษณะของข้อมูลประเภทเอกสารที่รูปแบบในการเขียนที่แตกต่างกันในหลากหลายรูปแบบ

2.2.1.4 โครงข่ายสามารถตอบคำถามที่ไม่เคยถูกถามมาก่อนได้ ทำให้แบบจำลองโครงข่ายประสาทเทียมเหมาะสำหรับสกัดวลีสำคัญที่สร้างโดยการเรียนรู้จากเอกสารตัวอย่างสามารถนำมาใช้เลือกวลีสำคัญจากเอกสารทดสอบได้

2.2.2 อัลกอริทึมสำหรับการเรียนรู้

อัลกอริทึมสำหรับการเรียนรู้ในงานวิจัยนี้คือ แม็ทพรอพากะชั่น ที่ใช้สำหรับการปรับค่าน้ำหนักในเส้นเชื่อมต่อระหว่างโหนดให้เหมาะสม โดยเป็นการปรับปรุงน้ำหนักคะแนนของเครือข่าย จากค่าความผิดพลาดที่เกิดขึ้น ทำให้เป็นเทคนิคการเรียนรู้ที่เหมาะสมสำหรับนำไปใช้ในการแก้ปัญหาที่เกี่ยวข้องกับการค้นหาใจความสำคัญที่ซ่อนอยู่ และเป็นเทคนิคการเรียนรู้ที่นิยมใช้เนื่องจากมีประสิทธิภาพและง่ายที่จะเป็นต้นแบบสำหรับโครงข่ายเพื่อการจำแนกต่อไป

2.2.3 ลักษณะของโครงข่าย

ลักษณะของโครงข่ายในงานวิจัยนี้คือ จำนวนโหนดอินพุต 1-5 โหนด จำนวนโหนดในชั้นซ่อน 1-5 โหนด และจำนวนโหนดในชั้นผลลัพธ์ 1 โหนด ดังนี้

2.2.3.1 จำนวนโหนดอินพุต

- อินพุตโหนดค่าน้ำหนักความถี่
- อินพุตโหนดค่าน้ำหนักตำแหน่งในประโยค
- อินพุตโหนดค่าน้ำหนักตำแหน่งในย่อหน้า
- อินพุตโหนดค่าน้ำหนักตำแหน่งในเอกสาร
- อินพุตโหนดค่าน้ำหนักหน้าที่คำ

2.2.3.2 จำนวนโหนดในชั้นซ่อน

จำนวนโหนดในชั้นซ่อนนั้น มีค่าจำนวนโหนดตั้งแต่ 1 ถึง 5 โหนด ซึ่งจะแตกต่างกันออกไปใช้ในการทดสอบแต่ละครั้ง

2.2.3.3 จำนวนโหนดในชั้นผลลัพธ์

จำนวนโหนดในชั้นผลลัพธ์นั้น มีค่าเท่ากับ 1 โหนดเสมอ โดยแทนผลในการทำนายว่าคำหรือวลีนั้นมีความสำคัญหรือไม่ (1 หรือ 0)

โดยการเลือกใช้จำนวนโหนดในชั้นอินพุตและในชั้นซ่อนนั้น ขึ้นอยู่กับแต่ละเงื่อนไขในการทดสอบ

2.2.4 รูปแบบของลักษณะค่าน้ำหนักเข้าสู่โครงข่าย

การหนดค่าน้ำหนักเข้าสู่โครงข่ายนั้น ผู้วิจัยได้พัฒนาโปรแกรมให้ค่าน้ำหนักที่สอดคล้องกับกฎไวยากรณ์ของภาษาไทย โดยเป็นการให้ค่าน้ำหนักที่สัมพันธ์กับรูปแบบข้อมูล และเป็นค่าน้ำหนักในการเชื่อมต่อของโหนดแต่ละโหนดระหว่างชั้นรับข้อมูลและชั้นซ่อน โดยกำหนดค่าให้มีค่าอยู่ในช่วงระหว่าง 0 - 1 ซึ่งเป็นค่าที่เหมาะสมสำหรับการเรียนรู้ของโครงข่ายประสาทเทียมเทียม

3. ผลการทดสอบ

จากผลการทดสอบในการนำเทคนิคของโครงข่ายประสาทเทียมเทียมเข้ามาใช้พัฒนาและปรับปรุงการสกัดคำหรือวลีสำคัญในเอกสาร เพื่อเพิ่มประสิทธิภาพของการค้นคืนดังนี้

3.1 ความถูกต้องของการสกัดสารสนเทศ

สำหรับการทดลองด้วยชุดข้อมูลทั้ง 3 ชุด ในการทดสอบในการทดสอบทั้ง 2 ประเภท คือ 1. การทดสอบการเลือกใช้พารามิเตอร์และจำนวนโหนดในชั้นซ่อน และ 2. การหาค่าที่เหมาะสมในแต่ละพารามิเตอร์คุณสมบัติของคำหรือวลี นั้น ได้ผลการทดสอบในด้านค่าความถูกต้องดังนี้

3.1.1 ผลการทดสอบประเภทที่ 1

การทดสอบเพื่อประเมินการเลือกใช้พารามิเตอร์และการเลือกใช้จำนวนโหนดในชั้นซ่อน สำหรับแต่ละความสัมพันธ์ระหว่างเอกสารชุดฝึกสอนและเอกสารชุดทดสอบ ได้ค่าความถูกต้องสูงสุด ดังนี้

3.1.1.1 ชุดการทดสอบที่ 1 ได้ค่าความถูกต้องดังนี้

ค่าความถูกต้องในการสอนสูงถึง 89.31%

ค่าความถูกต้องในการทดสอบสูงถึง 90.26%

3.1.1.2 ชุดการทดสอบที่ 2 ได้ค่าความถูกต้องดังนี้

ค่าความถูกต้องในการสอนสูงถึง 89.31%

ค่าความถูกต้องในการทดสอบสูงถึง 79.09%

3.1.1.2 ชุดการทดสอบที่ 2 ได้ค่าความถูกต้องดังนี้

ค่าความถูกต้องในการสอนสูงถึง 90.16%

ค่าความถูกต้องในการทดสอบสูงถึง 82.40%

3.1.2 ผลการทดสอบประเภทที่ 2

การทดสอบโดยการหาค่าตัวเลขในเริ่มต้นที่นำไปคูณกับค่าน้ำหนักอินพุตจริง เพื่อให้ได้ค่าน้ำหนักอินพุตที่เหมาะสมสำหรับการเรียนรู้ของโครงข่าย ที่จะทำให้โครงข่ายสามารถจำแนกข้อมูลได้แม่นยำที่สุด โดยได้ค่าความถูกต้องสูงสุด ดังนี้

3.1.2.1 ชุดการทดสอบที่ 1 ได้ค่าความถูกต้องดังนี้

ค่าความถูกต้องในการสอนสูงถึง 89.31%

ค่าความถูกต้องในการทดสอบสูงถึง 90.26%

3.1.2.2 ชุดการทดสอบที่ 2 ได้ค่าความถูกต้องดังนี้

ค่าความถูกต้องในการสอนสูงถึง 89.39%

ค่าความถูกต้องในการทดสอบสูงถึง 79.09%

3.1.2.2 ชุดการทดสอบที่ 3 ได้ค่าความถูกต้องดังนี้

ค่าความถูกต้องในการสอนสูงถึง 90.07%

ค่าความถูกต้องในการทดสอบสูงถึง 82.40%

3.2 การเลือกใช้พารามิเตอร์ที่เหมาะสม

การเลือกใช้พารามิเตอร์ที่เหมาะสมที่สุด คือ พารามิเตอร์ ค่าน้ำหนักความถี่การเกิดขึ้นของคำหรือวลีนั้น และการเลือกใช้พารามิเตอร์ร่วมกับการเรียนรู้ของโหนดในชั้นซ่อนของโครงข่าย 3 – 5 โหนดนั้น คือ การเลือกใช้พารามิเตอร์ในเงื่อนไขที่ 17 , 20 , 25 , 26 , 30

- เงื่อนไขที่ 17 - ค่าน้ำหนักความถี่ ค่าน้ำหนักตำแหน่งในประโยค และค่าน้ำหนักหน้าที่คำ

- เงื่อนไขที่ 20 - ค่าน้ำหนักความถี่ ค่าน้ำหนักตำแหน่งในย่อหน้า และน้ำหนักตำแหน่งในเอกสาร

- เงื่อนไขที่ 25 - ค่าน้ำหนักความถี่ ค่าน้ำหนักตำแหน่งในประโยค ค่าน้ำหนักตำแหน่งในย่อหน้า และค่าน้ำหนักหน้าที่คำ

- เงื่อนไขที่ 26 - ค่าน้ำหนักความถี่ ค่าน้ำหนักตำแหน่งในประโยค ค่าน้ำหนักตำแหน่งในย่อหน้า และค่าน้ำหนักตำแหน่งในเอกสาร

- เงื่อนไขที่ 30 - ค่าน้ำหนักความถี่ ค่าน้ำหนักตำแหน่งในประโยค ค่าน้ำหนักตำแหน่งในย่อหน้า ค่าน้ำหนักหน้าที่คำ และค่าน้ำหนักตำแหน่งในเอกสาร

4. สรุปผลการวิจัย

4.1 สรุปผลการทดสอบค่าความถูกต้อง

จากผลการวิจัยในบทที่ 6 และการให้ค่าน้ำหนักทั้ง 5 คุณสมบัติของคำหรือวลีสาคัญนั้น สามารถสรุปได้ว่า

4.1.1 การเลือกใช้คุณสมบัติของคำหรือวลีร่วมกันสำหรับการสอนให้กับโครงข่ายนั้น โดยสามารถทำการเลือกใช้พารามิเตอร์ร่วมกันได้อย่างเหมาะสมดังหัวข้อที่ 3.2 นั้น จะสามารถทำให้โครงข่ายสามารถจำแนกข้อมูลได้ค่าความถูกต้องสูงถึง 90.26% เพิ่มจำนวนคุณสมบัติที่ใช้ในการสอนให้กับโครงข่ายนั้น ถ้ายิ่งจำนวนคุณสมบัติ (ข้อมูลอินพุต) เข้าสู่โครงข่ายสำหรับการสอนให้กับโครงข่ายมากขึ้นจะยิ่งทำให้ประสิทธิภาพในด้านค่าความถูกต้องของการระบุคำหรือวลีสาคัญได้ดี

4.1.2 การกำหนดค่าน้ำหนักคุณสมบัติของคำหรือวลีที่ผู้วิจัยได้กำหนดนั้น (คูณด้วยค่า 1) มีความเหมาะสมต่อการเรียนรู้ของโครงข่าย และสามารถทำให้โครงข่ายแยกแยะข้อมูลมีความแม่นยำสูง

4.1.3 ปัจจัยในการเลือกใช้จำนวนพารามิเตอร์ การเลือกใช้พารามิเตอร์ร่วมกัน การเลือกใช้จำนวนโหนดในชั้นซ่อน(จำนวน 3-5 โหนด)ที่ทำให้โครงข่ายได้เรียนรู้มากพอและซับซ้อน นั้น มีผลต่อค่าความถูกต้อง

4.1.4 การเพิ่มจำนวนพารามิเตอร์ที่มากขึ้นนั้น ไม่มีผลโดยตรงต่อค่าความถูกต้องในการจำแนกของโครงข่าย

4.1.5 การกำหนดระดับค่าน้ำหนักมากน้อยในแต่ละคุณสมบัติค่านั้น มีความแตกต่างกันในแต่ละการเลือกใช้ข้อมูลสอนและทดสอบ

4.1.6 การเลือกใช้ออกสารชุดฝึกสอนและออกสารชุดทดสอบนั้น แม้จะใช้เอกสารที่เป็นเอกสารเดียวกัน ก็ไม่มีผลต่อค่าความถูกต้องของการสกัด

4.1.7 การใช้เอกสารชุดฝึกสอนและเอกสารชุดทดสอบเป็นเอกสารชุดเดียวกันนั้น มีผลให้ค่าความถูกต้องในการสกัดมีค่าสูง

4.1.8 ผลการวิจัยนี้สามารถบอกได้ว่าโครงข่ายประสาทเทียมแบบเพอร์เซพตรอนหลายชั้น หรือแบบป้อนไปข้างหน้าที่มีการเรียนรู้แบบแพร่กลับนั้น เหมาะสมต่อการเรียนรู้ในการค้นหาใจความสำคัญโดยการสกัดคำหรือวลีที่สามารถเป็นตัวแทนของเอกสาร ได้ดีถึง 90.26%

4.1.9 การเลือกใช้พารามิเตอร์ค่าน้ำหนักความถี่การเกิดขึ้นนั้น สามารถทำให้ค่าความถูกต้องที่เกิดขึ้นนั้นมีโอกาสที่ค่าความถูกต้องที่ได้มีค่าสูง โดยจากผลการทดสอบการเลือกใช้น้ำหนักความถี่นั้น ในการทดสอบที่ 3 ที่มีเอกสารชุดฝึกสอนและเอกสารชุดทดสอบเป็นเอกสารคนละโดเมนกันนั้น โครงข่ายสามารถเรียนรู้ได้ค่าความถูกต้องในข้อมูลชุดฝึกสอนถึง 89.71% และสามารถสกัดคำหรือวลีสำคัญในข้อมูลชุดทดสอบได้ค่าความถูกต้องถึง 80.00% ซึ่งมีค่าสูงกว่าเฉลี่ยในการทดสอบที่ 3 คือ ค่าความถูกต้องในการฝึกสอนเท่ากับ 89.27% และค่าความถูกต้องในการทดสอบเท่ากับ 79.20% โดยผู้วิจัยมีความคิดเห็นว่าการเลือกใช้น้ำหนักความถี่การเกิดขึ้นของคำหรือวลีเพียงอย่างเดียวนั้น ในบางกรณีมีความเพียงพอต่อการสกัดคำหรือวลีสำคัญในระดับที่ค่าความถูกต้องในการสกัดมีค่าสูง เพราะค่าน้ำหนักความถี่นั้นแสดงถึงการเน้นย้ำซ้ำๆ คำหรือวลีเหล่านั้น ว่าในเอกสารเหล่านั้นกำลังกล่าวถึงเรื่องใดหรือสิ่งใด แต่อย่างไรก็ตาม ค่าน้ำหนักอื่น ๆ ของคำหรือวลีนั้นก็มีความจำเป็นและมีความสำคัญที่สามารถส่งผลให้ค่าความถูกต้องในการสกัดมีค่าสูงได้ในกรณีการทดสอบอื่น ๆ และยังมีโอกาสที่ค่าความถูกต้องที่ได้จะสูงที่สุดในการทดสอบ โดยจากการทดสอบนั้นการเลือกใช้น้ำหนักพารามิเตอร์ทั้ง 5 ค่าน้ำหนักนั้นทำให้ค่าความถูกต้องที่ได้มีค่าสูงที่สุด

5. อุปสรรค

การเขียนเอกสารในปัจจุบันนี้มีด้วยกันหลากหลายลักษณะ ซึ่งอาจมีรูปแบบในการวางตำแหน่ง การใช้คำ การเขียนประโยค การเขียนย่อหน้า ที่แตกต่างออกไปจากแนวคิดของงานวิจัยนี้ โดยที่ถ้าเอกสารที่ทำการทดสอบนั้น มีรูปแบบที่ไม่ตรงตามเงื่อนไขเอกสารที่กำหนดไว้ในโปรแกรมการกำหนดค่าน้ำหนักอาจไม่สามารถกำหนดค่าน้ำหนักได้อย่างเหมาะสม

6. ข้อเสนอแนะงานวิจัย

6.1 การเลือกข้อมูลทดสอบนั้น ผู้วิจัยทำการพิจารณาเลือกส่วนของข้อมูลหรือส่วนของเอกสารด้วยตนเอง จึงอาจทำให้ข้อมูลบางส่วนนั้นไม่ถูกนำมาใช้ และอาจเป็นส่วนที่ทำให้ค่าความถูกต้องที่สูงกว่าเดิม ดังนั้นในขั้นตอนของการเลือกข้อมูลนั้น ควรใช้เทคนิคในการเลือกข้อมูลที่ทำให้ได้ค่าความถูกต้องที่สูงที่สุดในการทดสอบ เช่น เทคนิคเลือกข้อมูลแบบความเที่ยงตรง (K-fold Cross Validation) หรือ เทคนิคสุ่มข้อมูลแบบ Leave-one-out Cross Validation

6.2 การเพิ่มกฎการเรียนรู้เกี่ยวกับรูปแบบต่าง ๆ ของประโยคให้โครงข่ายได้เรียนรู้มากขึ้น เช่น ประโยคความเดียว , ประโยคความรวม , ประโยคสังกร ฯลฯ เพื่อให้สามารถระบุตำแหน่งใจความสำคัญของแต่ละรูปแบบของประโยค ซึ่งแต่ละรูปแบบก็จะมีการวางตำแหน่งใจความสำคัญของประโยคที่แตกต่างกันไป

6.3 การเขียนเอกสารในปัจจุบันนี้มีด้วยกันหลากหลายลักษณะ ซึ่งอาจมีรูปแบบในการวางตำแหน่ง การใช้คำ การเขียนประโยค การเขียนย่อหน้า ที่แตกต่างออกไปจากแนวคิดของงานวิจัยนี้ โดยที่ถึงแม้งานวิจัยจะพยายามหารูปแบบที่เป็นกลางของการเขียนเอกสารก็ตามที ดังนั้นถ้าผู้ที่จะนำงานวิจัยนี้ไปพัฒนาต่อไปนั้น อาจจะเป็นที่จะต้องเพิ่มกฎการเรียนรู้ให้กับโปรแกรมให้ค่าน้ำหนักมากขึ้น หรือให้เหมาะสมกับรูปแบบของเอกสารที่ต้องการค้นหาใจความสำคัญ แนวคิดในการค้นหาและกำหนดค่าน้ำหนักของงานวิจัยนี้ สามารถนำไปใช้ในการสกัดคำหรือวลีสำคัญที่เป็นเอกสารภาษาอังกฤษได้ เพราะเอกสารภาษาอังกฤษนั้นมีรูปแบบการเขียนที่คล้ายคลึงกับเอกสารภาษาไทย

6.4 ในการกำหนดระดับค่าน้ำหนักที่อ้างอิงหลักไวยากรณ์ภาษาไทยนั้น ในกระบวนการขั้นตอนของการให้ค่าน้ำหนักตำแหน่งในประโยค ค่าน้ำหนักตำแหน่งในย่อหน้า ค่าน้ำหนักตำแหน่งในเอกสาร และค่าน้ำหนักหน้าที่คำซึ่งเป็นค่าน้ำหนักที่ผู้วิจัยได้ตั้งสมมติฐานขึ้นมา นั้น ควรให้ผู้ใช้ในการกำหนดค่าน้ำหนักของคำหรือวลีในแต่ละคุณสมบัติเองได้ โดยจะกำหนดค่าที่

เลือกอัตโนมัติ(default) เป็นค่านำหนักในเบื้องต้นที่ผู้วิจัยได้ทำการกำหนดไว้ สำหรับในกรณีที่ผู้ใช้ไม่ต้องการกำหนดค่านำหนักเอง และในสำหรับค่านำหนักความถี่การเกิดขึ้นนั้นไม่จำเป็นต้องทำการกำหนดเองเพราะเป็นการนับจำนวนความถี่คำหรือวลีที่เกิดขึ้นในเอกสารนั้น ๆ

6.5 โดยแนวความคิดการให้ค่านำหนักคำหรือวลีในงานวิจัยนี้นั้น อาจจะไปประยุกต์ใช้กับระบบการค้นหาคำหรือเสิร์ชเอนจิน สำหรับเป็นเทคนิคที่ช่วยในการค้นหาเอกสารที่ตรงตามความต้องการของผู้ใช้จริง ๆ สำหรับเป็นการค้นหาจากเอกสารที่เสิร์ชเอนจินค้นคืนได้มาซ้ำอีกครั้ง เพื่อที่จะทำการเลือกเฉพาะเอกสารที่มีคำค้นตรงกับคำหรือวลีสำคัญจริงๆจากเอกสารที่เสิร์ชเอนจินค้นหามาได้ดังนี้

6.5.1 ในการค้นหาของเสิร์ชเอนจินนั้น เมื่อผู้ใช้ทำการป้อนคำค้นเข้าสู่ระบบการค้นหา นั้น ระบบเสิร์ชเอนจินจะทำการค้นหาทุกเอกสารที่มีคำหรือวลีนั้น ๆ ปรากฏอยู่มาทั้งหมด โดยที่บางครั้ง คำหรือวลีนั้น ๆ ไม่ได้เป็นคำหรือวลีสำคัญในเอกสาร

6.5.2 เอกสารที่ผ่านขั้นตอนการค้นคืนได้มานั้น นำมาผ่านการกลั่นกรองเทคนิคในการวิจัย สำหรับเพื่อค้นหาคำหรือวลีสำคัญในเอกสารเหล่านั้น ที่ตรงกับคำค้นที่ผู้ใช้ป้อนเข้าสู่ระบบการค้นหาจริง ๆ โดยที่ไม่ได้เป็นแค่เพียงคำหรือวลีในเอกสารและไม่มีความสำคัญ