

บทที่ 4

กรรมวิธีการเรียนรู้คำกำกวม

ประเภทของข้อความกำกวม

คำกำกวมในงานวิจัยนี้หมายถึงแต่ละแบบที่เป็นไปได้จากการตัดคำของข้อความกำกวมเช่น มากกว่า สามารถตัดคำได้เป็น มา กว่า กับ มาก ว่า ในที่นี้จะเรียก มา กว่า ว่าเป็นหนึ่งคำกำกวมและมากกว่า เป็นอีกหนึ่งคำกำกวม ซึ่งการตัดคำจะเป็น มา กว่า หรือ มาก ว่า นั้นจะต้องดูคำบริบทหรือคำที่อยู่ติดกันว่าควรจะถูกตัดคำได้เป็นแบบใด หรือข้อความ บอกว่า สามารถตัดคำได้เป็น บอก ว่า กับ บอก กว่า ซึ่งคำกำกวม บอก กว่า นั้นโอกาสที่จะเกิดขึ้นมีน้อยกว่าคำกำกวม บอก ว่า มากดังนั้นจึงสามารถแบ่งข้อความกำกวมออกเป็น 2 ประเภทคือ

1. แบบไม่ขึ้นอยู่กับบริบท (Context Independent) การตัดคำจากข้อความกำกวมแบบนี้ไม่จำเป็นต้องพึ่งบริบท เนื่องจากมีเพียงคำกำกวมคำเดียวที่เป็นไปได้ซึ่งคำอื่นมีโอกาสเกิดขึ้นได้น้อยมาก ซึ่งอาจจะเรียกข้อความกำกวมแบบนี้ได้ว่าเป็นข้อความที่ไม่กำกวมที่แท้จริง เช่นข้อความ กรอบ สามารถตัดคำได้เป็น กร อบ กับ กรอบ ซึ่งคำกำกวม กรอบ มีโอกาสที่เกิดขึ้นมากกว่า กร อบ อย่างไรก็ตามวิธีการตัดคำให้จำนวนน้อยที่สุดก็ยังตัดข้อความกำกวมแบบนี้ไม่ถูกต้อง 100% เช่น พอกกลางคืน ตัดคำได้เป็น พอก ลาง คืน ซึ่งที่ถูกต้องคือ พอ กลาง คืน ยานอก ตัดคำได้เป็น ยาน ออก ที่ถูกต้องคือ ยาน ออก

2. แบบขึ้นอยู่กับบริบท (Context Dependent) คำกำกวมที่เกิดขึ้นจากการตัดคำจากข้อความแบบนี้มีความเป็นไปได้ทุกคำ ดังนั้นจะต้องดูบริบทช่วยในการตัดคำ ถึงแม้ว่าข้อความกำกวมแบบนี้จะมีน้อยกว่าแบบแรก แต่แบบนี้จะมีความผิดพลาดจากการตัดคำได้มากกว่าเนื่องจากทุกคำกำกวมมีความเป็นไปได้ทุกคำ เช่น

มากกว่า	สามารถตัดคำได้เป็น	มาก ว่า	กับ	มา กว่า
ตากลม	สามารถตัดคำได้เป็น	ตา ก ลม	กับ	ตา กลม
ตัวเล็ก	สามารถตัดคำได้เป็น	ตัว เล็ก	กับ	ตัว เล็ก
ทางด้าน	สามารถตัดคำได้เป็น	ทาง ด้าน	กับ	ทาง ด้าน

การเรียนรู้คำกำกวมโดยใช้ระบบ FOIL และ RIPPER

1. ข้อมูลที่ใช้ในการเรียนรู้

ข้อมูลที่ใช้ในการเรียนรู้คำกำกวมและโปรแกรมการตัดคำแบบจำลองโครงข่ายประสาทเทียมในงานวิจัยนี้ได้มาจากคุณไพศาล เจริญพรสวัสดิ์ ที่ทำงานที่ห้องปฏิบัติการวิจัยและพัฒนาภาษาและซอฟต์แวร์ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ ซึ่งในงานวิจัยนี้ใช้ 4950 ประโยคซึ่งในแต่ละประโยคจะถือว่ามีความกำกวมที่ต้องการจะเรียนรู้เพียงคำเดียวในประโยคเท่านั้นส่วนคำอื่นๆถือว่าไม่เป็นคำกำกวม และในแต่ละประโยคได้ทำการแบ่งคำและประเภทย่อยของคำไว้เรียบร้อยแล้ว ข้อมูลที่ใช้ในการเรียนรู้คำกำกวมแต่ละคำแบ่งออกเป็นสองกลุ่มคือ กลุ่มข้อมูลสอนมีจำนวนร้อยละ 80 และกลุ่มข้อมูลทดสอบกฎที่เรียนรู้ได้จำนวนร้อยละ 20 และเก็บข้อมูลที่จะเรียนรู้แต่ละคำกำกวมไว้ในแต่ละแฟ้มข้อมูล

เนื่องจากในงานวิจัยนี้ได้ใช้คุณลักษณะของประเภทของคำและประเภทย่อยของคำ โดยประเภทและประเภทย่อยของคำได้มีงานวิจัยทางภาษาศาสตร์ [11] ทำการแบ่งประเภทของคำในภาษาไทยตามหน้าที่ของคำที่แตกต่างกันออกไปเป็น 14 ประเภทซึ่งมีทั้งหมด 47 ประเภทย่อยดังนี้

ประเภทของคำ	ประเภทย่อยของคำ	รายละเอียด	ตัวอย่าง
NOUN	NPRP	Proper noun	วินโดวส์95, โคโรนา, ไค้ก, พระอาทิตย์
	NCNM	Cardinal number	หนึ่ง, สอง, สาม, 1, 2, 3
	NONM	Ordinal number	ที่หนึ่ง, ที่สอง, ที่1, ที่2, ที่3
	NLBL	Label noun	1, 2, 3, 4, ก, ข, a, b
	NCMN	Common noun	หนังสือ, อาหาร, อาคาร, คน
	NTTL	Title noun	ดร., พลเอก
PRONOUN	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
VERB	VACT	Active verb	ทำงาน, ร้องเพลง, กิน
	VSTA	Stative verb	เห็น, รู้, คือ
	VATT	Attribute verb	อ้วน, ดี, สวย
AUXILIARY	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

ตารางที่ 4.1 ประเภทของคำ

ประเภทของคำ	ประเภทย่อยของคำ	รายละเอียด	ตัวอย่าง
AUXILIARY	XVBB	Pre-verb auxiliary, in imperative mood	กรุณา, จง, เชิญ, อย่า, ห้าม
	XVAE	Post-verb auxiliary	ไป, มา, ขึ้น
DETERMINER	DDAN	Definite determiner, after noun without classifier in between	นี้, นั้น, โน่น, ทั้งหมด
	DDAC	Definite determiner, allowing classifier in between	นี้, นั้น, โน่น, อยู่น
	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, เลือ 2 ตัว
	DONM	Determiner, ordinal number expression	ที่หนึ่ง, ที่สอง, ที่สุดท้าย
	ADVERB	ADVN	Adverb with normal form
ADVI		Adverb with iterative form	เร็วๆ, เสมอๆ, ช้าๆ

ตารางที่ 4.1 ประเภทของคำ (ต่อ)

ประเภทของคำ	ประเภทย่อยของคำ	รายละเอียด	ตัวอย่าง
ADVERB	ADVP	Adverb with prefixed form	โดยเร็ว
	ADVS	Sentential adverb	โดยปกติ, ธรรมดา
CLASSIFIER	CNIT	Unit classifier	ตัว, คน, เล่ม
	CLTV	Collective classifier	คู่, กลุ่ม, ผุ้, เชิง, ทาง, ด้าน, แบบ, รุ่น
	CMTR	Measurement classifier	กิโลกรัม, แก้ว, ชั่วโมง
	CFQC	Frequency classifier	ครั้ง, เทียว
	CVBL	Verbal classifier	ม้วน, มัด
CONJUNCTION	JCRG	Coordinating conjunction	และ, หรือ, แต่
	JCMP	Comparative conjunction	กว่า, เหมือนกับ, เท่ากับ
	JSBR	Subordinating conjunction	เพราะว่า, เนื่องจาก, ที่, แม้ว่า, ถ้า
PREPOSITION	RPRE	Preposition	จาก, ละ, ของ, ได้, บน
INTERJECTION	INT	Interjection	ไ้ย, ไ้, เออ, เอ้, อ้อ
PREFIX	FIXN	Nominal prefix	การทำงาน, ความสนุกสนาน
	FIXV	Adverbial prefix	อย่างรวดเร็ว
SENTENCE PARTICLE	EAFF	Ending for affirmative sentence	จ้ะ, จ้ะ, ค่ะ, ครับ, นะ, นำ, เถอะ
	EITT	Ending for interrogative sentence	หรือ, เหนอ, ไหม, มั้ย
NEGATOR	NEG	Negator	ไม่, มิได้, ไม่ได้, มิ
PUNCTUATION	PUNC	Punctuation	(.), " . . . :

ตารางที่ 4.1 (ต่อ)

2. ข้อมูลตัวอย่าง

ข้อมูลตัวอย่างแต่ละตัวอย่างประกอบด้วย 10 คุณลักษณะเรียงตามลำดับดังนี้คือ กลุ่มคำที่ 3 ถึงคำที่ 10 ก่อนหน้า คำที่ 2 ก่อนหน้า คำที่ 1 ก่อนหน้า คำที่ 1 ข้างหลัง คำที่ 2 ข้างหลัง กลุ่มคำที่ 3 ถึงคำที่ 10 ข้างหลัง ประเภทย่อยของคำที่ 2 ก่อนหน้า ประเภทย่อยของคำที่ 1 ก่อนหน้า ประเภทย่อยของคำที่ 1 ข้างหลัง และประเภทย่อยของคำที่ 2 ข้างหลังของคำกำกวมที่ต้องการเรียนรู้เช่น การเรียนรู้คำว่า มา กว่า จากประโยคที่อยู่ในรูป คำ/ประเภทย่อยของคำ และ BR แทนช่องว่าง 1 ช่องว่าง

โดยมีข้อมูลที่จะนำมาเรียนรู้ดังรูปที่ 4.1 ซึ่งจะนำไปเขียนข้อมูลในแฟ้มข้อมูลเพื่อที่จะนำไปเรียนรู้ด้วยระบบ FOIL ดังแสดงในรูปที่ 4.2 และข้อมูลในแฟ้มข้อมูลกับแฟ้มชื่อสำหรับการเรียนรู้ด้วยระบบ RIPPER แสดงในรูปที่ 4.3

ประเทศไทย/NPRP มี/VACT การ/FIXN ใช้/VACT ตัวอักษร/NCMN ภาษาไทย/NPRP มา/XVAE
กว่า/DIAQ BR/PUNC 700/DCNM BR/PUNC ปี/CMTR แล้ว/XVAE //
ดังนั้น/JSBR ทาง/RPRE กลุ่ม/NCMN วิจัย/VACT ไบโอเซนเซอร์/NPRP BR/PUNC สถาบัน
เทคโนโลยีพระจอมเกล้า/NPRP ธนบุรี/NPRP ซึ่ง/PREL มี/VSTA ประสพการณ์/NCMN ใน/RPRE
งาน/NCMN วิจัย/VACT ด้าน/NCMN นี้/DDAC มา/XVAE กว่า/JCMP BR/PUNC 8/DCNM
BR/PUNC ปี/CMTR จึง/XVBM ได้/XVAM ดำเนินการ/VACT วิจัย/VACT และ/JCRG พัฒนา/VACT
ไบโอเซนเซอร์/NPRP สำหรับ/RPRE วิเคราะห์/VACT ปริมาณ/NCMN ออร์กาโนฟอสฟอรัส/NPRP
ขึ้น/XVAE //
ใน/RPRE ด้าน/NCMN สิ่งแวดล้อม/NCMN นั้น/DDAC BR/PUNC ถูก/XVAM กล่าวถึง/VACT
มาก/ADVN ว่า/JSBR เป็น/VSTA สาขา/NCMN ที่/PREL มี/VSTA ข้อจำกัด/NCMN ทาง
สังคม/NCMN มาก/ADVN BR/PUNC เมื่อ/JSBR เทียบ/VSTA กับ/RPRE ด้าน/NCMN อื่นๆ/DIAC //

รูปที่ 4.1 ตัวอย่างข้อมูลในการเรียนรู้คำว่า มา_กว่า

Word: ประเทศไทย,มี,การ,ใช้,ตัวอักษร,ภาษาไทย,<space>,700,ปี,แล้ว,
สถาบันเทคโนโลยีพระจอมเกล้า,ธนบุรี,ซึ่ง,ประสพการณ์,ใน,งาน,วิจัย,ด้าน,นี้,8,จึง,ได้,ดำเนินการ,และ,
พัฒนา,สิ่งแวดล้อม,นั้น,ถูก,กล่าวถึง,เป็น,สาขา,ที่,ข้อจำกัด,ทางสังคม,มาก,เมื่อ,เทียบ.
Tag: ?, NPRP, NCNM, NONM, NLBL, NCMN, NTTL,
PPRS, PDMN, PNTR, PREL, VACT, VSTA, VATT,
XVBM, XVAM, XVMM, XVBB, XVAE,
DDAN, DDAC, DDBQ, DDAQ, DIAC, DIBQ, DIAQ, DCNM, DONM,
ADVN, ADVI, ADVP, ADVS, CNIT, CLTV, CMTR, CFQC, CVBL,
JCRG, JCMP, JSBR, RPRE, INT, FIXN, FIXV,
EAFF, EITT, NEG, PUNC.
List: [ประเทศไทย มี การ ใช้], [<space> ปี แล้ว],
[สถาบันเทคโนโลยีพระจอมเกล้า ธนบุรี ซึ่ง มี ประสพการณ์ ใน งาน วิจัย],
[<space> ปี จึง ได้ ดำเนินการ วิจัย และ พัฒนา], [ใน ด้าน สิ่งแวดล้อม นั้น <space>],
[ที่ มี ข้อจำกัด ทางสังคม มาก <space> เมื่อ เทียบ].

รูปที่ 4.2 ข้อมูลในแฟ้มข้อมูลที่จะนำไปเรียนรู้คำว่า มา_กว่า ด้วยระบบ FOIL

มา_กว่า(List,Word,Word,Word,Word,List,Tag,Tag,Tag,Tag)
 [ประเทศไทย มี การ ใช้,ตัวอักษร,ภาษาไทย,<space>,700,[<space> ปี แล้ว],NCMN,NPRP,PUNC,DCNM
 [สถาบันเทคโนโลยีพระจอมเกล้า ธนบุรี ซึ่ง มี ประสบการณ์ ใน งาน วิจัย],ด้าน,นี้,<space>,8,[<space> ปี จึง
 ได้ ดำเนินการ วิจัย และ พัฒนา],NCMN,DDAC,PUNC,DCNM
 .
 [ใน ด้าน สิ่งแวดล้อม นั้น <space>],ถูก,กล่าวถึง,เป็น,สาขา,[ที่ มี ข้อจำกัด ทางสังคม มาก <space> เมื่อ
 เทียบ],XVAM,VACT,VSTA,NCMN

รูปที่ 4.2 ข้อมูลในแฟ้มข้อมูลที่จะนำไปเรียนรู้คำว่า มา_กว่า ด้วยระบบ FOIL (ต่อ)

จากรูปที่ 4.2 จะประกอบด้วยคำทุกคำที่มีในข้อมูลตัวอย่างที่จะนำไปเรียนรู้ ประเภทย่อย
 ของคำทุกประเภทย่อย และกลุ่มของคำที่3ถึงคำที่10ทั้งก่อนหน้าและข้างหลังของข้อมูลตัวอย่างที่จะ
 นำไปเรียนรู้ และในส่วนถัดมาที่เขียนว่า มา_กว่า(List,Word,Word,Word,Word,List,Tag,Tag,Tag,Tag)
 หมายถึงชื่อคำกำกวมที่ต้องการเรียนรู้แล้วตามด้วยคุณลักษณะทั้ง 10 คุณลักษณะดังที่กล่าวมาแล้ว
 ในเครื่องหมายวงเล็บ ส่วนเครื่องหมาย ; จะทำการแบ่งข้อมูลที่เป็นตัวอย่างบวกที่อยู่ข้างบนเครื่องหมาย
 หมายออกจากข้อมูลที่เป็นตัวอย่างลบที่อยู่ข้างล่างของเครื่องหมาย

ประเทศไทย มี การ ใช้,ตัวอักษร,ภาษาไทย,space,700,space ปี แล้ว,NCMN,NPRP,PUNC,DCNM,มา_กว่า.
 สถาบันเทคโนโลยีพระจอมเกล้า ธนบุรี ซึ่ง มี ประสบการณ์ ใน งาน วิจัย,ด้าน,นี้,space,8,space ปี จึง ได้ ดำเนิน
 การ วิจัย และ พัฒนา,NCMN,DDAC,PUNC,DCNM, มา_กว่า.
 ใน ด้าน สิ่งแวดล้อม นั้น space,ถูก,กล่าวถึง,เป็น,สาขา,ที่ มี ข้อจำกัด ทางสังคม มาก space เมื่อ
 เทียบ,XVAM,VACT,VSTA,NCMN,มา_กว่า.

รูปที่ 4.3 ข้อมูลในแฟ้มข้อมูลที่จะนำไปเรียนรู้คำว่า มา_กว่า ด้วยระบบ RIPPER

มา_กว่า:
 pw310 : set.
 pw2 : atom.
 pw1 : atom.
 mw1 : atom.
 mw2 : atom.
 mw310 : set.
 pt2 : atom.
 pt1 : atom.

รูปที่ 4.4 ข้อมูลในแฟ้มชื่อที่จะนำไปเรียนรู้คำว่า มา_กว่า ด้วยระบบ RIPPER

mt1 : atom.
mt2 : atom.

รูปที่ 4.4 ข้อมูลในแฟ้มชื่อที่จะนำไปเรียนรู้คำว่า มา_กว่า ด้วยระบบ RIPPER (ต่อ)

ในรูปที่ 4.4 ประกอบด้วยชื่อคำกำกวมที่ต้องการเรียนรู้ ส่วนข้อความในบรรทัดถัดมาจะเป็นการนิยามชนิดข้อมูลของแต่ละคุณลักษณะที่ใช้ในแฟ้มข้อมูลโดยเรียงตามลำดับดังนี้คือ pw310 ใช้แทนกลุ่มของคำที่ 3 ถึงคำที่ 10 ก่อนหน้าและ mw310 ใช้แทนกลุ่มคำที่ 3 ถึงคำที่ 10 ข้างหลังของคำว่า มา_กว่า มีชนิดของข้อมูลเป็นแบบ set pw2 ใช้แทนคำที่ 2 ก่อนหน้า pw1 ใช้แทนคำที่ 1 ก่อนหน้า mw1 ใช้แทนคำที่ 1 ข้างหลัง และ mw2 ใช้แทนคำที่ 2 ข้างหลังของคำว่า มา_กว่า ซึ่งมีชนิดของข้อมูลเป็นแบบ atom ของคำว่า มา_กว่า ส่วน pt2 pt1 mt1 mt2 ใช้แทนประเภทย่อยของคำของคำที่ 2 ก่อนหน้า คำที่ 1 ก่อนหน้า คำที่ 1 ข้างหลัง คำที่ 2 ข้างหลังตามลำดับ ซึ่งมีชนิดของข้อมูลเป็นแบบเดียวกันคือ atom

3. ความรู้ภูมิหลัง ระบบ RIPPER ไม่มีการใช้ความรู้ภูมิหลังในการสร้างกลุ่มของกฎจึงไม่ต้องกำหนดความรู้ภูมิหลังในการเรียนรู้ ส่วนระบบ FOIL มีการใช้ความรู้ภูมิหลังในการสร้างอนุประโยคผลลัพธ์ และผลลัพธ์ที่ได้จากการเรียนรู้คำกำกวมต้องการออกมาเป็นคำ ประเภทของคำหรือประเภทย่อยของคำ กับระบบ FOIL ไม่สามารถเรียนรู้กลุ่มของคำคงที่ของคุณลักษณะได้เหมือนกับ RIPPER ดังนั้นความรู้ภูมิหลังของตัวอย่างในการเรียนรู้จึงต้องกำหนดความสัมพันธ์ของคำประเภทของคำ และประเภทย่อยของคำ โดยที่ความสัมพันธ์ของประเภทของคำและความสัมพันธ์ของประเภทย่อยของคำ จะเหมือนกันทุกคำกำกวมที่จะนำไปเรียนรู้ซึ่งจะมีความสัมพันธ์ของประเภทของคำ 10 ประเภทของคำและความสัมพันธ์ของประเภทย่อยของคำจะมี 47 ประเภทย่อยของคำดังที่แสดงในตารางที่ 4.1 ส่วนความสัมพันธ์ของคำและความสัมพันธ์ member ที่ใช้สำหรับหาคุณลักษณะของคำที่เป็นกลุ่มของคำที่ 3 ถึงคำที่ 10 ก่อนหน้าและหลังคำกำกวมจะขึ้นอยู่กับข้อมูลที่จะนำไปใช้ในการเรียนรู้คำกำกวมแต่ละคำ เช่นตัวอย่างความรู้ภูมิหลังในการเรียนรู้คำกำกวม มา_กว่า แสดงดังรูปที่ 4.5 (รายละเอียดแสดงในภาคผนวก ค)

*การ(Word)
การ
.
*ประเทศไทย(Word)
ประเทศไทย
.
...

รูปที่ 4.5 ตัวอย่างความรู้ภูมิหลังในแฟ้มข้อมูลในการเรียนรู้คำว่า มา_กว่า ด้วยระบบ FOIL

*เมื่อ(Word)
เมื่อ
•
*เทียบ(Word)
เทียบ
•
*NPRP(Tag)
NPRP
•
*NCNM(Tag)
NCNM
•

*NEG(Tag)
NEG
•
*PUNC(Tag)
PUNC
•
*NOUN(Tag)
NPRP
NCNM
NONM
NLBL
NCMN
NTTL
•
*PRONOUN(Tag)
PPRS
PDMN
PNTR
PREL
•

รูปที่ 4.5 ตัวอย่างความรู้ภูมิหลังในเพิ่มข้อมูลในการเรียนรู้คำว่า มา_กว่า ด้วยระบบ FOIL (ต่อ)

...
*PREFIX(Tag)
FIXN
FIXV
.
*PARTICLE(Tag)
EAFF
EITT
.
*member(Word,List)
ประเทศไทย,[ประเทศไทย มี การ ใช้]
มี,[ประเทศไทย มี การ ใช้]
การ,[ประเทศไทย มี การ ใช้]
ใช้,[ประเทศไทย มี การ ใช้]
<space>,[<space> ปี แล้ว]
ปี,[<space> ปี แล้ว]
แล้ว,[<space> ปี แล้ว]
สถาบันเทคโนโลยีพระจอมเกล้า,[สถาบันเทคโนโลยีพระจอมเกล้า ธนบุรี ซึ่ง มี ประสบการณ์ ใน งาน วิจัย]
ธนบุรี,[สถาบันเทคโนโลยีพระจอมเกล้า ธนบุรี ซึ่ง มี ประสบการณ์ ใน งาน วิจัย]
ซึ่ง,[สถาบันเทคโนโลยีพระจอมเกล้า ธนบุรี ซึ่ง มี ประสบการณ์ ใน งาน วิจัย]
มี,[สถาบันเทคโนโลยีพระจอมเกล้า ธนบุรี ซึ่ง มี ประสบการณ์ ใน งาน วิจัย]
ประสบการณ์,[สถาบันเทคโนโลยีพระจอมเกล้า ธนบุรี ซึ่ง มี ประสบการณ์ ใน งาน วิจัย]
ใน,[สถาบันเทคโนโลยีพระจอมเกล้า ธนบุรี ซึ่ง มี ประสบการณ์ ใน งาน วิจัย]
งาน,[สถาบันเทคโนโลยีพระจอมเกล้า ธนบุรี ซึ่ง มี ประสบการณ์ ใน งาน วิจัย]
วิจัย,[สถาบันเทคโนโลยีพระจอมเกล้า ธนบุรี ซึ่ง มี ประสบการณ์ ใน งาน วิจัย]
<space>,[<space> ปี จึง ได้ ดำเนินการ วิจัย และ พัฒนา]
ปี,[<space> ปี จึง ได้ ดำเนินการ วิจัย และ พัฒนา]
จึง,[<space> ปี จึง ได้ ดำเนินการ วิจัย และ พัฒนา]
ได้,[<space> ปี จึง ได้ ดำเนินการ วิจัย และ พัฒนา]
ดำเนินการ,[<space> ปี จึง ได้ ดำเนินการ วิจัย และ พัฒนา]
วิจัย,[<space> ปี จึง ได้ ดำเนินการ วิจัย และ พัฒนา]
และ,[<space> ปี จึง ได้ ดำเนินการ วิจัย และ พัฒนา]
พัฒนา,[<space> ปี จึง ได้ ดำเนินการ วิจัย และ พัฒนา]
ใน,[ใน ด้าน สิ่งแวดล้อม นั้น <space>]

รูปที่ 4.5 ตัวอย่างความรู้ภูมิหลังในแฟ้มข้อมูลในการเรียนรู้คำว่า มา_กว่า ด้วยระบบ FOIL (ต่อ)

ด้าน,[ใน ด้าน สิ่งแวดล้อม นั้น <space>
 สิ่งแวดล้อม,[ใน ด้าน สิ่งแวดล้อม นั้น <space>
 นั้น,[ใน ด้าน สิ่งแวดล้อม นั้น <space>
 <space>,[ใน ด้าน สิ่งแวดล้อม นั้น <space>
 ที่,[ที่ มี ข้อจำกัด ทางสังคม มาก <space> เมื่อ เทียบ]
 มี,[ที่ มี ข้อจำกัด ทางสังคม มาก <space> เมื่อ เทียบ]
 ข้อจำกัด,[ที่ มี ข้อจำกัด ทางสังคม มาก <space> เมื่อ เทียบ]
 ทางสังคม,[ที่ มี ข้อจำกัด ทางสังคม มาก <space> เมื่อ เทียบ]
 มาก,[ที่ มี ข้อจำกัด ทางสังคม มาก <space> เมื่อ เทียบ]
 <space>,[ที่ มี ข้อจำกัด ทางสังคม มาก <space> เมื่อ เทียบ]
 เมื่อ,[ที่ มี ข้อจำกัด ทางสังคม มาก <space> เมื่อ เทียบ]
 เทียบ,[ที่ มี ข้อจำกัด ทางสังคม มาก <space> เมื่อ เทียบ]

รูปที่ 4.5 ตัวอย่างความรู้อิมิตหลังในแฟ้มข้อมูลในการเรียนรู้คำว่า มา_กว่า ด้วยระบบ FOIL (ต่อ)

ผลการเรียนรู้

กฎที่ได้จากการเรียนรู้โดยระบบ FOIL และ RIPPER จะสามารถเขียนอยู่ในรูปเงื่อนไขได้เหมือนกัน เช่นกฎที่ได้จากการเรียนรู้คำกำกวมคำว่า มา_กว่า ดังนี้

FOIL

มา_กว่า:

มา_กว่า(A,B,C,D,E,F,G,H,I,J) :- DCNM(J).

ตาก_ลม:

ตาก_ลม(A,B,C,D,E,F,G,H,I,J) :- VERB(G).

ตาก_ลม(A,B,C,D,E,F,G,H,I,J) :- NCMN(G).

ตาก_ลม(A,B,C,D,E,F,G,H,I,J) :- PRONOUN(I).

ตาก_ลม(A,B,C,D,E,F,G,H,I,J) :- ยีน(C).

สามารถเขียนกฎที่ได้จากการเรียนรู้ได้ดังนี้

if (ประเภทของคำที่2 ข้างหลังเป็น DCNM) then มา_กว่า.

if ((ประเภทของคำที่1 ข้างหน้าเป็น VERB) or (ประเภทของคำที่1 ข้างหน้าเป็น NCMN) or

(ประเภทของคำที่1 ข้างหลังเป็น PRONOUN) or (คำข้างหน้าคำที่1เป็นคำว่า ยีน)) then ตาก_ลม.

โดยที่ VERB เป็นประเภทของคำที่จะรวมถึงประเภทย่อยของคำได้แก่ VACT VSTA และ VATT ด้วย และ PRONOUN ก็จจะรวมประเภทย่อยของคำได้แก่ PPRS PDMN PNTR และ PREL ด้วย

RIPPER

'มา_กว่า':-mt2=DCNM.

'ตาก_ลม' :- pw1!=มี, pw1!=เบิก' (22/0).

กฎที่ได้จากการเรียนรู้สามารถเขียนได้ดังนี้

if (ประเภทของคำที่2 ข้างหลังเป็น DCNM) then มา_กว่า.

if ((คำข้างหน้าคำที่1 ไม่ใช่คำว่า มี) and (คำข้างหน้าคำที่1 ไม่ใช่คำว่า เบิก)) then ตาก_ลม.

จากการนำกฎที่ได้จากการเรียนรู้ไปใช้ในการตัดคำกำกวมพบว่ากฎที่ได้ไม่สามารถช่วยการตัดคำกำกวมได้ในบางประโยค เช่น บริษัทเเปิดบริการมานานับ 10 ปี เมื่อผ่านการตัดคำแล้วจะพบว่ากฎที่ได้จากการเรียนรู้ไม่สามารถตัดคำให้ได้เป็นคำว่า มา_กว่า ได้ เนื่องจากประเภทย่อยของคำที่ 2 ข้างหลังไม่ใช่ DCNM แต่เป็น PUNC แทนซึ่งเป็นประเภทย่อยของช่องว่าง

จากการให้ข้อมูลตัวอย่างและกฎที่ได้จากการเรียนรู้ของระบบ FOIL และ RIPPER จะเห็นได้ว่าระบบ RIPPER สามารถเรียนรู้คุณลักษณะที่มีลักษณะเป็นกลุ่มคำของคำที่ 3 ถึงคำที่ 10 ก่อนหน้าและหลังได้โดยไม่ต้องมาใส่เป็นความรู้ภูมิหลังในการเรียนรู้ของระบบ FOIL แต่กฎที่ได้จากการเรียนรู้ด้วยระบบ FOIL จะเป็นกฎที่นิยามได้กว้างกว่า คือสามารถนิยามเป็นคำ ประเภทของคำ และประเภทย่อยของคำได้ ในขณะที่ RIPPER จะนิยามได้เฉพาะคำ และประเภทย่อยของคำเท่านั้น

ผลที่ได้จากการเรียนรู้คำกำกวมทั้งหมด 150 คำ (รายละเอียดแสดงในภาคผนวก ง) ของทั้งสองระบบวัดเป็นค่าเฉลี่ยร้อยละของจำนวนประโยคที่ทำการตัดคำกำกวมถูกต้องจากจำนวนประโยคทั้งหมด ของกลุ่มข้อมูลสอนกับกลุ่มข้อมูลทดสอบ ดังแสดงในตารางที่ 4.2

	กลุ่มข้อมูลสอน (Training Set)	กลุ่มข้อมูลทดสอบ (Test Set)
ระบบ FOIL	94.92 %	77.13 %
ระบบ RIPPER	92.79 %	70.08 %

ตารางที่ 4.2 ผลที่ได้จากการเรียนรู้

การพัฒนาโมดูลย่อยจากการเรียนรู้โดยใช้ระบบ FOIL

โปรแกรมการตัดคำด้วยวิธีการตัดคำแบบจำลองไตรแกรมที่นำมาใช้ในงานวิจัยนี้ได้ใช้ขั้นตอนวิธีวิเทอร์บี (Viterbi Algorithm) มาประยุกต์ใช้กับการกำกับประเภทย่อยของคำของแบบจำลองไตรแกรม [3] ซึ่งการทำงานของขั้นตอนวิธีวิเทอร์บีแสดงในรูปที่ 4.1 โดยจะมีการสร้างแถวลำดับ (array) ขนาด $N \times N \times T$ จำนวน 2 ชุดโดย N คือจำนวนประเภทย่อยของคำที่เป็นไปได้ทั้งหมด และ T คือจำนวนคำในประโยคที่จะนำมากำกับประเภทย่อยของคำ โดยที่แถวลำดับชุดแรกคือแถวลำดับ $seqscore[i][j][t]$ ซึ่งจะทำการเก็บค่าความน่าจะเป็นที่ดีที่สุดของการกำกับประเภทย่อยของคำของ w_1, \dots, w_t โดยคำที่ w_t กับ w_{t-1} จะมีประเภทย่อยของคำเป็น L_i และ L_j ตามลำดับ ส่วนแถวลำดับ

ชุดที่สองคือ $\text{backptr}[i][j][t]$ จะเก็บประเภทย่อยของคำของคำ $t-2$ เมื่อคำที่ t และ $t-1$ มีประเภทย่อยของคำเป็น L_i และ L_j ตามลำดับ

การพัฒนาโมดูลย่อยจากการเรียนรู้โดยใช้ระบบ FOIL เป็นการนำกลุ่มของอนุประโยคผลลัพธ์ที่ได้จากการเรียนรู้ของแต่ละคำกำกวมของกลุ่มข้อมูลสอน ซึ่งในงานวิจัยนี้เรียนรู้คำกำกวมทั้งหมดจำนวน 150 คำ มาเป็นเงื่อนไขช่วยในการตัดคำกำกวมในแต่ละประโยคของประโยคแบบต่างๆที่คะแนนดีที่สุด N ประโยคที่ผ่านการตัดคำด้วยวิธีการตัดคำแบบจำลองไตรแกรม ที่มีคำในพจนานุกรมจำนวน 10,000 คำและคำที่มีความน่าจะเป็นของแต่ละคำในแต่ละประเภทของคำจำนวน 136,922 คำ จากประโยคนำเข้า 1 ประโยคโดยที่แต่ละประโยคนั้นถูกตัดคำออกเป็นคำกับประเภทย่อยของคำ โดยจะให้คะแนนแต่ละประโยคของ N ประโยคที่ได้จากการตัดคำด้วยวิธีแบบจำลองไตรแกรมมีค่าเท่ากับ 0 และเมื่อตรวจสอบพบว่าในประโยคมีคำกำกวมและเป็นไปตามเงื่อนไขที่เรียนรู้ด้วย FOIL แล้วจะทำการบวกคะแนนของประโยคด้วย 1 สำหรับแต่ละคำกำกวมแล้วจะทำการเรียงลำดับคะแนนของแต่ละประโยคจากมากไปน้อย ประโยคที่เป็นคำตอบจะเป็นประโยคที่มีคะแนนมากที่สุด แต่ถ้าคะแนนของประโยคมีค่าเท่ากันจะใช้ค่าความน่าจะเป็นของประโยคเป็นตัวตัดสินแทน และประโยคที่จะเป็นคำตอบจะเป็นประโยคที่มีค่าความน่าจะเป็นมากที่สุด ซึ่งการวัดความถูกต้องการตัดคำกำกวมเป็นการดูเฉพาะคำกำกวมในประโยคที่เป็นคำตอบเท่านั้น โดยถือว่าประโยค 1 ประโยคจะมีคำกำกวมเพียงคำเดียวเท่านั้น ดังแสดงผลในตารางที่ 4.3

กำหนดให้ w_1, \dots, w_T เป็นลำดับคำในประโยค L_1, \dots, L_n เป็นประเภทย่อยของคำที่เป็นไปได้ $\text{Prob}(w_i | L_i)$ คือค่าความน่าจะเป็นของคำศัพท์ w_i เมื่อกำหนดให้มีประเภทย่อยของคำเป็น L_i และค่าความน่าจะเป็นของไตรแกรมคือ $\text{Prob}(L_k | L_i, L_j)$ ดังนั้นให้หาลำดับของประเภทย่อยของคำ C_1, \dots, C_T ที่เป็นของลำดับคำในประโยคที่มีความน่าจะเป็นมากที่สุด

Initialization Step

for $i=1$ to N do

for $j=1$ to N do

$$\text{seqscore}[i][j][1] = \text{Prob}(W_1 | L_i) \times \text{Prob}(L_i | \phi) \times \text{Prob}(W_2 | L_j) \\ \times \text{Prob}(L_j | L_i, \phi)$$

$$\text{backptr}[i][j][2] = 0$$

Iteration Step

for $t=3$ to T do

for $j=1$ to N do

รูปที่ 4.6 ขั้นตอนวิธีวิเทอร์บี (Viterbi Algorithm)

for $k=1$ to N do

$$\text{seqscore}[j][k][t] = \max_{i=1, N} (\text{seqscore}[j][i][t-1])$$

$$\times \text{Prob}(L_k | L_j, L_i) \times \text{Prob}(W_i | L_k)$$

$$\text{backptr}[j][k][t] = \text{ค่า } i \text{ ที่ทำให้ค่าสมการที่ผ่านมาเป็นค่าที่มากที่สุด}$$

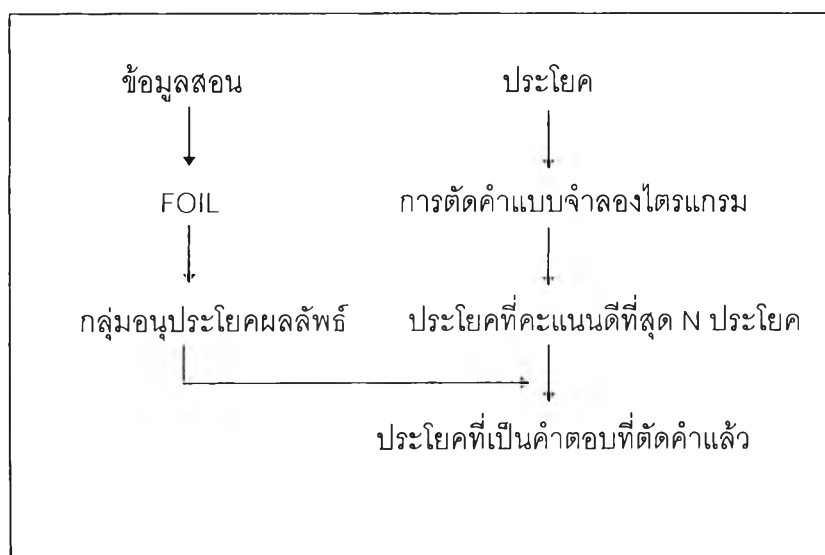
Sequence Identification Step

$C[T] = k$ and $C[T-1] = j$ โดยที่ j และ k นั้นทำให้ $\text{seqscore}[j][k][T]$ มีค่ามากที่สุด

for $i=T-2$ to 1 do

$$C[i] = \text{backptr}[C[i+1]][C[i+1]][i+1]$$

รูปที่ 4.6 ขั้นตอนวิธีวิเทอริ (Viterbi Algorithm)



รูปที่ 4.7 ขั้นตอนการตัดคำโดยใช้อนุประโยคผลลัพธ์จาก FOIL

	ข้อมูลสอน		ข้อมูลทดสอบ	
	แบบขึ้นกับบริบท	แบบไม่ขึ้นกับบริบท	แบบขึ้นกับบริบท	แบบไม่ขึ้นกับบริบท
FOIL	89.93 %	97.42 %	83.27 %	88.24 %
การตัดคำแบบจำลองไตรแกรม	87.19 %	92.27 %	80.54 %	82.35 %

ตารางที่ 4.3 ผลการทดลอง

คำสั่งเริ่มทำงานของโปรแกรมคือ

c:\trigram

Input file: ชื่อแฟ้มข้อมูลนำเข้า

Output file: ชื่อแฟ้มข้อมูลผลลัพธ์

ถ้าเป็นโปรแกรมที่มีโมดูลย่อยที่พัฒนาขึ้นมาใหม่จะใช้คำสั่ง foiltri แทน trigram

จากตารางที่ 4.3 จะเห็นได้ว่าการนำคุณลักษณะที่ได้จากการเรียนรู้ด้วยระบบ FOIL เข้ามาใช้ในการตัดคำกำกวมจากข้อความกำกวมทั้งแบบขึ้นอยู่กับบริบทและแบบไม่ขึ้นอยู่กับบริบทจะให้ความถูกต้องมากกว่าวิธีการตัดคำแบบจำลองไตรแกรม ซึ่งผลความถูกต้องในการตัดคำกำกวมจากข้อความกำกวมแบบขึ้นอยู่กับบริบทนั้นมีค่าไม่สูงมากนัก เหตุผลหนึ่งคือข้อมูลที่ใช้ในการเรียนรู้คุณลักษณะของคำกำกวมด้วยระบบ FOIL มีจำนวนไม่มากนักซึ่งอาจจะทำให้ไม่มีความหลากหลายของข้อมูล และบางคุณลักษณะอาจมีข้อมูลจำนวนน้อยเกินไปในการเรียนรู้จึงทำให้ผลที่ได้จากการเรียนรู้ไม่มีคุณลักษณะเช่นนี้ออกมา ซึ่งคุณลักษณะนั้นอาจเป็นคุณลักษณะที่สามารถใช้ช่วยการตัดคำกำกวมได้ นอกจากนี้ยังมีคำกำกวมคำอื่นๆที่ยังไม่ได้นำไปเรียนรู้เช่นคำว่า มากกว่า มากกว่า เหล่านี้ เหล่านี้ เพิ่มขึ้น เพิ่มขึ้น เป็นต้น

ดังนั้นในงานวิจัยนี้สรุปว่าการนำคุณลักษณะที่ได้จากการเรียนรู้ด้วยโปรแกรมตรรกะเชิงอุปนัยสามารถช่วยการตัดคำกำกวมได้ดีกว่าวิธีการตัดคำแบบจำลองไตรแกรม

ตัวอย่างแฟ้มข้อมูลและแฟ้มผลลัพธ์ของขั้นตอนการตัดคำ ที่มีข้อมูลในแฟ้มข้อมูลที่แสดงในรูปที่ 4.8 ซึ่งจะนำไปผ่านวิธีการตัดคำแบบจำลองไตรแกรมแล้วจะได้ผลลัพธ์ดังแสดงในรูปที่ 4.9 แล้วนำแฟ้มข้อมูลเดียวกันนี้ไปผ่านวิธีการตัดคำแบบจำลองไตรแกรมที่มีการนำอนุประโยคผลลัพธ์ที่ได้จากการเรียนรู้ด้วยระบบ FOIL มาช่วยในการตัดคำ ซึ่งจะได้ผลลัพธ์ดังแสดงในรูปที่ 4.10

ข้อมูลที่ได้เป็นข้อความที่เลือกเก็บจากหนังสือพิมพ์และวารสารต่างๆ จำนวนทั้งสิ้น 95 รายการ
 ขอบเขตของตัวแปรภายนอกหรือฟังก์ชันเริ่มต้นณจุดที่ประกาศ จนถึงสิ้นสุด
 สถาบันพัฒนาข้าราชการพลเรือนได้จัดทำหนังสือรวมบทความทางวิชาการประจำปีในโอกาสครบรอบปีที่ 9 ของ
 การก่อตั้งสถาบัน
 ปาฐกถานี้เป็นปาฐกถาที่แสดงในการประชุมของสมาคมทางวิชาการแห่งหนึ่งที่โรงแรมเอเชียพญาไทวันพุธ
 ที่ 18 ตุลาคม 2510
 องค์การวางแผนในระดับต่างๆนี้ก็ยังไม่ประสบผลดีเท่าที่ควร
 เขาควรจะพัฒนาตนเองและลูกน้องเพื่อที่จะทำงานของหน่วยงานให้มีผลดีให้มากที่สุด
 กรอบตรงกลางนั้น เป็นพื้นที่ซึ่งผู้ใช้สามารถวาดรูปเครือข่าย
 นายก. ได้จ่ายเงินมากให้รัฐแต่อาจได้ผลตอบแทนน้อยกว่านายข. ซึ่งให้เงินน้อยกว่าตน

รูปที่ 4.8 ข้อมูลในแฟ้มข้อมูล

- ✗ ข้อมูล/NCMN|ที่/|PREL|ได้/XVAE|เป็น/VSTA|ข้อความ/NCMN|ที่/|PREL|เลือก/VACT|เก็บ/VACT|จาก/RPRE|หนังสือ/NCMN|พิมพ์/VACT|และ/JCRG|วารสาร/NCMN|ต่างๆ/PUNC|BR/PUNC|จำนวน/NCMN|ทั้งสิ้น/DDAN|BR/PUNC|95/NCMN|BR/PUNC|รายการ/CNIT|
- ✗ ขอบเขต/NCMN|ของ/RPRE|ตัว/CNIT|แปร/VACT|ภายนอก/VATT|หรือ/EITT|ฟังก์ชัน/NCMN|เริ่มต้น/VSTA|ณ/NCMN|จุด/CNIT|ที่/|PREL|ประกาศ/VACT|BR/PUNC|จน/JSBR|สิ้นสุด/VSTA|
- ✗ สถาบัน/CNIT|พัฒนา/VACT|ข้าราชการพลเรือน/NCMN|ได้/XVAE|จัดทำ/VACT|หนังสือ/NCMN|รวม/VACT|บทความ/NCMN|ทาง/RPRE|วิชาการ/NCMN|ประจำปี/VATT|ปี/CMTR|ใน/RPRE|โอกาส/NCMN|ครบ/ADV|รอบ/CFQC|ปี/CMTR|ที่/|PREL|9/PUNC|ของ/RPRE|การ/FIXN|ก่อตั้ง/VACT|สถาบัน/CNIT|
- ✗ ปาฐกถา/VSTA|นี้/DDAC|เป็น/VSTA|ปาฐกถา/VSTA|ที่/|PREL|แสดง/VACT|ใน/RPRE|การ/FIXN|ประชุม/VACT|ของ/RPRE|สมาคม/NCMN|ทาง/RPRE|วิชาการ/NCMN|การ/FIXN|แห่ง/CNIT|หนึ่ง/DCNM|ที่/|PREL|โรง/CMTR|แรม/NCMN|เอเชียพญาไท/PUNC|วัน/CMTR|พุทธ/NCMN|ที่/|PREL|18/NCMN|ตุลาคม/NCMN|2510/NPRP|
- ✗ องค์/CMTR|กร/NCMN|วางแผน/VACT|ใน/RPRE|ระดับ/NCMN|ต่างๆ/DCNM|นี้/DDAC|ก็/JSBR|ยัง/XVBM|ไม่/NEG|ประสบ/VSTA|ผลดี/NCMN|เท่า/CMTR|ที่/|PREL|ควร/XVMM|
- ✗ เขา/PPRS|ควร/XVMM|จะ/XVBM|พัฒนา/VACT|ตนเอง/PDMN|และ/JCRG|ลูกน้อง/NCMN|เพื่อที่จะ/JSBR|ทำงาน/VACT|ของ/RPRE|หน่วยงาน/NCMN|ให้/JSBR|มี/VSTA|ผลดี/NCMN|ให้/JSBR|มาก/ADV|ที่สุด/JCMP|
- ✓ กรอบ/NCMN|ตรง/VATT|กลาง/VATT|นั้น/DDAC|BR/PUNC|เป็น/VSTA|พื้นที่/NCMN|ซึ่ง/JSBR|ผู้ใช้/NCMN|สามารถ/XVAM|วาด/VACT|รูป/NCMN|เครือข่าย/NCMN|
- ✓ นายก/NCMN|.NCMN|ได้/XVAE|จ่าย/VACT|เงิน/NCMN|มา/XVAE|ก/NCMN|ให้/JSBR|รัฐ/NCMN|แต่/JCRG|อาจ/XVMM|ได้/XVAE|ผลตอบแทน/NCMN|น้อย/VATT|กว่า/JCMP|นาย/NTTL|ช/DCNM|.PUNC|ซึ่ง/JSBR|ให้/JSBR|เงิน/NCMN|น้อย/VATT|กว่า/JCMP|ตน/PPRS|

รูปที่ 4.9 ผลลัพธ์ที่ได้จากเพิ่มข้อมูลที่ผ่านมาด้วยวิธีการตัดคำแบบจำลองไตรแกรม

- ✓ ข้อมูล/NCMN|ที่/|PREL|ได้/XVAE|เป็น/VSTA|ข้อความ/NCMN|ที่/|PREL|เลือก/VACT|เก็บ/VACT|จาก/RPRE|หนังสือพิมพ์/NCMN|และ/JCRG|วารสาร/NCMN|ต่างๆ/NPRP|BR/PUNC|จำนวน/NCMN|ทั้งสิ้น/DDAN|BR/PUNC|95/PUNC|BR/PUNC|รายการ/CNIT|
- ✓ ขอบเขต/NCMN|ของ/RPRE|ตัวแปร/NCMN|ภายนอก/VATT|หรือ/EITT|ฟังก์ชัน/NCMN|เริ่ม/VSTA|ต้น/CMTR|ณ/NCMN|จุด/CNIT|ที่/|PREL|ประกาศ/VACT|BR/PUNC|จน/JSBR|สิ้นสุด/VSTA|
- ✓ สถาบัน/CNIT|พัฒนา/VACT|ข้าราชการพลเรือน/NCMN|ได้/XVAE|จัดทำ/VACT|หนังสือ/NCMN|รวม/VACT|บทความ/NCMN|ทาง/RPRE|วิชาการ/NCMN|ประจำปี/NCMN|ใน/RPRE|โอกาส/NCMN|ครบ/ADV|รอบ/CFQC|ปี/CMTR|ที่/|PREL|9/NCMN|ของ/RPRE|การ/FIXN|ก่อตั้ง/VACT|สถาบัน/CNIT|
- ✓ ปาฐกถา/VSTA|นี้/DDAC|เป็น/VSTA|ปาฐกถา/VSTA|ที่/|PREL|แสดง/VACT|ใน/RPRE|การ/FIXN|ประชุม/VACT|

รูปที่ 4.10 ผลลัพธ์ที่ได้จากเพิ่มข้อมูลที่ผ่านมาด้วยวิธีการตัดคำที่ใช้อนุประโยคผลลัพธ์ที่ได้จากการ

เรียนรู้ด้วยระบบ FOIL

ของ/RPRE|สมาคม/NCMN|ทาง/RPRE|วิชาการ/NCMN|แห่ง/CNIT|หนึ่ง/DCNM|ที่/PREL|
 โรง/CMTR|แรม/NCMN|เอเชียพญาไท/NCMN|วัน/CMTR|พร/NCMN|ที่/PREL|18/NCMN|ตุลาคม/NCMN|
 2510/NPRP|
 ✓ องค์กร/NCMN|วางแผน/VACT|ใน/RPRE|ระดับ/NCMN|ต่างๆ/PUNC|นี้/DDAC|ก็/JSBR|ยัง/XVBM|ไม่/NEG|
 ประสบ/VSTA|ผล/NCMN|ดี/VATT|เท่า/CMTR|ที่/PREL|ควร/XVMM|
 ✓ เขา/PPRS|ควร/XVMM|จะ/XVBM|พัฒนา/VACT|ตนเอง/PDMN|และ/JCRG|ลูกน้อง/NCMN|เพื่อที่จะ/JSBR|
 ทำงาน/VACT|ของ/RPRE|หน่วยงาน/NCMN|ให้/JSBR|มี/VSTA|ผล/NCMN|ดี/VATT|ให้/JSBR|มาก/ADVN|
 ที่สุด/JCMP|
 ✕ กรอ/VACT|ป/PUNC|ตรง/VATT|กลาง/VATT|นั้น/DDAC|BR/PUNC|เป็น/VSTA|พื้นที่/NCMN|ซึ่ง/JSBR|
 ผู้ใช้/NCMN|สามารถ/XVAM|วาด/VACT|รูป/NCMN|เครือข่าย/NCMN|
 ✕ นายก/NCMN|. /NCMN|ได้/XVAE|จ่าย/VACT|เงิน/NCMN|มาก/ADVN|ให้/JSBR|รัฐ/NCMN|แต่/JCRG|
 อาจ/XVMM|ได้/XVAE|ผล/NCMN|ตอบแทน/VACT|น้อย/VATT|กว่า/JCMP|นา/NCMN|ยช/DCNM|. /NPRP|
 ซึ่ง/JSBR|ให้/JSBR|เงิน/NCMN|น้อย/VATT|กว่า/JCMP|ตน/PPRS|

รูปที่ 4.10 ผลลัพธ์ที่ได้จากเพิ่มข้อมูลที่ผ่านด้วยวิธีการตัดคำที่ใช้อนุประโยคผลลัพธ์ที่ได้จากการ
 เรียนรู้ด้วยระบบ FOIL (ต่อ)