

บทที่ 4

ผลการวิจัยและอภิปรายผล

ในบทนี้จะอธิบายผลการวิจัยและอภิปรายผลที่ได้ดังมีรายละเอียดดังนี้

ผลการวิจัย

1. แบบจำลองที่เหมาะสมของระบบ BN-LVCSR

ผลลัพธ์จากการทดลองขั้นตอนการออกแบบโครงสร้างระบบ BN-LVCSR ในบทที่ 3 จะเป็นผลลัพธ์จากการถอดรหัสซึ่งเป็นคำตอบของการรู้จำเสียงนั้น เป็นลำดับคำที่มีความน่าจะเป็นสูงสุดเมื่อนำมาหาเปอร์เซ็นต์ความถูกต้องและเปอร์เซ็นต์ความแม่นยำ จะได้ผลดังนี้

ตารางที่ 4.1

ตารางเปรียบเทียบผลการทดลองในบทที่ 3

ชุดข้อมูล	LOTUS	BN	LOTUS + BN
จำนวนประโยค	20,984	9,115	30,099
จำนวนคำ	390,584	170,219	560,803
จำนวนคำศัพท์	5,412	7,123	9,602
%Correct	64.41%	68.64%	66.95%
%Accuracy	52.97 %	63.56%	61.02%

จากตารางที่ 4.1 จะเห็นว่าระบบต้นแบบมีเปอร์เซ็นต์ความถูกต้อง 64.41% เมื่อนำข้อมูลคลังข้อมูลข่าวภาษาไทย (BN Corpus) มาเพิ่มให้กับข้อมูล LOTUS Corpus จะทำให้เปอร์เซ็นต์ค่าความถูกต้องเพิ่มมากขึ้นเพียง 2.54% แต่เมื่อเปรียบเทียบกับคลังข้อมูลข่าวภาษาไทยเพียงอย่างเดียวจะให้เปอร์เซ็นต์ค่าความถูกต้องเพิ่มขึ้น 4.23% เพราะข้อมูลของคลังข้อมูลข่าวภาษาไทยมีความใกล้เคียงกับชุดข้อมูลทดสอบ ดังนั้นเปอร์เซ็นต์ความถูกต้องที่ดีที่สุดจะเป็น 68.64% ซึ่งผลลัพธ์ได้นี้ยังไม่เป็นที่น่าพึงพอใจ เปอร์เซ็นต์ความถูกต้องควรจะมีค่ามากกว่า 70% ซึ่งผู้วิจัยจะทำการปรับปรุงระบบต่อไป

1.1 สมมติฐานที่ 1

ในการฝึกฝนแบบจำลองภาษา ผู้วิจัยคิดว่าถ้าทำการเพิ่มข้อมูลสำหรับการฝึกฝนแบบจำลองภาษาให้มากขึ้น จะทำให้ค่าสถิติต่างๆของแบบจำลองครบมากยิ่งขึ้นและระบบจะเรียนรู้คำศัพท์ได้มากขึ้น ซึ่งข้อมูลที่จะนำมาเพิ่มนั้นเป็นข้อมูลประเภทข่าว โดยนำมาจากแหล่งอื่น การที่เพิ่มข้อมูลประเภทข่าวนั้นเป็นเพราะโครงสร้างข้อมูลจะใกล้เคียงกับชุดข้อมูลการทดสอบซึ่งเป็นข้อมูลข่าว อาจจะทำให้ความถูกต้องในการถอดรหัสมากยิ่งขึ้น

การทดลองของสมมติฐานที่ 1

ข้อมูลข่าวที่นำมาเพิ่มนั้นเป็น คลังข้อมูลข่าวจากหนังสือพิมพ์ไทยรัฐ (Newspaper) ซึ่งได้ข้อมูลนี้เก็บรวบรวมมาจากเนคเทค ประกอบด้วย 699,606 ประโยค 10,853,503 คำ และ 70,855 คำศัพท์

สำหรับคลังข้อมูลข่าวภาษาไทย ที่ใช้ฝึกฝนแบบจำลองภาษานั้นจะนำเอาคำที่ไม่มีคำอ่านอยู่ในพจนานุกรมเสียงอ่าน เช่น คำว่า [laugh] ซึ่งเป็นเสียงหัวเราะในข่าว โดยจะนำเอาเฉพาะคำออกไป จะแทนคลังข้อมูลข่าวภาษาไทย ชุดนี้ว่า BN Full Corpus

ข้อมูลที่จะนำมาใช้ฝึกฝนแบบจำลองภาษาเพิ่มนั้นจะเปรียบเทียบข้อมูลที่แตกต่างกัน 3 แบบ ได้แก่

1. ข้อมูลของ BN Full Corpus ซึ่งประกอบด้วย 13,044 ประโยค 224,167 คำ 10,314 คำศัพท์ และคำศัพท์ของพจนานุกรมเสียงอ่าน จากคลังข้อมูลข่าวภาษาไทย 38,238 คำศัพท์ (แทนแบบจำลองนี้ด้วย LM1)
2. ข้อมูลของ BN Full Corpus กับ Newspaper ซึ่งประกอบด้วย 574,220 ประโยค 8,778,769 คำ 18,643 คำศัพท์ และคำศัพท์ของพจนานุกรมเสียงอ่านจากคลังข้อมูลข่าวภาษาไทย 38,238 คำศัพท์ (แทนแบบจำลองนี้ด้วย LM2)
3. ข้อมูลของ BN Full Corpus กับ Newspaper ซึ่งประกอบด้วย 679,629 ประโยค 10,514,704 คำ 44,303 คำศัพท์ และคำศัพท์ของพจนานุกรมเสียงอ่านจากคลังข้อมูลข่าวภาษาไทย รวมกับ Newspaper 65,000 คำศัพท์ (แทนแบบจำลองนี้ด้วย LM3)

ผลการทดลองของสมมติฐานที่ 1

ผลลัพธ์จากการทดลองของสมมติฐานที่ 1 จะได้เปอร์เซ็นต์ความถูกต้องและเปอร์เซ็นต์ความแม่นยำ ดังนี้

ตารางที่ 4.2

ตารางเปรียบเทียบผลการทดลองของสมมติฐานที่ 1

ชุดข้อมูล	LM1	LM2	LM3
จำนวนประโยค	13,044	574,220	679,629
จำนวนคำ	224,167	8,778,769	10,514,704
จำนวนคำศัพท์	10,314	18,643	44,303
จำนวนคำศัพท์ ของพจนานุกรม	38,238	38,238	65,000
%Correct	67.80 %	66.53 %	64.83 %
%Accuracy	65.68 %	61.86 %	60.59 %

จากตารางที่ 4.2 แสดงให้เห็นว่าการเพิ่มข้อมูลข่าวที่ใช้ฝึกฝนแบบจำลองภาษาไม่สามารถเพิ่มเปอร์เซ็นต์ความถูกต้องและเปอร์เซ็นต์ความแม่นยำได้ สาเหตุเป็นเพราะเมื่อฝึกฝนแบบจำลองภาษาให้รู้คำศัพท์มากขึ้น ทำให้ส่วนของการถอดรหัสคำสำคัญเกิดทางเลือกของความน่าจะเป็นที่จะเกิดของคำมากขึ้น เมื่อทางเลือกมีมากขึ้นก็จะทำให้เกิดความผิดพลาดมากขึ้น และเนื่องจากชุดข้อมูลการทดสอบนั้นเป็นคำหรือกลุ่มคำสั้นๆ จึงไม่ได้ประโยชน์จากแบบจำลองภาษามากนัก

จากผลการทดลองของสมมติฐานที่ 1 แสดงให้เห็นว่าการเพิ่มข้อมูลข่าวที่ใช้ในการฝึกฝนแบบจำลองภาษาและคำศัพท์ในพจนานุกรมเสียงอ่านไม่ช่วยเพิ่มเปอร์เซ็นต์ความถูกต้องและเปอร์เซ็นต์ความแม่นยำขึ้นได้

1.2 สมมติฐานที่ 2

นำแบบจำลองจากบทที่ 3 มาใช้ในการปรับปรุง คือข้อมูลของคลังข้อมูลข่าว 9,115 ประโยค 170,219 คำ 7,123 คำศัพท์ และคำศัพท์ของพจนานุกรมเสียงอ่าน 9,842 คำ จะทำการปรับปรุงโดยลดคำศัพท์ในพจนานุกรมเสียงอ่าน

การทดลองของสมมติฐานที่ 2

จากการที่ผู้วิจัยได้สังเกตผลลัพธ์ที่ได้จากการถอดรหัสนั้น คำที่ได้มานั้นมักจะเป็นคำสั้นๆ ที่ไม่มีความหมาย ไม่น่าจะเป็นคำสำคัญได้ โดยความน่าจะเป็นของการที่จะเกิดคำสั้นๆ เหล่านั้นมีมากกว่าคำสำคัญ ดังนั้นผู้วิจัยจึงลดคำศัพท์ในพจนานุกรมเสียงอ่าน โดยคำที่เอาออกจากพจนานุกรมเสียงอ่านนี้จะเป็นคำหยุด (Stop word list) ซึ่งก็คือ คำสั้นๆ ที่ไม่มีความหมาย ไม่น่าจะนำมาเป็นคำสำคัญได้ เช่น คำเชื่อมและคำบุพบทต่างๆ (การ, ความ, ก็, จึง ฯลฯ) คำหยุดที่จะตัดออกจากพจนานุกรมเสียงอ่านนั้นมี 125 คำศัพท์ ดังนั้นพจนานุกรมเสียงอ่านที่ใช้ในการถอดรหัสจะมีคำศัพท์ 9,717 คำ

ข้อมูลที่จะนำมาใช้ฝึกฝนแบบจำลองภาษาของสมมติฐานที่ 2 คือ ข้อมูลของคลังข้อมูลข่าวภาษาไทยซึ่งประกอบด้วย 9,115 ประโยค 170,219 คำ 7,123 คำศัพท์ และคำศัพท์ของพจนานุกรมเสียงอ่าน 9,717 คำ

ผลการทดลองของสมมติฐานที่ 2

การทดลองจากสมมติฐานที่ 2 แบบจำลองที่ปรับปรุงโดยลดคำศัพท์ในพจนานุกรมเสียงอ่านแล้วจะได้ค่าเปอร์เซ็นต์ความถูกต้องเท่ากับ 78.32% และเปอร์เซ็นต์ความแม่นยำเท่ากับ 74.34% ดังแสดงในตารางที่ 4.3

ตารางที่ 4.3

ตารางเปรียบเทียบผลการทดลองของสมมติฐานที่ 2

ชุดข้อมูล	แบบจำลองที่เหมาะสม
จำนวนประโยค	9,115
จำนวนคำ	170,219
จำนวนคำศัพท์	7,123
จำนวนคำศัพท์ของพจนานุกรม	9,717
%Correct	78.32 %
%Accuracy	74.34 %

ซึ่งค่า เปอร์เซนต์ความถูกต้องและเปอร์เซนต์ความแม่นยำ 78.32%, 74.34% เป็นผลที่ผู้วิจัยยอมรับได้ดังนั้น จึงนำแบบจำลองนี้ซึ่งเป็นแบบจำลองที่ดีที่สุดนำไปใช้งานในการค้นหาคำสำคัญต่อไป

2. ขั้นตอนการค้นหาโดยใช้คำสำคัญ

โครงสร้างของลักษณะคำในภาษาไทยนั้น จะไม่มีการแยกคำ ทำให้วิธีการแยกคำเป็นสิ่งที่จำเป็น เพื่อช่วยในการค้นหาคำสำคัญ ในการค้นหาคำสำคัญในภาษาไทยนั้นนอกจากต้องการการแยกคำแล้ว ยังต้องตัดคำหยุดออกไปก่อนที่จะค้นหาคำสำคัญด้วย ซึ่งในงานวิทยานิพนธ์นี้ ผลลัพธ์ที่ได้จากการถอดรหัสนั้นได้มีการตัดคำหยุดออกไปแล้ว ดังนั้นผลลัพธ์ที่ได้จากการถอดรหัสนั้นสามารถนำไปใช้สำหรับการค้นหาคำสำคัญได้เลย

เนื่องจากขั้นตอนจากระบบ BN-LVCRS นั้นไม่ได้ให้ความถูกต้อง 100% ดังนั้นการค้นหาคำสำคัญจะพบปัญหาว่า การค้นหาคำสำคัญทั้งคำนั้น ถ้าการถอดรหัสออกมาได้ถูกต้องทั้งคำจะสามารถค้นหาคำสำคัญและแสดงประโยคออกมาได้ แต่ถ้าการถอดรหัสออกมาไม่ถูกต้องทุกคำ ผลลัพธ์จากการค้นหาอาจจะไม่แสดงประโยคใดออกมา ทำให้ผลลัพธ์ที่ได้นั้นผิดพลาดได้ ดังนั้นการใช้คำสำคัญ 10 คำ จากวิธีการ N-best โดย N=10 ในการค้นหาย่อมให้ผลลัพธ์ดีกว่าใช้คำสำคัญ 1 คำ และคำสำคัญทั้งคำนั้นอาจจะไม่ได้ถูกต้องทั้งหมด ดังนั้นการใช้คำสำคัญย่อย (Sub Keyword) มาช่วยนั้นการค้นหาด้วยนั้น จะทำให้ประสิทธิภาพในการค้นหาข้อมูลเพิ่มมากขึ้น ซึ่งคำสำคัญย่อยนั้นได้มาจากการแยกคำจากคำสำคัญ

ตารางที่ 4.4

ตัวอย่างของคำสำคัญและคำสำคัญที่ได้จากวิธีการ 10-best

คำสำคัญ	กระทรวง พาณิชย	ประกาศ	ประกาย ไฟ
1-best	กระทรวง พาณิชย	ลักลอบ	ประกาย ราย
2-best	กระทู พาณิชย	รับทราบ	ประกาย ร้าย
3-best	กระทู ฝาน	รับรอง	ประกาย ไร่
4-best	ประท้วง พาณิชย	รัดเกล้า	ประกาย ตาย
5-best	กระทู รัน	ประกาศ	ร่างกาย ราย
6-best	กระทู พันเอก	ปรับ ออก	ประกาย ไร่
7-best	ครับ ฝู พาณิชย	ปรับ แรก	ประกาย หาย
8-best	กระทู คัน	กลับ แรก	ร่างกาย ร้าย
9-best	รับรู้ พาณิชย	ปรับ รอบ	ร่างกาย ไร่
10-best	กระทู ค้าน	กลับ รอบ	ประกาย ใต้

ตารางที่ 4.4 แสดงตัวอย่างของคำสำคัญและคำสำคัญที่ได้จากวิธีการ 10-best โดยคำว่า “กระทรวง พาณิชย” จะพบได้ใน 1-best คำว่า “ประกาศ” จะพบได้ใน 5-best และคำว่า “ประกาย ไฟ” จะไม่พบใน N-best ใด แต่จะมีคำว่า “ประกาย” ซึ่งเป็นคำสำคัญย่อยที่ถูกต้อง

ดังนั้นจะทำการพัฒนาแอปพลิเคชันตามโครงสร้างที่ได้ออกแบบไว้ เพื่อให้สะดวกต่อการใช้งาน โดยระบบ BN-LVCRS จะได้ผลลัพธ์เป็น 10 คำสำคัญ ซึ่งในแต่ละคำสำคัญจะมีกลุ่มของคำสำคัญย่อยที่ได้มีการจัดลำดับความสำคัญโดยใช้ความถี่

ภาพที่ 4.1

ผลลัพธ์จากการค้นหาคำสำคัญว่า “กระทรวง พาณิชย์”

The screenshot shows a web application interface with a search bar at the top. Below the search bar, there is a section titled "Your list:" containing a table with two columns: "Keyword" and "Content". The "Keyword" column lists various terms related to the Ministry of Commerce, and the "Content" column displays the corresponding search results. The results are organized into a table with multiple rows and columns, showing a list of keywords and their associated content.

Keyword	Content
กระทรวง พาณิชย์	ไอศกานนท์ รัฐมนตรี ช่วย ว่าการ กระทรวง พาณิชย์ ประกาศ สะ เร่ง ภาษี และ เปลี่ยน ชื่อ องค์การ ค้างสินค้า
กระทรวง พาณิชย์	พาณิชย์ ชัด ชัด จำนวนย สินค้า ราคา ถูก หัก สิบ ทั่ว ประเทศ
กระทรวง พาณิชย์	พาณิชย์ เปิด เวที ระดม ความเห็น จาก ทุก ภาค ส่วน ที่ เกี่ยวข้อง ชัด ทำ ข้าว ไทย เพื่อ การ พัฒนา ที่ ยั่งยืน
กระทรวง พาณิชย์	ริ ว่าการ กระทรวง พาณิชย์ กลับ มอง ว่า เหตุ ระเบิด ที่ เกิด ขึ้น ถือ
กระทรวง พาณิชย์	เสียง ที่ สำคัญ ต่อ เศรษฐกิจ ไทย
กระทรวง พาณิชย์	กระทรวง พาณิชย์ หอใจ มุอคำ การ คำ ไทย ออสดรณีย์ ที่ ชาย ตัว กว่า ร้อยละ สิบ ี่ ใน ปี ที่ ผ่าน มา โดย มี มุอคำ รวม เกือบ สาม แสน ล้าน บาท
กระทรวง พาณิชย์	เช่นเดียวกับ นาย วิเศษมา เมืองสุช อดีต รัฐมนตรี ช่วย ว่าการ กระทรวง พาณิชย์
กระทรวง พาณิชย์	ใน ระหว่าง ที่ นาย เสรีโกกร จิระเมธย์ รัฐมนตรี ว่าการ กระทรวง พาณิชย์ ไป ร่วม
กระทรวง พาณิชย์	สัมมนา ร่วมมือ ร่วมใจ พัฒนา คำ ปลีก ไทย ครั้ง ที่ สอง ที่ จังหวัด นครราชสีมา
กระทรวง พาณิชย์	รัฐมนตรี ว่าการ กระทรวง พาณิชย์ ยืนยัน ว่า รัฐบาล ได้ พิจารณา ร่าง ข้อตกลง
กระทรวง พาณิชย์	เปิด เขตการค้าเสรี ไทย ญี่ปุ่น อย่าง รอบคอบ แล้ว

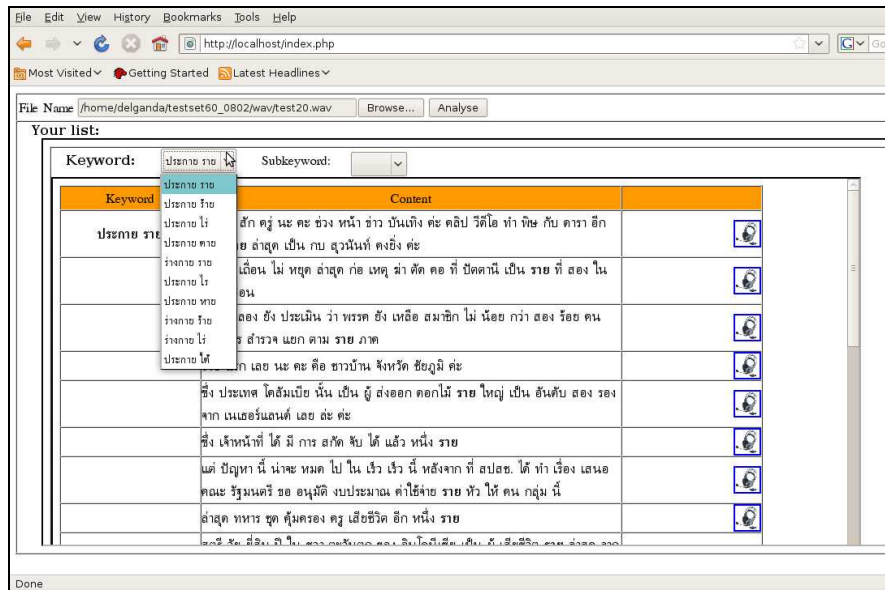
ภาพที่ 4.2

ผลลัพธ์จากการค้นหาคำสำคัญว่า “ประกาศ”

The screenshot shows a web application interface with a search bar at the top. Below the search bar, there is a section titled "Your list:" containing a table with two columns: "Keyword" and "Content". The "Keyword" column lists various terms related to the word "ประกาศ" (Announcement), and the "Content" column displays the corresponding search results. The results are organized into a table with multiple rows and columns, showing a list of keywords and their associated content.

Keyword	Content
ประกาศ	ที่ทางการ ต้อง ประกาศ ภาวะฉุกเฉิน และ อยุทธ ประชชน บาง ส่วน ไป ยัง
ประกาศ	ปออคกัย
ประกาศ	ที่ บริเวณ ด่าน อุนสาวรีย์ ท้าวสุรนารี และ ด่าน เวที กลาง แจ่ง ด้าน ซ้ำ เป็น
ประกาศ	ที่ สำหรับ ชัด งาน การ ประกาศ ผล รางวัล ภาพยนตร์ แห่งชาติ หรือ สุวรรณ
ประกาศ	ทองคำ ครั้ง ที่ สิบ หก ประจำ
ประกาศ	ก็ตาม สำหรับ การ ประกาศ ผล ครั้ง นี้ ชาว บันเทิง โดทวิ จะ ติดตาม ผล มา
ประกาศ	กัน ให้ทราบ อย่าง แน่นนอน ค่ะ
ประกาศ	นาง อนุช ไอศกานนท์ รัฐมนตรี ช่วย ว่าการ กระทรวง พาณิชย์ ประกาศ สะ เร่ง
ประกาศ	ปรับ ภาษีสินค้า และ เปลี่ยน ชื่อ องค์การ ค้างสินค้า
ประกาศ	วันนี้ กระทรวง วัฒนธรรม ประกาศ รายชื่อ เชิดชู เกียรติ พิธป็น แห่งชาติ ปี สอง
ประกาศ	พันห้าร้อย ี่ สิบเก้า จำนวน เก้า ล้าน จาก สาม สาขา ค่ะ
ประกาศ	ตอนนี้ เรา มา ต่อ กัน ที่ การ ประกาศ ผล รางวัล พิธป็น แห่งชาติ ประจำปี สอง
ประกาศ	พันห้าร้อย ี่ สิบเก้า ได้ มี การ แบ่ง ออก เป็น สาม สาขา ด้วย กัน นะ ค่ะ
ประกาศ	ศึกษากาณณ ี่ ำปยุชการ ดำรงแ่ง ชาติ ประกาศ ชิดเส้น ดำรงล้า เกิด ไ้ใหม่

ภาพที่ 4.3
ผลลัพธ์จากการค้นหาคำสำคัญว่า “ประกายไฟ”



จากภาพที่ 4.3 การค้นหาคำสำคัญว่า “ประกายไฟ” จะไม่พบในคำสำคัญทั้ง 10 คำ ดังนั้นจะ
ใช้คำสำคัญย่อยมาช่วยในการค้นหา โดยจัดลำดับความถี่ของคำสำคัญย่อยตามคำสำคัญทั้ง 10
คำ จะแสดงดังตารางที่ 4.5

ตารางที่ 4.5
ตารางแสดงความถี่ของคำสำคัญย่อยตามคำสำคัญว่า “ประกายไฟ” ทั้ง 10 คำ

คำสำคัญย่อย	ความถี่
ประกาย	7
ร่างกาย	3
ไร่	2
ร้าย	2
ราย	2
ไร	1
ได้	1
หาย	1
ตาย	1

ภาพที่ 4.4

ผลลัพธ์จากการจัดลำดับตามความถี่ของคำสำคัญย่อย

File Edit View History Bookmarks Tools Help

http://localhost/index.php

Most Visited Getting Started Latest Headlines

File Name /home/delganda/testset60_0802/wav/test20.wav Browse... Analyse

Your list:

Keyword: ประกาย ราย Subkeyword: ไฟ

Keyword	ประกาย ราย	ไฟ	ประกาย ราย	ไฟ
ประกาย ราย	คำ พัก สัก คุ้ นะ ค่ะ ช่วง พัก	ไฟ	ไฟ	ไฟ
	แล้ว ราย อำสุด เป็น กบ ดูว	ไฟ	ไฟ	ไฟ
	ใจ ได้ เดือน ไม่ หุตุ อำสุด	ไฟ	ไฟ	ไฟ
	รอบ เดือน	ไฟ	ไฟ	ไฟ
	นาย จำลอง ยัง ประเด็น ว่า พ	ไฟ	ไฟ	ไฟ
	จาก การ ส่ววจ แยก ตาม ราย	ไฟ	ไฟ	ไฟ
	ราย แรก เลข นะ ค่ะ คือ ขาว	ไฟ	ไฟ	ไฟ
	ซึ่ง ประเทศ โดสมเบ๊ย นั้น เป็น	ไฟ	ไฟ	ไฟ
	จาก เนเธอร์แลนด์ เลข อ่ะ ค่ะ	ไฟ	ไฟ	ไฟ
	ซึ่ง เจ้าหน้าที่ ได้ มี การ สัก รับ	ไฟ	ไฟ	ไฟ
	แต่ ปัญหา นี้ น่าจะหมด ไป ใน	ไฟ	ไฟ	ไฟ
	คณะ รัฐมนตรี ขอ อนุมัติ	ไฟ	ไฟ	ไฟ
	อำสุด ทหาร ชุด คุ้มครอง	ไฟ	ไฟ	ไฟ

Done

คำสำคัญที่ต้องการค้นหาคือคำว่า “ประกาย ไฟ” ดังนั้นจะลองเลือกคำว่า “ประกาย” ซึ่ง
จะแสดงผลดังนี้

ภาพที่ 4.5

ผลลัพธ์จากการเลือกคำสำคัญย่อยว่า “ประกาย”

The screenshot shows a web browser window with a search application. The address bar shows 'http://localhost/index.php'. Below the browser window, there is a section titled 'Your list:' containing a table of search results. The table has two columns: 'Keyword' and 'Content'. The 'Keyword' column contains the word 'ประกาย'. The 'Content' column contains a paragraph of Thai text. A blue circle highlights the word 'ประกาย' in the text. Below the table, there is a link labeled '1'.

Keyword	Content
ประกาย	เจ้าหน้าที่ ระบุว่า ต้นแมงป่อง นกจาก โรง นก 5 คต ว่า ดูปรกณ ตัว นำ ไปทำ ภาย ใน ชั่วครู่ แล้ว นกประกาย ไหวจะเดิน ไป ติด สารเคมี ที่ ไวไฟ

1

ผลลัพธ์ที่ได้จากการเลือกคำสำคัญย่อยว่า “ประกาย” จะค้นหาข้อมูลประโยคขึ้นมา มีคำว่า “ประกาย ไฟ” ซึ่งตรงกับสิ่งที่เราต้องการค้นหา

อภิปรายผลการทดลอง

จากผลการทดลองทั้งหมดที่ทำมาจะเห็นได้ว่าระบบของ BN-LVCSR แบบจำลองที่ดีที่สุดคือ แบบจำลองในสมมติฐานที่ 2 ได้เปอร์เซ็นต์ความถูกต้องและเปอร์เซ็นต์ความแม่นยำ 78.32%, 74.34% ซึ่งให้ผลลัพธ์ที่ดีกว่างานวิจัยในด้านนี้ที่ผ่านมา ทั้งนี้เปอร์เซ็นต์ความถูกต้อง 78.32% อาจะมาจากการที่ใช้วิธี 10-best และการตัดคำหยุดออกจากพจนานุกรมเสียงอ่าน ซึ่งช่วยให้ความถูกต้องมากขึ้น การเพิ่มข้อมูลข่าวสำหรับการฝึกฝนแบบจำลองภาษาให้เรียนรู้คำศัพท์มากขึ้น ไม่ได้ช่วยให้เปอร์เซ็นต์ความถูกต้องและเปอร์เซ็นต์ความแม่นยำเพิ่มขึ้น แต่ทำให้เปอร์เซ็นต์ความถูกต้องและเปอร์เซ็นต์ความแม่นยำลดลงอีกด้วย เพราะเมื่อมีคำศัพท์มากขึ้นก็จะทำให้เกิดความผิดพลาดมากขึ้นด้วย ส่วน 22.68% ที่มีความผิดพลาดนั้นอาจจะเกิดจากชุดข้อมูลที่ใช้ในการทดสอบนั้นเป็นเพียงคำวลีสั้นๆ และมีขนาดเล็ก ทำให้คำศัพท์ที่ใช้ในการทดสอบมีจำนวนน้อย ดังนั้นถ้าเพิ่มชุดข้อมูลทดสอบให้มีขนาดใหญ่ขึ้น คำศัพท์มากขึ้น และลักษณะการพูดให้เป็นประโยคยาวๆ อาจจะทำให้มีเปอร์เซ็นต์ความถูกต้องและเปอร์เซ็นต์ความแม่นยำเพิ่มมากขึ้นได้

ในส่วนของการค้นหาคำสำคัญนั้น จะมีการใช้คำสำคัญย่อยมาช่วยในการค้นหาคำสำคัญเพื่อให้ตรงกับความต้องการของผู้ใช้มากที่สุด โดยคำสำคัญย่อยนั้นจะเรียงลำดับตามความถี่ของเกิดคำ ถ้าความถี่ของคำสำคัญย่อยมีมากอาจจะสันนิษฐานได้ว่าคำสำคัญที่ต้องการค้นหานั้น มีคำสำคัญย่อยที่มีความถี่สูงสุดเป็นองค์ประกอบก็ได้ ดังนั้นการใช้คำสำคัญย่อยมาช่วยในการค้นหาคำสำคัญนั้นจะช่วยเพิ่มประสิทธิภาพในการค้นหามากยิ่งขึ้น