สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

รายงานการวิจัยฉบับสมบูรณ์

การระบุผู้พูดจากหนังสือนิทาน
Speaker Identification from Story Book

รศ.ดร. พรฤดี เนตโสภากุล

King Mongkut's Institute of Technology Ladkrabang

RCH
W 276 ก
2556

ฉชหมู่.....
เลขทะเบียน 138001
วัน.เดือน.ปี - 8 ก.ย. 2558

b 12681490
.b.....
.i.....

ได้รับทุนสนับสนุนงานวิจัยจากเงินรายได้คณะเทคโนโลยีสารสนเทศ
ประจำปีงบประมาณ 2556
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ชื่อโครงการ การระบุผู้พูดจากหนังสือนิทาน
แหล่งเงิน เงินรายได้คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ประจำปีงบประมาณ 2556 จำนวนเงินที่ได้รับการสนับสนุน 50,000 บาท
ระยะเวลาการทำวิจัย ตั้งแต่ 1 ตุลาคม พ.ศ. 2555 ถึง 1 พฤศจิกายน พ.ศ. 2556
ชื่อ-สกุลหัวหน้าโครงการ
รศ.ดร. พรฤดี เนติโสภากุล
คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

บทคัดย่อ

ในหนังสือนิยาย นิทานหรือบทละคร จะประกอบด้วยข้อความที่เป็นบทพูด และส่วนของการบรรยายที่บ่งบอกเกี่ยวกับตัวละครที่กำลังพูด แต่บางบทพูดอาจจะไม่มีส่วนของการบรรยายปรากฏชัดเจน ไม่มีตัวละครที่พูดปรากฏอย่างชัดเจน หรือมีตัวละครมากกว่าหนึ่งตัว สำหรับงานวิจัยที่เกี่ยวกับการระบุผู้พูดจากหนังสือนิยายที่ผ่านมา ส่วนใหญ่จะใช้สร้างรายชื่อตัวละครที่สร้างขึ้นโดยผู้ใช้งานเพื่อใช้ในการระบุ ดังนั้นเมื่อต้องการนำหนังสือเรื่องใหม่ จำเป็นต้องสร้างรายชื่อตัวละครใหม่ งานวิจัยนี้จึงนำเสนอการระบุผู้พูดโดยไม่จำเป็นต้องสร้างรายชื่อตัวละครทั้งหมดในเรื่องก่อน โดยการสร้างฐานข้อมูลคำกริยาที่บ่งบอกอาการพูด และจัดกลุ่มคำกริยาเหล่านั้นตามลำดับความสำคัญ จากนั้นใช้เรกูลา เอกเพรสชัน (Regular Expression) และกฎฮิวริสติก (Heuristic Rule) ร่วมกับฐานข้อมูลคำกริยาที่ได้สร้างขึ้น เพื่อการสกัดชื่อตัวละครที่มีบทพูดในเรื่องเพื่อใช้เป็นตัวเลือกในการระบุผู้พูด และการระบุผู้พูดโดยใช้ความสำคัญของคำกริยาที่ได้จัดกลุ่มไว้ ผลการระบุผู้พูดพบว่า ระบบสามารถระบุผู้พูดได้ถูกต้อง 85%

คำสำคัญ: การระบุผู้พูด, การสกัดตัวละคร, คำกริยาที่บ่งบอกอาการพูด, บทพูด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

Research Title: Speaker Identification from Story Book
Researcher: Assoc. Prof. Ponrudee Netisopakul, Ph.D
Faculty: Information Technology. **Department:** Information Technology.

ABSTRACT

The text of story books consists of quote texts and narrative texts. If there is a pair of a quote text and a narrative text, usually the narrative text will have the speaker of the quote text. However, for some quote texts, there is no associative narrative text or have many narrative texts or have many actors in the narrative text. Most of previous research on speaker identification employed an avatar list, which is a list of all actors in the story created manually. This approach is inappropriate for our work, which aims to mark the speaker of each quote for 40 children stories and also for new coming stories. Hence, this research proposes a speaker identification process without an avatar list. The main idea is to employ a set of speech verbs and their priorities as being a speaker. In addition, regular expression and heuristic rules are used to select the most probable speaker of each quote. The experimental result shows the correctness of our approach to be 85%

Keywords: speaker identification, actor extraction, speech verbs, quote text

King Mongkut's Institute of Technology Ladkrabang

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

กิตติกรรมประกาศ

งานวิจัยเรื่อง การระบุผู้พูดจากหนังสือนิทาน ได้รับทุนสนับสนุนการวิจัยจากจากแหล่งเงินรายได้คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ประจำปีงบประมาณ พ.ศ. 2556 ผู้วิจัยขอขอบคุณพระคุณคณะฯและสถาบันฯที่ให้การสนับสนุนทุนวิจัยมา ณ ที่นี้

รศ.ดร. พรฤดี เนติโสภากุล

King Mongkut's Institute of Technology Ladkrabang

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

สารบัญ

หน้า

บทคัดย่อ I
ABSTRACT II
กิตติกรรมประกาศ III
สารบัญ IV
สารบัญตาราง V
สารบัญภาพ VI
บทที่ 1 บทนำ 1
1.1 ความเป็นมาและความสำคัญของปัญหา 1
1.2 วัตถุประสงค์ของการวิจัย 1
1.3 ขอบเขตของการวิจัย 1
1.4 วิธีดำเนินการวิจัย 1
1.5 ประโยชน์ที่คาดว่าจะได้รับ 1
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 2
2.1 กฎแบบลำดับขั้นสำหรับการระบุผู้พูดในหนังสือนิยาย (Hierarchical Rule Generalization for Speaker Identification in Fiction Books) 2
2.2 การระบุผู้พูดในหนังสือนิยายโดยกระบวนการจัดลำดับความสำคัญ (A naïve, salience-base method for speaker identification in fiction books) 5
2.3 การระบุผู้พูดในหนังสือเด็กสำหรับการสังเคราะห์เสียง (Identifying Speakers in Children's Stories for Speech Synthesis) 8
2.4 การจำแนกข้อความประเภทนิยายตามตัวละครแบบกึ่งอัตโนมัติ (Semi-automatic Novel Text Classification based on Character) 9
บทที่ 3 วิธีดำเนินการวิจัย 12
3.1 การสำรวจและการสร้างคลังคำกริยาที่บ่งบอกอาการพูด 12
3.2 การเตรียมข้อมูลเพื่อเข้าสู่กระบวนการระบุผู้พูด 13
3.3 การระบุผู้พูดให้กับบทพูดโดยอัตโนมัติ 16
บทที่ 4 ผลการวิจัย 22
4.1 การวัดประสิทธิภาพการสกัดคำที่เป็นตัวละครในวลีที่อยู่ติดกับบทพูดด้วยกฎ และการระบุผู้พูดด้วยตัวละครที่สกัดได้ 23
4.2 การวัดประสิทธิภาพการสกัดคำที่เป็นตัวละครด้วยกฎ และการกำกับชื่อตัวละครจากรายชื่อที่สร้างขึ้นในวลีที่อยู่ติดกับบทพูด และการระบุผู้พูดด้วยตัวละครที่สกัดและกำกับได้ 26
4.3 การวัดประสิทธิภาพการสกัดคำที่เป็นตัวละครด้วยกฎ และการกำกับชื่อตัวละครจากรายชื่อที่สร้างขึ้นสำหรับทุกวลี และการระบุผู้พูดด้วยตัวละครสกัดและกำกับได้ทั้งหมด 29
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ 32
5.1 สรุปผลการวิจัย 32
5.2 ข้อเสนอแนะ 32

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บรรณานุกรม 33
ภาคผนวก ก 36
ภาคผนวก ข 41
ประวัตินักวิจัย 46

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

สารบัญตาราง

ตารางที่ **หน้า**

3.1 แสดงจำนวนคำกริยาในแต่ละคลาส 13
3.2 แสดงความถี่ของนามวลีที่เป็นชื่อตัวละครและอยู่ติดกับคำกริยาที่บ่งบอกอาการพูด 16
3.3 แสดงความถี่ความสัมพันธ์ของตำแหน่งวลีที่แสดงผู้พูดกับความใกล้ชิดของวลีนั้นกับบทพูด 20
4.1 การสกัดคำที่เป็นตัวละครในวลีที่อยู่ติดกับบทพูด และการระบุผู้พูดจากตัวละครนั้น 24
4.2 การสกัดคำที่เป็นตัวละครร่วมกับการกำกับชื่อตัวละครในวลีที่อยู่ติดกับบทพูด และการระบุผู้พูดจากตัวละครนั้น 27
4.3 การสกัดคำที่เป็นตัวละครร่วมกับการกำกับชื่อตัวละครในเรื่องทั้งหมด และการระบุผู้พูดจากตัวละครนั้น 30
ก.1 คำกริยาที่แสดงอาการพูด และตำแหน่งวลีที่พบคำกริยานั้น และตำแหน่งของชื่อตัวละครที่กระทำกริยานั้นกับคำกริยา 37
ข.1 โครงสร้างนามวลีที่เป็นชื่อตัวละครจากการสำรวจนิทานภาษาไทยสำหรับเด็ก 40 เรื่อง 42

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

สารบัญภาพ

ภาพที่ **หน้า**

2.1 โครงสร้างข้อมูลแบบลำดับขั้น....2
2.2 แสดงตัวอย่างกฎสำหรับหาคำกริยาที่บ่งบอกอาการพูด....3
2.3 แสดงตัวอย่างกฎสำหรับหาชื่อตัวละคร....4
2.4 แสดงตัวอย่างกฎสำหรับระบุบทพูดกับผู้พูด....4
2.5 แสดงการรวมกันของกฎและการแจกแจงกฎ....5
2.6 แสดงกริยาที่แสดงอาการพูด....6
2.7 แสดงภาพสมมุติของกริยาที่เชื่อมโยงกัน....6
2.8 แสดงกระบวนการตัดสินใจในการแก้ปัญหาสรรพนามที่อ้างถึง....7
2.9 แสดงการกำกับบทสนทนาภายในเรื่อง....8
2.10 แสดงแบบโครงสร้างที่เป็นไปได้สำหรับสองบทพูด....10
2.11 แสดงกลุ่มโครงสร้างที่ถูกกำกับหน้าที่ของคำเรียบร้อยแล้ว....10
2.12 แสดงการกำหนดตัวละคร (<CHAR>) ให้เป็นผู้พูด (<UCHAR>)....10
3.1 แสดงแผนผังการทำงานทั้งหมดของการระบุผู้พูด....12
3.2 แสดงผลลัพธ์ที่ได้จากโปรแกรมตัดคำด้วยเทคนิคไฮบริด....14
3.3 แสดงผลลัพธ์กำกับหน้าที่ของคำและกำกับคำกริยาที่บ่งบอกอาการพูด....15
3.4 แสดงตัวอย่างของรวมคำนามวลี....16
3.5 แสดงกฎที่ใช้ในการสกัดชื่อตัวละคร....17
3.6 แสดงขั้นตอนการทำงานของการสกัดและกำกับชื่อตัวละคร....18
3.7 แสดงตัวอย่างการกำกับชื่อตัวละครที่อาจจะเป็นผู้พูด....18
3.8 แสดงตัวอัลกอริทึมในการกำหนดการกระทำให้กับตัวละคร....19
3.9 แสดงตัวอัลกอริทึมในการระบุผู้พูดให้กับบทพูด....20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

การระบุผู้พูดจากหนังสือนิทานเป็นขั้นตอนหนึ่งที่สำคัญ ในงานวิจัยด้านการทำความเข้าใจภาษามนุษย์ เพื่อนำไปสู่งานวิจัยเกี่ยวกับการแปลงข้อความให้เป็นเสียงพูดของตัวละครในหนังสือนิทาน ซึ่งได้รับความสนใจจากนักวิจัยเป็นจำนวนมาก เนื่องจากสามารถนำไปพัฒนาเป็นหนังสือเสียงที่ยังมีไม่เพียงพอ

ปัญหาหนึ่งของการแปลงข้อความให้เป็นเสียงพูดของตัวละครในหนังสือนิทาน คือ ผู้ฟังไม่สามารถแยกแยะน้ำเสียงตัวละครได้ ทำให้ขาดอรรถรสในการฟัง เนื่องจากองค์ประกอบของข้อความจากหนังสือนิทานประกอบด้วย ตัวละครที่ปรากฏในรูปของชื่อคน สัตว์ หรือสิ่งอื่นๆที่สร้างขึ้นในเรื่อง ข้อความที่เป็นบทพูด และข้อความส่วนการบรรยาย ดังนั้น การระบุตัวละคร การแยกแยะระหว่างข้อความที่เป็นบทพูดกับข้อความส่วนบรรยาย และการระบุบทพูดให้กับตัวละคร จึงมีความจำเป็นอย่างมาก

1.2 วัตถุประสงค์ของการวิจัย

1.2.1 เพื่อศึกษาเทคนิคและอัลกอริทึมที่ใช้ในการระบุผู้พูดจากหนังสือนิทาน

1.2.2 เพื่อวัดประสิทธิภาพของเทคนิคและอัลกอริทึมที่ใช้ในการระบุผู้พูดจากหนังสือนิทาน

1.3 ขอบเขตของการวิจัย

1.3.1 ศึกษาและเปรียบเทียบเทคนิค 2 เทคนิคที่ใช้ในการระบุผู้พูด

1.3.2 กำหนดคลาส (Classes) ของคำกริยาที่แสดงอาการพูด (Speech Verb) จากหนังสือนิทานภาษาไทยสำหรับเด็ก 40 เรื่องโดยผู้ใช้งาน

1.3.3 นิทานที่นำมาประมวลผลอยู่ในรูปแบบของไฟล์ข้อความ (.txt) ภาษาไทย มีการกำกับข้อความที่เป็นบทพูดด้วยเครื่องหมายอัญประกาศ

1.3.4 ทดลองและวัดประสิทธิภาพจากการระบุผู้พูดจากหนังสือนิทาน 40 เรื่อง

1.4 วิธีดำเนินการวิจัย

1.4.1 ศึกษาและรวบรวมข้อมูลที่เกี่ยวข้องกับการระบุผู้พูดจากหนังสือนิทาน หนังสือนิยาย และบทละคร

1.4.2 คิดค้นและปรับปรุงเทคนิคและอัลกอริทึมให้เหมาะสมโดเมน

1.4.3 พัฒนาเทคนิคและอัลกอริทึมที่ได้คิดค้น

1.4.4 วัดประสิทธิภาพและเปรียบเทียบเทคนิคและอัลกอริทึมที่ใช้ในการระบุผู้พูดจากหนังสือนิทาน

1.4.5 จัดทำรายงานผลการดำเนินการ

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1.5.1 ได้รับเทคนิคและอัลกอริทึมที่ใช้ในการระบุผู้พูดจากหนังสือนิทาน

1.5.2 ได้รับข้อความที่มีการติดแท็กกำกับ บทพูดที่ตรงกับผู้พูด เพื่อนำไปเป็นอินพุท (Input) ในการสังเคราะห์เสียง

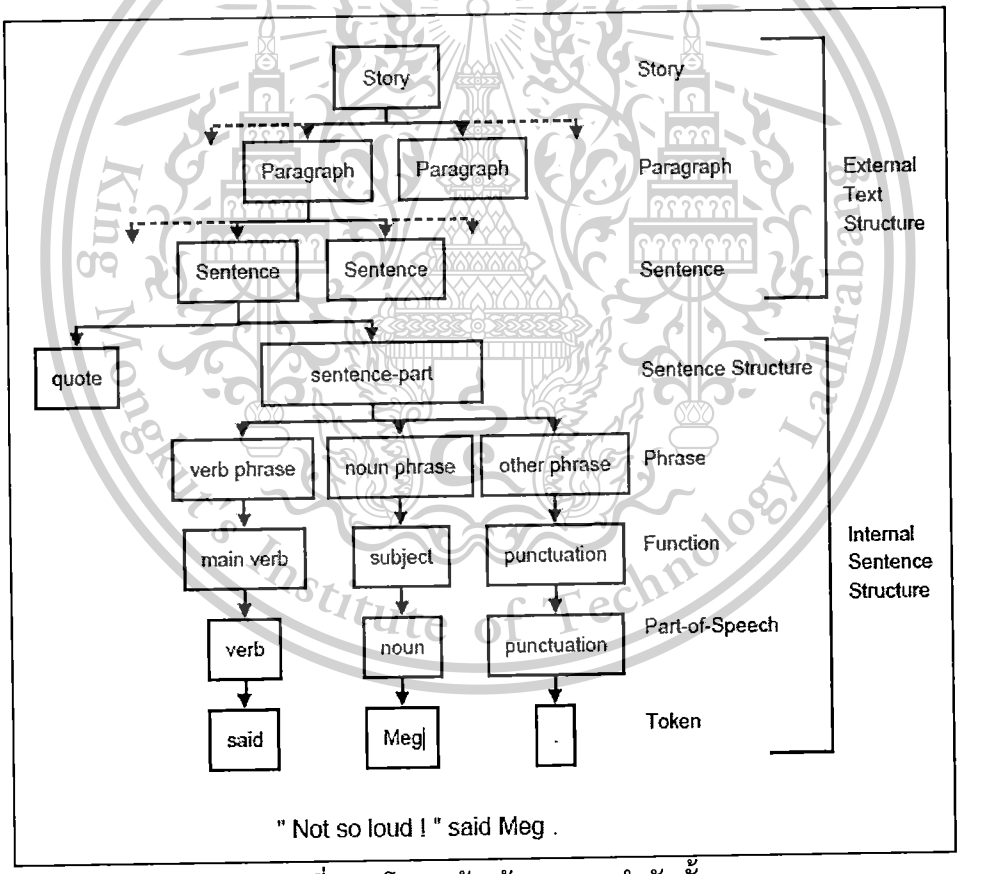
1.5.3 เป็นขั้นตอนพื้นฐานเพื่อนำไปสู่ในงานวิจัยขั้นต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 กฎแบบลำดับขั้นสำหรับการระบุผู้พูดในหนังสือนิยาย (Hierarchical Rule Generalization for Speaker Identification in Fiction Books)

งานวิจัยนี้ (Glass และ Bangay, 2006) นำเสนอการระบุผู้พูดจากหนังสือนิยาย โดยใช้กฎและรูปแบบของโครงสร้างประโยคแบบลำดับขั้น โดยการจัดโครงสร้างข้อความภายในเรื่อง ดังภาพที่ 2.1 คือ โครงสร้างภายนอก (External Text Structure) เป็นการจัดโครงสร้างข้อมูลโดยรวมทั้งหมด คือ เนื้อเรื่อง (Story) จะถูกแบ่งเป็นพารากราฟ (Paragraphs) แต่ละพารากราฟจะประกอบด้วยประโยค (Sentences) และ โครงสร้างระดับล่างหรือโครงสร้างภายใน (Internal Sentence Structure) ซึ่งเป็นส่วนที่นำไปพิจารณา โดยมีการใช้เครื่องมือสำหรับกำกับหน้าที่ของคำ (Part-of-Speech Tagger) และเครื่องมือสำหรับการวิเคราะห์ประโยค (Syntactic Parser Tool) จากภายนอก เพื่อช่วยในการจัดโครงสร้างของประโยคที่เป็นส่วนของการบรรยาย (Sentence Part)



ภาพที่ 2.1 โครงสร้างข้อมูลแบบลำดับขั้น
ที่มา : (Glass และ Bangay, 2006)

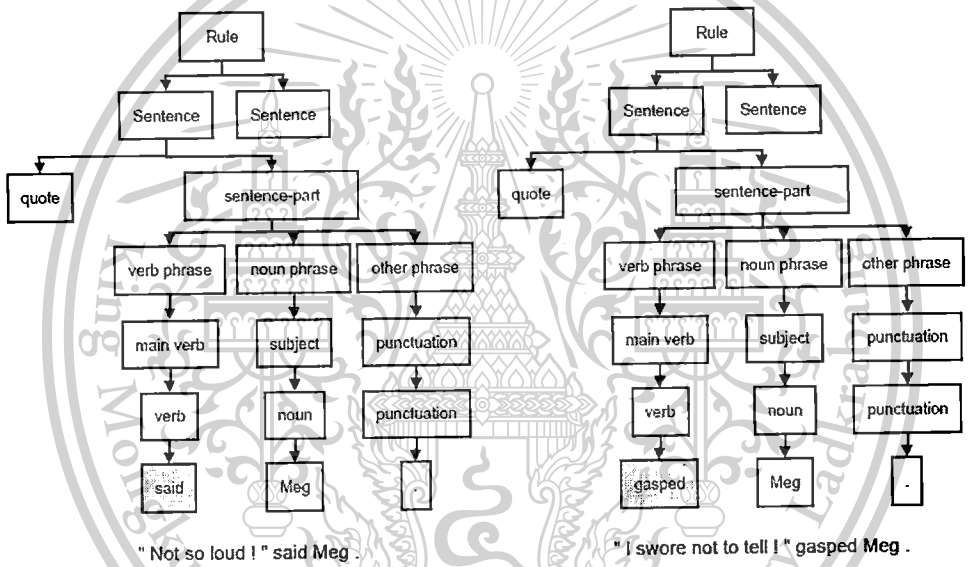
งานวิจัยนี้นำเสนอการใช้ คำกริยาที่บ่งบอกอาการพูด (Speech Verb) ซึ่งคำกริยาเหล่านั้นอาจจะอยู่ในประโยคก่อนหน้าบทพูด หรือตามหลังบทพูดก็ได้ แล้วยังใช้รายชื่อตัวละครที่สร้างขึ้นโดยมนุษย์ ช่วยในการระบุผู้พูดให้กับบทพูด มีรายละเอียดดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2.1.1 การสร้างรายชื่อตัวละคร (Character Identification) ในงานวิจัยนี้จำเป็นต้องมีการสร้างรายชื่อตัวละคร (Avatar List) เพื่อใช้ในการระบุผู้พูดด้วยมนุษย์

2.1.2 สร้างกฎสำหรับการสกัดข้อมูล (Build Rule for Script Extraction) กฎที่ใช้ในการสกัดคำกริยาที่บ่งบอกอาการพูด ตัวละคร และผู้พูด นั้นถูกสร้างด้วยผู้วิจัยเองจากหนังสือที่จะใช้เรียนรู้ (Seeding Text)

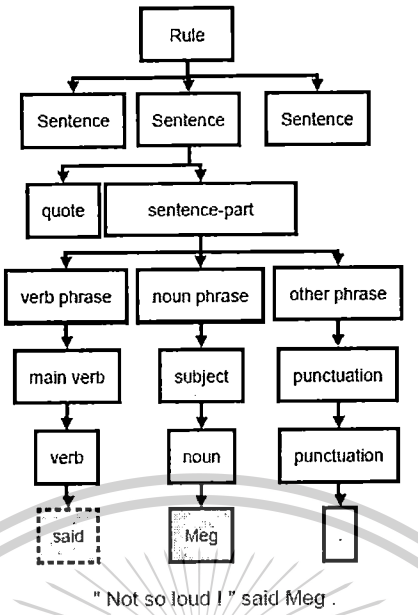
กฎสำหรับหาคำกริยาที่บ่งบอกอาการพูด (Speech verb) การสร้างกฎสำหรับการหาคำกริยาที่บ่งบอกอาการพูด เริ่มจากการเลือกกฎขึ้นมาหนึ่งกฎ (Seed Example) โดยสิ่งที่สนใจคือ คำกริยาที่แสดงอาการพูด (Speech Verb) จากนั้นเลือกกฎที่สองมาพิจารณา ซึ่งมีลักษณะคล้ายกับกฎแรกแตกต่างกันเพียงคำที่อยู่ในตำแหน่งคำกริยาที่บ่งบอกอาการพูด ถ้าทุกตำแหน่งตรงกันก็สามารถหาคำกริยาที่บ่งบอกอาการพูดได้จากตำแหน่งที่แตกต่างกันนั้นเอง แสดงดังภาพที่ 2.2 กฎทางซ้ายมือกฎแรก และกฎทางขวามือคือกฎที่สอง จะได้คำตอบของกฎที่สองคือ “gasped” เป็นกริยาที่แสดงอาการพูด



ภาพที่ 2.2 แสดงตัวอย่างกฎสำหรับหาคำกริยาที่บ่งบอกอาการพูด
ที่มา : (Glass และ Bangay, 2006)

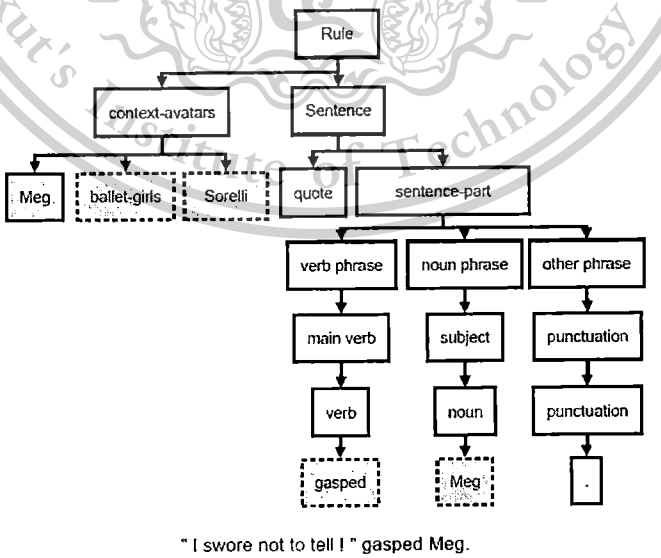
กฎสำหรับหาตัวละคร (Actor token) ในงานวิจัยนี้จะมีการแก้ปัญหาเกี่ยวกับการใช้สรรพนามแทนชื่อตัวละคร (Resolving Ambiguity) เรียบร้อยแล้ว ส่วนกฎที่ใช้คล้ายกับกฎที่ใช้กับการหาคำกริยาที่บ่งบอกอาการพูด แสดงดังภาพที่ 2.3 โดยจะใช้คำกริยาที่ได้จากในส่วนแรกหาชื่อคำที่แสดงถึงตัวละคร (Actor token) ดังนั้นหากผลลัพธ์ที่ได้จากกฎในส่วนแรกผิดพลาด อาจจะส่งผลทำให้ผลลัพธ์จากกฎนี้ผิดพลาดด้วย แต่งานวิจัยนี้มีการใช้รายชื่อตัวละครที่สร้างโดยผู้งาน ทำให้ลดความผิดพลาดในส่วนนี้ลงได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



ภาพที่ 2.3 แสดงตัวอย่างกฎสำหรับหาชื่อตัวละคร
ที่มา : (Glass และ Bangay, 2006)

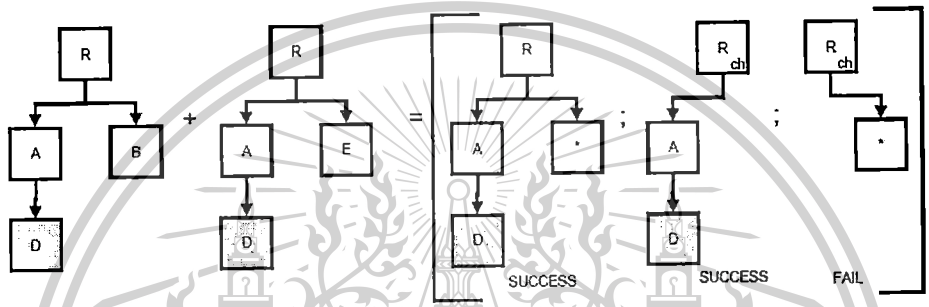
กฎสำหรับระบุผู้พูด (Speaker) ให้กับบทพูด (Quote) ในส่วนนี้มีการเพิ่มรายชื่อตัวละครที่ปรากฏในบริบท (Context-avatars) ซึ่งรายชื่อเหล่าต้องมีการเรียงลำดับชื่อตัวละครในรายชื่อ โดยเรียงจากตัวละครที่ถูกอ้างอิงล่าสุด แสดงดังภาพที่ 2.4 วิธีเลือกผู้พูดนั้นจะใช้ตัวละครที่ถูกสกัดได้ และคำกริยาที่บ่งบอกอาการพูด หากชื่อตัวละครที่สกัดได้ตรงกับชื่อตรงกับชื่อตัวละครในรายชื่อของตัวละครที่ปรากฏในบริบท ตัวละครตัวนั้นจะเป็นผู้พูดบทพูดนั้นๆ ถ้าหากไม่มีคำที่แสดงถึงตัวละครในประโยคที่พิจารณาอยู่ จะนำประโยคก่อนหน้ามาทำการพิจารณาจนพบชื่อตัวละคร ที่ตรงกับชื่อตัวละครในรายชื่อที่เพิ่มเข้ามา



ภาพที่ 2.4 แสดงตัวอย่างกฎสำหรับระบุบทพูดกับผู้พูด
ที่มา : (Glass และ Bangay, 2006)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2.1.3 รวมกฎ (Merging) เนื่องจากการดำเนินการในข้อ 2.1.2 ต้องสร้างกฎสำหรับทุกๆบทพูด ทำให้มีกฎมากมายในการระบุคำกริยา ชื่อตัวละคร และผู้พูด งานวิจัยนี้จึงนำเสนอการรวมกฎ เพื่อลดจำนวนกฎให้น้อยลง แต่ยังสามารถใช้กฎที่มีจำนวนน้อยระบุคำกริยา ชื่อตัวละคร และผู้พูดได้เหมือนเดิม โดยจะใช้วิธีการเพิ่มไวการ์ด (Wild-cards) เป็นการนำกฎมาเปรียบเทียบกัน โดยในทุกๆ ตำแหน่ง ถ้าตำแหน่งที่ไม่เหมือนกันสามารถใส่ไวการ์ด (Wild-cards) แทน โดยในตำแหน่งที่เป็นพรีเซิร์ฟ (Preserved Node) และตำแหน่งที่เป็นคำตอบ (Answer Node) หรือตำแหน่งที่เหนือทั้งสองตำแหน่งนั้น จะไม่ถูกแทนด้วยโหนดที่เป็นไวการ์ด (Wild-cards Node) และทั้งสองกฎต้องมีตำแหนงที่เป็นพรีเซิร์ฟ (preserved node) ที่เหมือนกัน แสดงดังภาพที่ 2.5 ในโหนดที่เป็นสีเทา (D) จะได้ผลลัพธ์ดังกฎที่สามจากทางซ้ายมือ ที่โหนด B และ E ถูกแทนด้วยไวการ์ด



ภาพที่ 2.5 แสดงการรวมกันของกฎและการแจกแจงกฎ
ที่มา : (Glass และ Bangay, 2006)

2.1.4 การแจกแจงกฎ (Enumerating all sequences of matching children) เมื่อเปรียบเทียบกฎเพื่อใส่ไวการ์ดแล้ว จะทำการแจกแจงจากส่วนต่างๆโครงสร้างเพื่อสร้างเป็นโครงสร้างย่อยที่ใช้แทนกฎทั้งสอง แสดงดังภาพที่ 2.5 จากโหนด B และ E ถูกแทนด้วยไวการ์ด ทำการแยกโครงสร้างของกฎได้เป็นกฎย่อยสองกฎ ซึ่งกฎทางขวามือนั้นไม่สามารถใช้ได้เพราะกฎจะสกัดผิดพลาดเนื่องจากไวการ์ดสามารถจะแทนคำใดๆก็ได้

จากงานวิจัยนี้พบว่างานวิจัยนี้มีนำเสนอในการสกัดสารสนเทศในรูปแบบโครงสร้างที่เป็นลำดับชั้นในการวิเคราะห์โดยใช้แพทเทิลแมตชิ่ง (Pattern Matching) โดยการใช้กฎที่มีจำนวนน้อยแต่ครอบคลุมสำหรับทุกแพทเทิล อีกทั้งยังการแก้ไขปัญหาการใช้สรรพนามแทนชื่อตัวละคร และสามารถแก้ปัญหาการไม่ปรากฏของคำกริยาที่บ่งบอกอาการพูดและผู้พูดที่ไม่ชัดเจน แต่งานวิจัยนี้จำเป็นต้องสร้างรายชื่อของตัวละครโดยมนุษย์ก่อน อีกทั้งงานวิจัยนี้จำเป็นต้องมีชุดฝึกสอน ทำให้ผลลัพธ์ที่ได้ผลดีกับหนังสือที่มีผู้แต่งคนเดียวกับหนังสือที่ใช้เป็นชุดฝึกสอน แต่เมื่อใช้หนังสือที่มีผู้แต่งแตกต่างจากชุดฝึกสอนได้ผลลัพธ์ที่ไม่ค่อยดี ดังนั้นผลของการทดลองต้องขึ้นกับหนังสือที่ใช้เป็นชุดฝึกสอนด้วย

2.2 การระบุผู้พูดในหนังสือนิยายโดยกระบวนการจัดลำดับความสำคัญ (A naïve, salience-base method for speaker identification in fiction books)

งานวิจัยนี้ (Glass และ Bangay, 2007) นำเสนอการสกัดชื่อที่เป็นตัวละคร (Actor), การสกัดคำกริยาที่แสดงอาการพูด (Speech-Verb) ด้วยการให้คะแนนตามคุณสมบัติต่างๆ (Scoring Technique) และการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ระบุผู้พูดที่เกี่ยวข้องกับบทพูดนั้น โดยการใช้ฮิวริสติก (Heuristic) ซึ่งฮิวริสติกที่ใช้นั้น สามารถแก้ปัญหาการใช้สรรพนามแทนชื่อตัวละครได้ด้วย รายละเอียดและวิธีการ มีขั้นตอนดังนี้

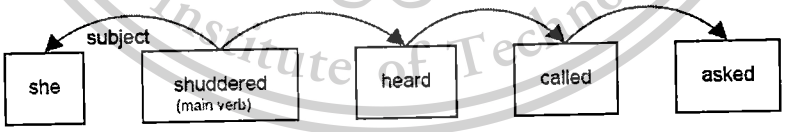
2.2.1 การกำกับคำกริยาที่แสดงอาการพูด (Speech-Verb Annotation) ทำการค้นหาคำกริยา (Verb) ในประโยคที่อยู่รอบบทสนทนา ผลลัพธ์ที่ได้จะเป็นรายการคำกริยา ที่อาจเป็นคำกริยาที่แสดงอาการพูดของบทสนทนา ที่กำลังพิจารณาอยู่ โดยการพิจารณาคุณสมบัติของคำกริยาตามคุณสมบัติ คือ คำกริยาหลัก (Main Verb) : คำที่ถูกกำกับว่าเป็นคำกริยาหลักจาก Connexor FDG parser, ไฮเปอร์เนียม (Hypernym) : คำกริยาที่เป็นซับเซ็ต (Subset) ของคำ “คอมมูนิเคชัน (communication)” “เวอร์บาไลซ์ (verbalise)” หรือ บรีธ (breath) ตามคลังข้อมูลเวิดเน็ต (WordNet), คำกริยาในประโยคที่อยู่ติดกับบทสนทนา (Adjacent Sentence) และ คำกริยาที่อยู่ใกล้ชิดกับบทสนทนา (Proximity to quote) : คำกริยาที่เกิดขึ้นก่อนหรือตามหลังบทสนทนา

ทุกคุณสมบัติมีการให้คะแนนคุณสมบัติละ1 คะแนน แต่คำกริยาที่อยู่ใกล้ชิดกับบทสนทนา จะคิดตามสัดส่วนของระยะทางระหว่างคำกริยากับบทสนทนา มีค่าอยู่ระหว่าง 0 ถึง 1 คำกริยาคำใดมีคะแนนสูงสุดจะถูกเลือกให้เป็นคำกริยาที่แสดงอาการพูดของบทสนทนาที่กำลังพิจารณาอยู่นั้น

2.2.2 การกำกับตัวละคร (Actor Annotation) การระบุตัวละครมีความคิดพื้นฐานที่ว่า คำที่แสดงถึงตัวละครมีการพึ่งพากับคำกริยาที่แสดงอาการพูด ซึ่งสามารถระบุได้จากโปรแกรมที่ใช้แจกแจงประโยค (Parser) ในบางประโยคคำกริยาที่เชื่อมต่อกัน (Verb - Chains) ดังแสดงในภาพที่ 2.6 คำกริยาจำนวนมากสามารถเชื่อมไปหาคำกริยาหลักคำเดียว แล้วจะสามารถหาประธาน หรือกรรมของกริยานั้นได้

She **shuddered** when she **heard** little Jammes speak of the ghost , **called** her a " silly little fool " and then , as she was the first to believe in ghosts in general , and the Opera ghost in particular , at once **asked** for details :
" Have you seen him ? "

ภาพที่ 2.6 แสดงกริยาที่แสดงอาการพูด
ที่มา : (Glass และ Bangay, 2007)



ภาพที่ 2.7 แสดงภาพสมมุติของกริยาที่เชื่อมโยงกัน
ที่มา : (Glass และ Bangay, 2007)

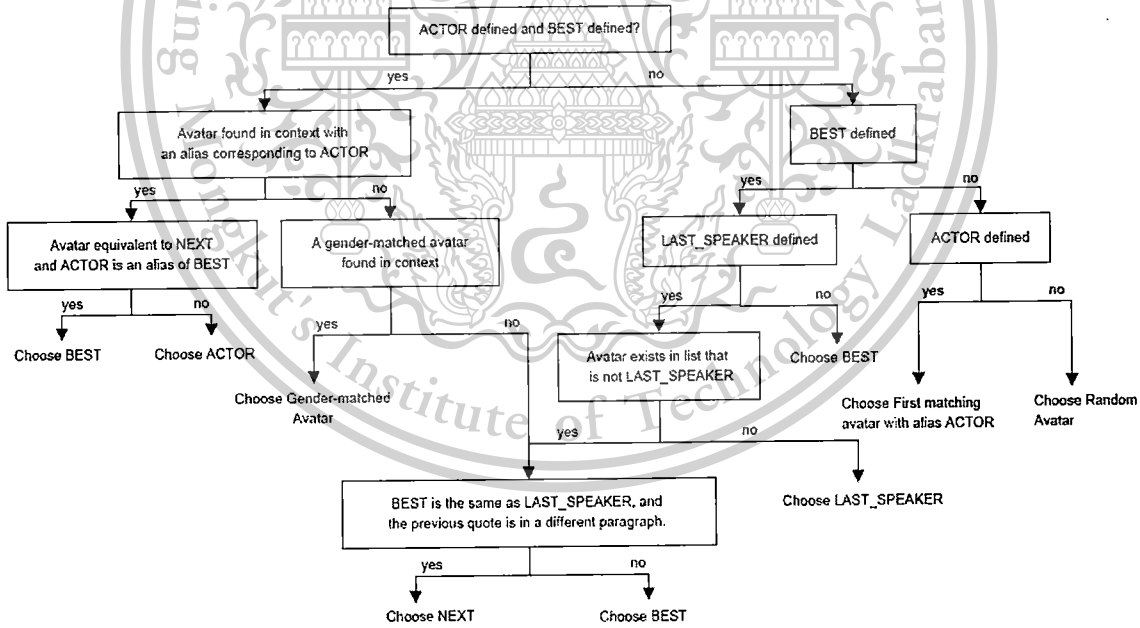
ความผิดพลาดในการแจกแจงประโยคในประโยคที่มีความซับซ้อน ขึ้นอยู่กับโครงสร้างที่ได้จากโปรแกรมแจกแจงประโยค ส่งผลให้เทคนิคการจัดลำดับที่เคยใช้ในการค้นหาตัวละครขึ้นอยู่กับคำกริยา ซึ่งเป็นเพียงหนึ่งคุณสมบัติที่มีการให้คะแนน ดังนั้นนอกจากการใช้วิธีการแจกแจงประโยคแล้ว ในการกำกับชื่อตัวละครยังใช้คุณสมบัติอื่นๆ ร่วมด้วย คือ คำนั้นเป็นประธาน (Subject) หรือ กรรม (Object): หน้าที่ของคำเป็นประธานหรือกรรมของกริยาหลักจากการแจกแจงประโยค, กรองจากหน้าที่ของคำ (Part-of-Speech Filter): คำนามที่เป็นลูกหลานของหมวดบุคคล (Person) ในคลังข้อมูลเวิดเน็ต (Word Net), กรองจากชื่อเฉพาะ (Proper Noun Filter): คำที่ขึ้นต้นด้วยตัวอักษรตัวพิมพ์ใหญ่เว้นแต่คำแรกที่ขึ้นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเอกสารนี้ทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ประโยค, ชื่อตัวละครที่เป็นตัวย่อ (Abbreviation Filter) และ ระยะห่างจากกริยา (Distance from Verb) ซึ่งดูจากความใกล้ชิดกับคำกริยาที่แสดงอาการพูด

ทุกคุณสมบัติมีการให้คะแนนคุณสมบัติละ 1 คะแนน ยกเว้นระยะห่างจากกริยาที่มีการนำ 0.1% มาคูณกับคะแนนที่มีอยู่ สำหรับแต่ละคำที่อยู่ระหว่างคำที่แสดงถึงตัวละคร และคำกริยาที่แสดงอาการพูด

2.2.3 การเลือกผู้พูด (Speaker Resolution) ตัวละครที่ถูกระบุในขั้นตอนการกำกับชื่อตัวละคร ตรงกับชื่อตัวละครในรายชื่อตัวละคร (Avatar List) ที่ถูกสร้างโดยผู้ใช้งาน ตัวละครนั้นจะถูกกำหนดให้เป็นผู้พูด สำหรับบทพูดที่กำลังพิจารณาอยู่ อย่างไรก็ตามชื่อตัวละครที่ได้อาจอยู่ในรูปแบบที่แตกต่างกัน คือ ชื่อที่ใช้อ้างอิงได้โดยตรง (Direct Reference) สรรพนามที่ใช้อ้างถึง (Pronominal Anaphora) คำนามที่ใช้อ้างถึง (Nominal Anaphora) และ คำที่ใช้อ้างถึงที่ไม่ชี้เฉพาะ (Deictic Anaphora) การเลือกตัวละครจากรายชื่อที่สร้างในกรณีแรกนั้นง่าย แต่อีกสามกรณีต้องการข้อมูลอย่างอื่นร่วมด้วย คือ อ้างจากบริบทของบทพูด (Quote Context) เนื่องจากบริบทที่อยู่รอบๆ บทสนทนาสามารถอธิบายเหตุการณ์ต่างๆ จึงการนำมาช่วยในการเพิ่มหรือลดคะแนนคือ พิจารณาชื่อตัวละครที่ปรากฏในบริบทนั้น ที่ละตัวจนครบทุกตัว ถ้าชื่อตัวละครที่ถูกกำกับตรงกับชื่อตัวละคร หรือนามแฝงของตัวละครในรายชื่อที่ถูกสร้างขึ้นด้วยมนุษย์ให้คะแนนชื่อตัวละครนั้น 1 คะแนน ส่วนคำที่เหลือก็จะคูณด้วย 0.9 และคำที่มี “ ‘s ” ตามหลังหรือคำนั้นอยู่หลังคำบุพบท (Preposition-Complement) คะแนนที่ได้ก็จะลบด้วย 0.3 แล้วเรียงลำดับชื่อตัวละครในบริบทนั้นจากนั้นใช้กฎการตัดสินใจดังภาพที่ 2.8 ในการระบุผู้พูด



ภาพที่ 2.8 แสดงกระบวนการตัดสินใจในการแก้ปัญหาสรรพนามที่อ้างถึง
ที่มา : (Glass และ Bangay, 2007)

ถ้าคำที่อ้างถึงตัวละครเป็นคำสรรพนาม หรือคำนามที่ใช้แทนชื่อตัวละคร มีกระบวนที่ตัดสินใจตามตัวละครในบริบทของบทสนทนาที่ควรถูกเลือก ซึ่งแตกต่างจากอัลกอริทึมที่ใช้แก้ปัญหาแบบเดิม ซึ่งในที่นี้เกี่ยวข้องกับการตัดสินใจเลือกผู้พูดจากการระบุตัวละคร ไม่ใช่การแก้ปัญหาการใช้คำนามและสรรพนามเรียกชื่อแทนทั้งหมด มี 4 ตัวแปรคือ Actor: อ้างถึงตัวละครที่ถูกระบุไว้ก่อนหน้านี้ อาจไม่มีถ้าบทสนทนา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารนี้ทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ไม่มีกริยาแสดงอาการพูด และคำที่ใช้อ้างถึงตัวละคร, BEST: อ้างถึงตัวละครอันดับแรกรายชื่อที่เกี่ยวข้องกับบริบทของบทสนทนา อาจไม่มีถ้าไม่มีบริบท, NEXT: อ้างถึงตัวละครอันดับสองรายชื่อที่เกี่ยวข้องกับบริบทของบทสนทนา ถูกใช้เมื่อบทสนทนามีการสลับกันพูด และ LAST_SPEAKER: อ้างถึงผู้พูดที่ถูกก่อนหน้านี้ อาจไม่มีค่าเมื่อบทสนทนาปัจจุบันเป็นบทสนทนาแรก

งานวิจัยนี้นำเสนอการระบุผู้พูดจากหนังสือนิยาย ซึ่งเป็นงานวิจัยที่ต่อเนื่องจากงานวิจัยก่อนหน้านี้ โดยนำเสนองานวิจัยที่ไม่ต้องอาศัยชุดฝึกสอน ด้วยวิธีการให้คะแนนความสำคัญตามคุณสมบัติต่างๆ นอกจากนี้ยังใช้กฎการตัดใจสินสำหรับการแก้ปัญหาการใช้สรรพนามแทนชื่อตัวละคร แม้จะไม่ได้แก้ปัญหาทั้งหมดก็ตาม แต่งานวิจัยนี้ยังคงมีการสร้างรายชื่อของตัวละครโดยผู้วิจัยไว้ก่อน และ จากงานวิจัยนี้พบว่าปัญหาจากเครื่องมือที่ใช้ในวิเคราะห์ประโยคมีผลต่อความถูกต้องในการระบุผู้พูดด้วย

2.3 การระบุผู้พูดในหนังสือเด็กสำหรับการสังเคราะห์เสียง (Identifying Speakers in Children's Stories for Speech Synthesis)

งานวิจัยนี้ (Zhang, Black, และ Sproat, 2003) นำเสนอเกี่ยวกับการระบุผู้พูดในหนังสือสำหรับเด็ก พวกเขาเรียกระบบนี้ว่า อีเอสพีอีอาร์ (ESPER) ซึ่งมีแนวคิดและขั้นตอนการวิเคราะห์ดังนี้

2.3.1 การระบุบทพูด (Identifying Spoken Speech) โดยงานวิจัยนี้ มีการกำกับบทพูดภายในเนื้อเรื่องโดยพิจารณาจากเครื่องหมายอัญประกาศ (Single or Double quote) ซึ่งงานวิจัยนี้มีการพิจารณาบทพูด (Quote) ที่ซ้อนกันอยู่ภายในด้วย แสดงดังภาพที่ 2.9

```
The farmer described it to his wife:
<QUOTE TYPE="NEW"> "The tail-feathers
of the fowl were very short, and it
winked with both its eyes, and said
<QUOTE TYPE="NEW"> "Cluck, cluck."</QUOTE>
What were the thoughts of the fowl as it
said this I cannot tell you..." </QUOTE>
```

ภาพที่ 2.9 แสดงการกำกับบทสนทนาภายในเรื่อง
ที่มา : (Zhang, Black, และ Sproat, 2003)

นอกจากนั้นแล้วงานวิจัยนี้ยังมีการระบุชนิดของบทพูด (Identifying Quoted Speech Type) ซึ่งจะแบ่งออกเป็น 2 ชนิดคือ บทพูดที่ต่อเนื่องกับบทพูดก่อนหน้านี้ (Continuation Quote) หรือว่าเป็นบทพูดที่ไม่ต่อเนื่อง (New Quote) เพื่อลดความผิดพลาดของการระบุผู้พูด โดยมีสมมุติฐานที่ว่า หากบทพูดที่กำลังพิจารณา เป็นบทพูดที่ต่อเนื่องกับบทพูดก่อนหน้านี้ ก็จะใช้ตัวละครตัวเดียวกันกับบทพูดก่อนหน้านี้เป็นผู้พูดของบทพูดปัจจุบัน ซึ่งงานวิจัยนี้นำเสนอ ดิสซิชันทรี (Decision Tree) ในการแบ่งประเภทของบทพูด โดยมีคุณสมบัติที่ใช้สำหรับการพิจารณาคือ คำขึ้นต้นของบทพูด, คำลงท้ายของบทพูด, ตัวอักษรในคำแรกของบทพูดขึ้นเป็นตัวอักษรพิมพ์ใหญ่, บทพูดเป็นวลีแรกในพารากราฟ, บทพูดเป็นบทพูดแรกในพารากราฟ, วรรคของบทพูดก่อนหน้าที่อยู่ในย่อหน้าเดียวกัน, จำนวนวรรคของคำที่อยู่ก่อนบทพูดและจำนวนที่คั่นระหว่างบทพูดก่อนหน้ากับบทพูดปัจจุบัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

2.3.2 การระบุตัวละคร (Character Identification) งานวิจัยนี้นำเสนอเทคนิคที่เกี่ยวกับการสกัดชื่อเฉพาะ (Named Entities Extraction) โดยการใช้เทคนิคแพทเทิลแมตชิ่ง (Pattern Matching) สำหรับการสกัดชื่อตัวละครที่เป็นชื่อเฉพาะ และมีการใช้เครื่องมือกำกับหน้าที่ของคำที่ชื่อว่า Festival Speech Synthesis System เพื่อช่วยสกัดชื่อตัวละครที่ไม่ได้มีชื่อเป็นชื่อเฉพาะ เช่น the old washerwoman, the Dodo

2.3.3 การระบุผู้พูด (Identifying the Speakers) งานวิจัยนี้สร้างกฎอย่างง่ายในการระบุผู้พูดให้กับบทพูดจากสมมุติฐานที่ว่า ผู้พูดและบทพูดอยู่ในย่อหน้าเดียวกัน และชื่อตัวละครที่ปรากฏในบทพูดไม่ใช่ผู้พูดบทพูดนั้น ทำให้ได้เป็นกฎอย่างง่ายว่า ชื่อตัวละครที่นำหน้าบทพูดในย่อหน้าเดียว ตัวละครตัวนั้นเป็นผู้พูด หรือถ้าไม่มีให้ใช้ตัวละครที่ตามหลังบทพูดนั้น

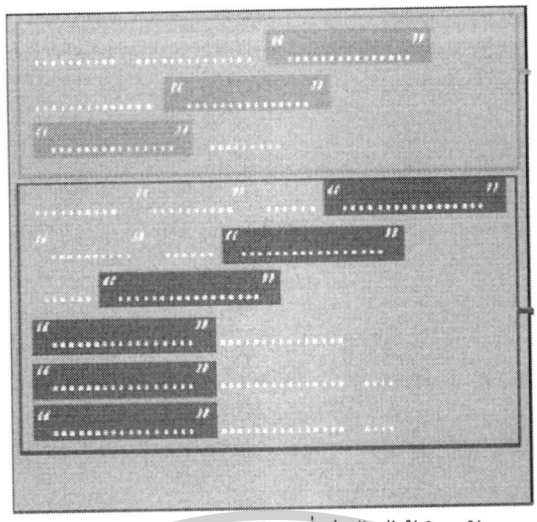
จากการศึกษาวิจัยนี้พบว่างานวิจัยนี้สามารถระบุชื่อตัวละครได้อัตโนมัติ และไม่ต้องมีการสร้างรายชื่อตัวละครขึ้นมาก่อน และมีแนวความคิดที่น่าสนใจคือการระบุประเภทของบทพูด แต่ปัญหาของงานวิจัยนี้พบว่า สมมุติฐานที่ใช้ในการระบุผู้พูดนั้นไม่มีความยืดหยุ่นและยังไม่เพียงพอ ซึ่งโดยทั่วแล้วผู้พูดอาจจะไม่จำเป็นต้องอยู่พารากราฟเดียวกันกับบทพูดก็ได้ นอกจากนี้การระบุผู้พูดไม่ได้พิจารณาส่วนอื่นร่วมด้วยเช่น การใช้คำกริยาที่บ่งบอกอาการพูด เป็นต้น และการแบ่งประเภทของบทพูดที่ผิดพลาด อาจทำให้ส่งผลต่อการระบุผู้พูดที่ผิดพลาดต่อเนื่องได้ด้วย

2.4 การจำแนกข้อความประเภทนิยายตามตัวละครแบบกึ่งอัตโนมัติ (Semi-automatic Novel Text Classification based on Character)

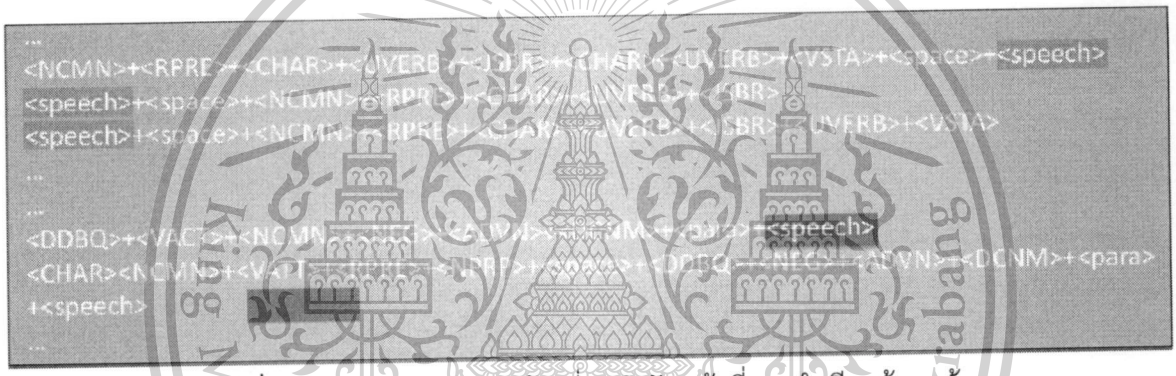
งานวิจัยด้านการระบุผู้พูดจากข้อความภาษาไทย (ณัฐธิดา, สุภรณี, และ อรณี, 2011) สำรวจพบเพียงงานนี้งานเดียว โดยงานนี้คำนวณความน่าจะเป็นของโครงสร้างประโยคที่ผ่านการกำกับ POS ด้วยวิธีไตร-แกรม (Trigram) โดยมีการเพิ่ม tag <CHAR> และ tag <UCHAR> ให้กับข้อความต้นฉบับ หมายถึง ตัวละคร และผู้พูด ตามลำดับ งานนี้มีข้อจำกัดหลายประการ คือ ประโยคที่ใช้ต้องเป็นประโยคอย่างง่ายเท่านั้น ผลลัพธ์และความแม่นยำขึ้นอยู่กับขนาดและลักษณะของชุดคลังประโยค โดยชุดคลังประโยคฝึกสอน ต้องผ่านการกำกับจากมนุษย์ก่อน

วิธีการของงานวิจัยนี้ เริ่มจากการแบ่งกลุ่มของโครงสร้างประโยค ที่จะใช้ในการคำนวณหาความน่าจะเป็น โดยกลุ่มของโครงสร้างประโยคจะประกอบด้วย วลีที่อยู่ติดกับบทพูดและบทพูด ดังภาพที่ 2.10 แสดงกลุ่มโครงสร้างสามแบบ สำหรับบทพูดที่หนึ่ง คือ แบบที่ 1 มีบทบรรยายสองประโยคตามด้วยบทพูด แบบที่สอง มีบทบรรยายหนึ่งประโยคก่อนบทพูด และแบบที่สาม คือมีบทบรรยายหลังบทพูด และ 6 แบบสำหรับบทพูดที่สอง คือมีทั้งบทบรรยายและบทพูดก่อนบทพูดที่กำลังสนใจ และรูปแบบอื่นๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.



ภาพที่ 2.10 แสดงแบบโครงสร้างที่เป็นไปได้สำหรับสองบทพูด
ที่มา : (ณัฐธิดา, สุภรณี, และ อรณี, 2011)



ภาพที่ 2.11 แสดงกลุ่มโครงสร้างที่ถูกกำกับหน้าที่ของคำเรียบร้อยแล้ว
ที่มา : (ณัฐธิดา, สุภรณี, และ อรณี, 2011)

เมื่อได้สร้างกลุ่มแบบโครงสร้างที่เหมาะสมแล้ว จึงกำกับหน้าที่ของคำ (Part-of-Speech) ในประโยคที่เป็นส่วนการบรรยาย เมื่อกำกับหน้าที่ของคำเรียบร้อยแล้ว จึงกำกับตัวละครด้วย <CHAR> จาก avatar list ที่มีอยู่ นอกจากนี้ ในชุดฝึกสอน จะต้องกำกับ <UCAHR> แสดงผู้พูดให้ถูกต้องด้วย เพื่อใช้คำนวณความน่าจะเป็น ของโครงสร้างประโยคที่มีความเป็นไปได้สูงสุดสำหรับตำแหน่ง <UCAHR> ในประโยคที่นำเข้ามา

<NCMN>+<CHAR>+<UVERB>+<JSBR>+<UCHAR>+<UVERB>+<VSTA>+<space>+<speech>

<NCMN>+<UCHAR>+<UVERB>+<JSBR>+<CHAR>+<UVERB>+<VSTA>+<space>+<speech>

ภาพที่ 2.12 แสดงการกำหนดตัวละคร (<CHAR>) ให้เป็นผู้พูด (<UCHAR>)
ที่มา : (ณัฐธิดา, สุภรณี, และ อรณี, 2011)

งานวิจัยนี้เป็นงานวิจัยเดียวที่พบสำหรับการระบุผู้พูดให้ตรงกับบทพูดในภาษาไทย แต่งานวิจัยนี้ยังมีข้อจำกัด คือ จำเป็นต้องสร้างรายชื่อตัวละครและรหัสตัวละครด้วยมนุษย์ สามารถประมวลผลได้เฉพาะประโยคอย่างง่าย และเนื่องจากต้องมีคลังข้อมูลขนาดเล็กสำหรับคำนวณค่าทางสถิติ ทำให้ผลลัพธ์ที่

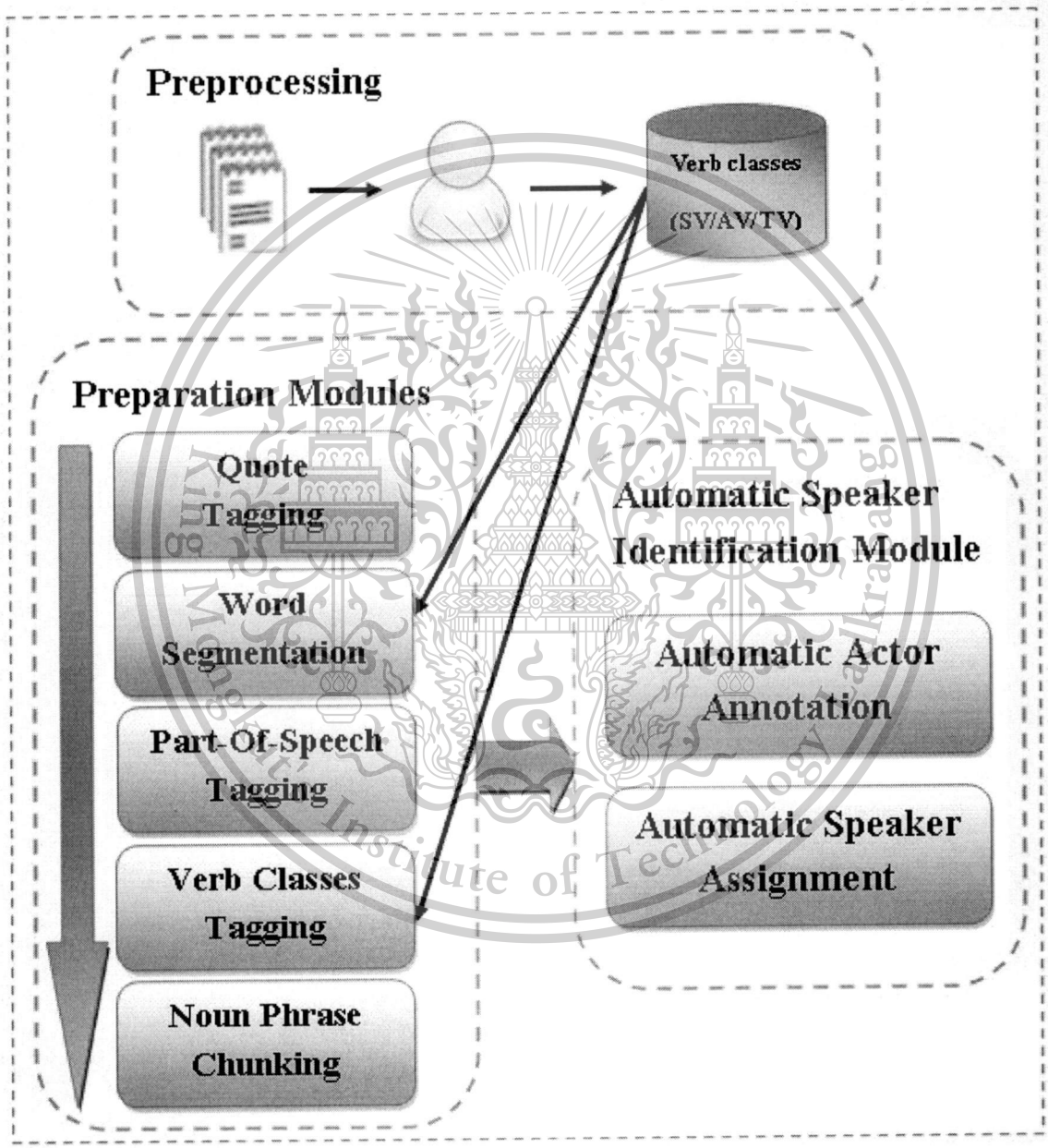
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ได้นั้นขึ้นอยู่กับโครงสร้างของประโยคที่จัดเก็บในคลังข้อมูลด้วย นอกจากนี้งานวิจัยนี้ยังใช้เพียงเทคนิคพื้นฐานในการคำนวณ คือ ไตร-แกรม (Trigram) โดยไม่มีการปรับปรุงหรือใช้ร่วมกับเทคนิคอื่นๆ ผลลัพธ์ที่ได้จึงไม่ดีเท่าที่ควร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 3
วิธีการดำเนินงานวิจัย

ในบทนี้ จะกล่าวถึงการเก็บรวบรวมคำกริยา ที่บ่งบอกอาการพูด (Speech Verbs) ขั้นตอนในการระบุตัวละคร ที่คาดว่าจะเป็นผู้พูด รวมทั้งกระบวนการและอัลกอริทึม ที่ใช้ในการระบุผู้พูดให้กับบทพูด ซึ่งแผนผังการทำงานของระบบ แสดงดังภาพที่ 3.1



ภาพที่ 3.1 แสดงแผนผังการทำงานทั้งหมดของการระบุผู้พูด

3.1 การสำรวจและการสร้างคลังคำกริยาที่บ่งบอกอาการพูด (Preprocessing)

จากการศึกษาและเก็บรวบรวมข้อมูล จากหนังสือนิทานภาษาไทยสำหรับเด็ก พบว่า แต่ละบทพูดมักมีบทบรรยายที่แสดงผู้พูดและอาการที่พูด โดยบทบรรยายดังกล่าว อาจอยู่ก่อนหน้าหรือหลังบทพูดก็ได้ เช่น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

“กล้าดียังไงมาปลุกข้า” ราชสีห์คำรามด้วยความโกรธ จากตัวอย่าง จะได้คำกริยาที่บ่งบอกอาการพูด คือ “คำราม” และ ตัวละครที่พูดคือ “ราชสีห์” ดังนั้น ผู้วิจัยจะทำการรวบรวมคำกริยาที่ปรากฏในลักษณะนี้เพื่อใช้เป็นคำบ่งชี้ผู้พูด ในการระบุตัวละครที่อาจจะเป็นผู้พูด คำกิริยาดังกล่าว สามารถแบ่งออกได้เป็น 4 คลาส (Classes) คือ

- คำกริยาที่บ่งบอกอาการพูดอย่างชัดเจน (Speech Verb: SV) เช่น พูด, กล่าว, ถาม, ตอบ, ...
- คำกริยาที่แสดงถึงความคิด หรืออาการเพ้อฝัน (Thinking Verb: TV) เช่น คิด, รำพึง, ฝันเฟื่อง, ...
- คำกริยาที่ไม่แสดงอาการพูดอย่างชัดเจน แต่มีอาการที่แสดงเป็นนัย (Action Verb: AV) เช่น พยักหน้า, แปลกใจ, ประหลาดใจ, ...
- คำกริยาอื่นๆ นอกเหนือจาก 3 คลาส ข้างต้น (Other Verb: Other)

จากการรวบรวมวลี ที่เป็นบทบรรยาย ที่มีผู้พูดและมีคำกิริยาเหล่านี้ชัดเจน จำนวน 725 วลี จากหนังสือนิทานภาษาไทยสำหรับเด็กจำนวน 40 เรื่อง จากจำนวนบทพูด 825 บทพูด พบว่า วลีส่วนใหญ่จะอยู่ในพารากราฟเดียวกันกับบทพูด คิดเป็น 92.97% วลีส่วนใหญ่อยู่ติดกับบทพูด คิดเป็น 96.69% จากวลีเหล่านี้ สามารถรวบรวมคำกริยาที่บ่งบอกอาการพูดได้ 91 คำ แสดงดังตารางที่ ก.1 นอกจากนี้ ยังมีคำกริยาที่บ่งบอกอาการพูดที่ได้จากผู้เชี่ยวชาญมนุษย์ รวมคำกริยาที่บ่งบอกอาการพูดที่รวบรวมไว้ทั้งหมด 107 คำ ซึ่งจัดเก็บไว้ในรูปแบบของ XML file โดยแต่ละคำจะกำกับคลาสข้างต้นไว้เรียบร้อยแล้ว จำนวนคำกริยาในแต่ละคลาส แสดงดังตารางที่ 3.1

ตารางที่ 3.1 แสดงจำนวนคำกริยาในแต่ละคลาส

Classes	Count	Examples
SV	43	พูด (say), ว่า (said that), ...
TV	10	คิด (think), สงสัย (doubt), ...
AV	54	ยิ้ม (smile), พยักหน้า (nod), ...
Other	-	คอย (wait), วิ่ง (ran), ...

3.2 การเตรียมข้อมูลเพื่อเข้าสู่กระบวนการระบุผู้พูด (Preparation Modules)

กระบวนการเตรียมข้อมูลจะทำการกำกับวลีที่เป็นบทพูด และวลีที่เป็นส่วนบรรยาย แล้วนำวลีที่เหล่านั้นไปตัดคำ พร้อมกำกับหน้าที่ของคำ (Part-of-Speech) แล้วกำกับคลาสของคำกริยา (Verb Classes) ตามฐานข้อมูลที่ได้สร้างไว้ และการรวมคำเพื่อสร้างนามวลี (Noun Phrase) ที่อาจจะเป็นชื่อของตัวละครภายในเรื่อง ซึ่งนิทานที่นำมาประมวลผลจะอยู่ในรูปข้อความตัวอักษรที่เก็บไว้ที่ไฟล์ .txt เมื่อการประมวลในกระบวนนี้แล้ว ข้อความเหล่านั้นจะถูกกับและเก็บไว้ในรูปแบบของ XML file

3.2.1 การกำกับบทพูด (Quote Tagging)

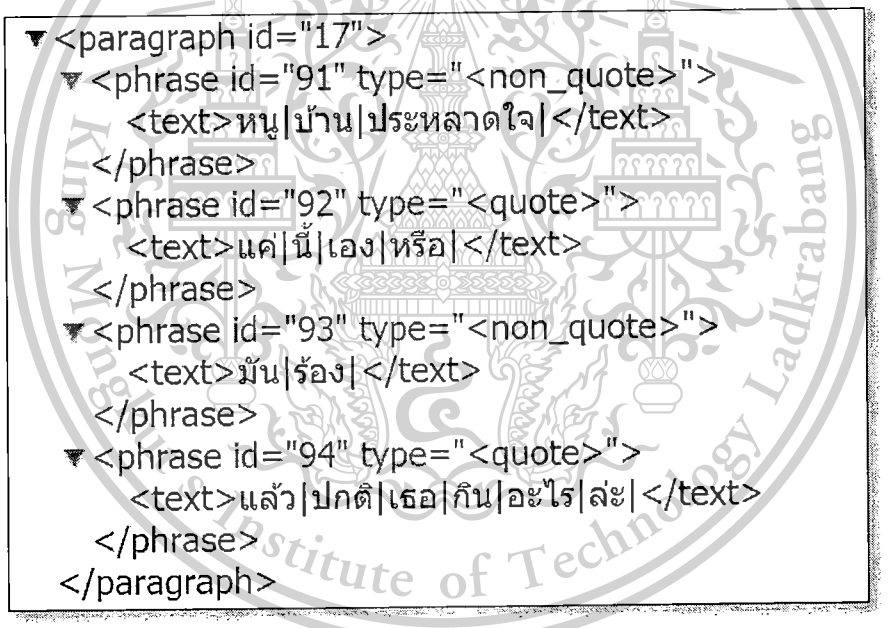
การกำกับวลีที่เป็นบทพูด (Quote) ในงานวิจัยนี้ จะทำการกำกับวลีที่อยู่ภายใต้เครื่องหมายอัญประกาศ (Double quote) ให้เป็นบทพูด เช่น *หนูบ้านประหลาดใจ “แค่นี้เองหรือ” มันร้อง “แล้วปกติเธอกินอะไรล่ะ”* ดังนั้นวลี “แค่นี้เองหรือ” และ “แล้วปกติเธอกินอะไรล่ะ” จะถูกกำกับให้เป็นบทพูด ดังแสดงในภาพที่ 3.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น ห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

3.2.2 การตัดคำ (Word Segmentation)

ในการตัดของงานวิจัยนี้ ใช้เครื่องมือจากภายนอกช่วยในการตัดคำ โดยใช้โปรแกรมตัดคำจากการแข่งขัน “การวัดเปรียบเทียบสมรรถนะเพื่อพัฒนามาตรฐานการประมวลผลภาษาไทย (Benchmark for Enhancing the Standard for Thai language Processing 2010: BEST 2010)” ที่มีชื่อว่า “การแบ่งคำภาษาไทยด้วยเทคนิคไฮบริด (Thai Word Segmentation a Hybrid Approach)” (วรศักดิ์, พีระศักดิ์, และ ธนพล, 2010) ซึ่งเป็นการผสมกันของ เทคนิคการแบ่งคำภาษาไทยด้วยดิกชันนารีจากคำที่ยาวที่สุดก่อน และเลือกคำจากหลักของภาษาไทย แล้วการตัดคำด้วยเทคนิคการแบ่งคำภาษาไทยด้วยดิกชันนารีจากคำที่ยาวที่สุด ซึ่งในเทคนิคนี้จะใช้ฐานข้อมูลของคำกริยาที่บ่งบอกอาการพูดที่ถูกเตรียมไว้เป็นดิกชันนารี เพื่อความถูกต้องของการกำกับคำกริยาที่บ่งบอกอาการพูดในขั้นตอนต่อไป

จากภาพที่ 3.2 เป็นตัวอย่างผลลัพธ์ที่ได้จากการตัดคำในกระบวนการนี้ เช่น *หนูบ้านประหลาดใจ* จะถูกตัดเป็น “หนู”, “บ้าน” และ “ประหลาดใจ” *แค่นี้เองหรือ* จะถูกตัดเป็น “แค่”, “นี้”, “เอง” และ “หรือ”



ภาพที่ 3.2 แสดงผลลัพธ์ที่ได้จากโปรแกรมตัดคำด้วยเทคนิคไฮบริด

3.2.3 การกำกับหน้าที่ของคำ (POS-of-Tagging)

การกำกับหน้าที่ของคำนั้นมีความสำคัญเป็นอย่างมากสำหรับการประมวลทางด้านภาษา เพราะการกำกับหน้าที่ของคำแต่ละคำนั้นจะทำให้สามารถวิเคราะห์ถึงโครงสร้างของประโยคได้ ดังนั้นในการวิจัยนี้จึงจำเป็นต้องมีการกำกับหน้าที่ให้คำแต่ละคำที่ถูกตัดมาแล้ว เพื่อช่วยในการวิเคราะห์หาตัวละครและผู้พูดในนิทาน ในงานวิจัยนี้ได้ใช้เครื่องมือภายนอกที่มีชื่อว่า “Jitar” ซึ่งใช้เทคนิคแบบจำลองมาร์โคฟแบบซ่อนเร้น (HMM) โดยใช้ค่าสถิติจาก ไตร-แกรม (Trigram) (Satayamas, 2012) และใช้คลังข้อมูลจาก NAiST Lab (NAiST Lab, 2011) ช่วยในการกำกับหน้าที่ของคำในวลีที่เป็นการบรรยาย (Narrative Phrase) ดังตัวอย่างที่แสดงในภาพที่ 3.3 จะพบว่า คำว่า “หนู” และ “บ้าน” จะถูกกำกับหน้าที่ของคำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

เป็น“คำนามทั่วไป (ncn)” ส่วน “ประหลาดใจ” จะถูกกำกับหน้าที่ของคำเป็น “คำกริยาที่ไม่ต้องการกรรม (vi)”

```
<paragraph id="17">
  <phrase id="91" type="<non_quote>">
    <text>หนู|บ้าน|ประหลาดใจ|</text>
    <word pos="ncn">หนู</word>
    <word pos="ncn">บ้าน</word>
    <word class="TV" pos="vi">ประหลาดใจ</word>
  </phrase>
  <phrase id="92" type="<quote>">
    <text>แค่|นี้|เอง|หรือ|</text>
  </phrase>
  <phrase id="93" type="<non_quote>">
    <text>มัน|ร้อง|</text>
    <word pos="pper">มัน</word>
    <word class="SV" pos="vi">ร้อง</word>
  </phrase>
  <phrase id="94" type="<quote>">
    <text>แล้ว|ปกติ|เธอ|กิน|อะไร|ล่ะ|</text>
  </phrase>
</paragraph>
```

ภาพที่ 3.3 แสดงผลลัพธ์กำกับหน้าที่ของคำและกำกับคำกริยาที่บ่งบอกอาการพูด

3.2.4 การกำกับคลาสของคำกริยา (Verb Classes Tagging)

ในการกำกับคลาสของคำกริยา (Verb Classes) จะทำการกำกับเฉพาะคำกริยาที่อยู่ในวลีบรรยาย (Narrative Phrase) ที่อยู่ติดกับบทพูด (Adjacent Phrase) เท่านั้น เนื่องจากวลีที่เป็นการบรรยายที่อยู่ติดกับบทพูดส่วนใหญ่ จะบรรยายถึงผู้พูดของบทพูดแต่ละบท การกำกับจะเปรียบเทียบคำกริยาที่อยู่วลีนั้น กับรายการคำกริยาที่ได้เตรียมไว้แล้ว แล้วกำกับคลาสให้กับคำกริยานั้น ดังแสดงในภาพที่ 3.3 เช่นคำว่า “ประหลาดใจ” มีการกำกับหน้าที่ของคำเป็นคำกริยาที่ไม่ต้องการกรรม และมีการกำกับคลาสของคำกริยาเป็น “คำกริยาที่แสดงถึงความคิด หรืออาการเพ้อฝัน (TV)” ส่วนคำว่า “ร้อง” มีการกำกับหน้าที่ของคำเป็นคำกริยาที่ไม่ต้องการกรรม และมีการกำกับคลาสของคำกริยาเป็น “คำกริยาที่บ่งบอกอาการพูดอย่างชัดเจน (SV)”

3.2.5 การรวมคำนาม (Noun Phrase Chunking)

เนื่องจากนามวลี (Noun Phrase) ในภาษาไทยสามารถเกิดขึ้นได้หลายรูปแบบ และภาษาไทยไม่มีการแบ่งขอบเขตของคำที่แน่นอนเหมือนกับภาษาอังกฤษ ที่ใช้ช่องว่างในการแบ่งขอบเขตของคำ ทำให้นามวลีถูกแยกออกจากกันในขั้นตอนการตัดคำ ดังแสดงในภาพที่ 3.3 คำว่า “หนู” และ “บ้าน” ถูกแบ่งแยกออกจากกัน ซึ่งทั้งสองคำควรจะรวมเป็นคำเดียวกันคือ “หนูบ้าน” ซึ่งเป็นชื่อของตัวละครที่ปรากฏในนิทานเรื่องหนึ่ง จากปัญหาดังกล่าว งานวิจัยนี้จึงต้องมีการรวมคำนามวลีที่ต้องการ โดยใช้ Regular Express Rule ในรวมคำนามที่ต้องการ

จากภาพที่ 3.4 แสดงตัวอย่างของรวมคำนามวลีที่ถูกแยกออกจากกันด้วยเครื่องหมาย “|” ในขั้นตอนการตัดคำ เช่น “หนูบ้าน”ที่ถูกแยกเป็น “หนู” กับ “บ้าน”, “ญาติของมัน” ถูกแยกเป็น “ญาติ”

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

“ของ” “มัน”, “ผู้นำข่าว” ถูกแยกเป็น “ผู้” “นำ” “ข่าว” และ “คุณนกไปรษณีย์” ถูกแยกเป็น “คุณ” “ไปรษณีย์” “นก”

```
▼ <phrase id="2" type="<non_quote>">
  ▼ <text>
      หนู|บ้าน|อุทาน|เมื่อ|ได้|รับ|บัตร|เชิญ|จาก|ญาติ|ของ|มัน|
    </text>
    <word pos="NP">หนูบ้าน</word>
    <word class="SV" pos="vi">อุทาน</word>
    <word pos="conj">เมื่อ</word>
    <word class="Other" pos="prev">ได้</word>
    <word class="Other" pos="vt">รับ</word>
    <word pos="NP">บัตร</word>
    <word class="Other" pos="vt">เชิญ</word>
    <word pos="prep">จาก</word>
    <word pos="NP">ญาติของมัน</word>
  </phrase>
▼ <phrase id="3" type="<non_quote>">
  ▼ <text>
      ซึ่ง|ผู้นำ|ข่าว|นี้|มา|แจ้ง|ก็|คือ|คุณ|ไปรษณีย์|นก|นั่นเอง|
    </text>
    <word pos="prel">ซึ่ง</word>
    <word pos="NP">ผู้นำข่าว</word>
    <word pos="det">นี้</word>
    <word class="Other" pos="vpost">มา</word>
    <word class="Other" pos="vt">แจ้ง</word>
    <word pos="conj">ก็</word>
    <word class="Other" pos="vcs">คือ</word>
    <word pos="NP">คุณไปรษณีย์นก</word>
    <word pos="part">นั่นเอง</word>
  </phrase>
```

ภาพที่ 3.4 แสดงตัวอย่างของรวมคำนามวลี

3.3 การระบุผู้พูดให้กับบทพูดโดยอัตโนมัติ (Automatic Speaker Identification Module)

ในการระบุผู้พูดให้กับบทพูดมี 2 กระบวนที่สำคัญ คือ การสกัดและกำกับตัวละครที่เกี่ยวข้องกับพูด และการเลือกตัวละครที่เป็นคนพูดบทพูดนั้น

3.3.1 การสกัดและกำกับตัวละครที่เกี่ยวข้อง (Automatic Actor Annotation)

การสกัดและกำกับตัวละครภายในเรื่องเป็นปัญหาหนึ่งที่สำคัญ สำหรับการระบุผู้พูด (Speaker Identification) ซึ่งงานวิจัยส่วนใหญ่จะใช้วิธีการของการสกัดชื่อเฉพาะ (Named Entity Extraction) แต่งานวิจัยนี้ไม่ได้ใช้วิธีของการสกัดชื่อเฉพาะ เนื่องจากชื่อตัวละครส่วนใหญ่เป็นคำนามทั่วๆไป เช่น “หนูบ้าน”, “หนูนา”, “ราชสีห์”, “หนู”, “ลา”, “คนตัดไม้”, “กบหนุ่ม” จากการศึกษาและสำรวจในการสร้างคำกริยาที่บ่งบอกอาการพูด พบว่าคำกริยาเหล่านั้นส่วนใหญ่จะตามหลังและอยู่ติดกับนามวลีที่เป็นชื่อตัวละครที่เป็นผู้พูด 89.5% ดังแสดงในตารางที่ 3.2

ตารางที่ 3.2 แสดงความถี่ของนามวลีที่เป็นชื่อตัวละครและอยู่ติดกับคำกริยาที่บ่งบอกอาการพูด

ตำแหน่งของนามวลี	ความถี่	เปอร์เซ็นต์ (%)
อยู่ติดกับคำกริยาที่บ่งบอกอาการพูด	649	89.5
อยู่ห่างกับคำกริยาที่บ่งบอกอาการพูดโดยมีคำอื่นคั่น	64	8.8
ไม่สามารถระบุได้	12	1.7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

ดังนั้นจึงเกิดแนวความคิดที่จะใช้ Regular Expression ในการสกัดชื่อตัวละคร (Actor Extraction) จากสมมุติฐานที่ว่า "นามวลีที่เป็นชื่อตัวละคร มักจะตามด้วยคำกริยาที่บ่งบอกอาการพูด" ซึ่งกฎที่ใช้ในการสกัดชื่อตัวละครแสดงดังภาพที่ 3.5 ซึ่งใช้กฎทั้งหมดจำนวน 13 กฎ ชื่อตัวละครที่สามารถสกัดได้จะถูกเก็บไว้ในไฟล์รายชื่อตัวละคร (List of Actors) ของแต่ละเรื่อง ดังภาพที่ 3.6 พร้อมทั้งมีการกำกับการกระทำ (Action) ให้กับตัวละครที่สามารถสกัดตามคลาสของคำกริยาที่ใช้ในการสกัดชื่อตัวละครตัวนั้น

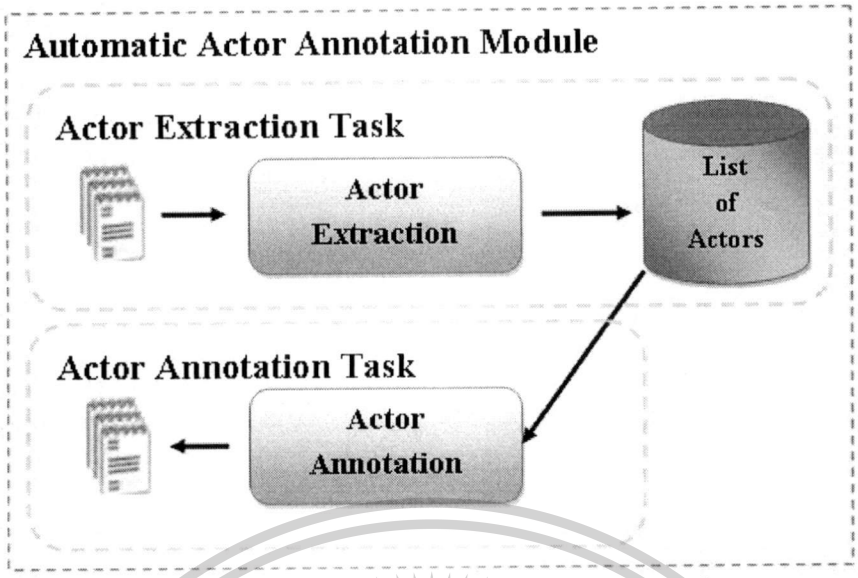
```
Rule1:   {(NP|pper) (SV|TV|AV)}
Rule2:   {(NP|pper) (vi|vt|prev|neg|adv|conj) (SV|TV|AV)}
Rule3:   {(NP|pper) (vt) (conj|adv) (SV|TV|AV)}
Rule4:   {(NP|pper) (prev) (prev) (SV|TV|AV)}
Rule5:   {(NP|pper) (prel) (vt) (SV|TV|AV)}
Rule6:   {(NP|pper) (conj) (prev|vi|vt|adv) (SV|TV|AV)}
Rule7:   {(NP|pper) (conj) (vt) (pref1) (SV|TV|AV)}
Rule8:   {(NP|pper) (conj) (prev) (prev) (SV|TV|AV)}
Rule9:   {(NP|pper) (prev) (vt) (conj) (SV|TV|AV)}
Rule10:  {(NP|pper) (vi) (vpost) (vi) (SV|TV|AV)}
Rule11:  {(NP|pper) (vi) (vi) (vpost) (SV|TV|AV)}
Rule12:  {(NP|pper) (vt) (vpost) (conj) (SV|TV|AV)}
Rule13:  {(NP|pper) (conj) (vt) (prec) (SV|TV|AV)}
```

ภาพที่ 3.5 แสดงกฎที่ใช้ในการสกัดชื่อตัวละคร

หลังจากการสกัดชื่อตัวละครในเบื้องต้นและสร้างรายชื่อตัวละครแล้ว งานวิจัยนี้จะนำรายชื่อตัวละครที่สร้างไว้ไปกำกับชื่อตัวละครในตำแหน่งอื่นๆ (Actor Annotation) ที่ไม่สามารถสกัดได้ โดยใช้วิธีการกำกับคำที่เหมือนกับชื่อตัวละครที่อยู่ในรายชื่อที่สร้างไว้จากวิธีการสกัดชื่อตัวละคร ซึ่งการสกัดและกำกับตัวละครด้วยวิธีนี้ ทำให้ชื่อตัวละครที่ได้ล้วนแต่เป็นตัวละครที่มีบทพูด เพราะชื่อตัวละครที่ไม่มีบทพูดนั้นจะไม่ถูกสกัดออกมา ทำให้การเลือกผู้พูดนั้นมีความถูกต้องมากขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

138001



ภาพที่ 3.6 แสดงขั้นตอนการทำงานของการสกัดและกำกับชื่อตัวละคร

จากภาพที่ 3.7 “หนูบ้าน” ถูกกำกับว่าเป็นตัวละครที่อาจเป็นผู้พูด (candidate = “yes”) มีการกระทำเป็น TV (action = “TV”) ได้จากคลาสของคำกริยา “ประหลาดใจ” ที่มีคลาสเป็น “TV” และคำว่า “มัน” ถูกกำกับว่าเป็นตัวละครที่อาจเป็นผู้พูด (candidate = “yes”) มีการกระทำเป็น SV (action = “SV”) ได้จากคลาสของคำกริยา “ร้อง” ที่มีคลาสเป็น “SV” ในงานวิจัยนี้ยังไม่มีการแก้ปัญหาการใช้สรรพนามแทนตัวบุคคล ดังนั้น คำว่า “มัน” จึงเสมือนเป็นชื่อตัวละครชื่อหนึ่ง

```
▼<paragraph id="17">
  ▼<phrase id="91" type="<non_quote>">
      <text>หนู|บ้าน|ประหลาดใจ|</text>
      <word action="TV" actorid="1" candidate="yes" pos="NP" verb="ประหลาดใจ">หนูบ้าน</word>
      <word class="TV" pos="vi">ประหลาดใจ</word>
    </phrase>
  ▼<phrase id="92" type="<quote>">
      <text>แค|นี่|เอง|หรือ|</text>
    </phrase>
  ▼<phrase id="93" type="<non_quote>">
      <text>มัน|ร้อง|</text>
      <word action="SV" actorid="1" candidate="yes" pos="pper" verb="ร้อง">มัน</word>
      <word class="SV" pos="vi">ร้อง</word>
    </phrase>
  ▼<phrase id="94" type="<quote>">
      <text>แล้ว|ปกติ|เธอ|กิน|อะไร|ละ|</text>
    </phrase>
</paragraph>
```

ภาพที่ 3.7 แสดงตัวอย่างการกำกับชื่อตัวละครที่อาจจะเป็นผู้พูด

3.3.2 การระบุผู้พูดให้บทพูดโดยอัตโนมัติ (Automatic Speaker Assignment)

ก่อนการระบุผู้พูดให้กับบทพูดจำเป็นต้องรู้การกระทำของตัวละคร ซึ่งกำกับตามชนิดของคำกริยาที่ได้แบ่งไว้ ดังนั้นยังมีอีกขั้นตอนหนึ่งที่สำคัญก่อนการระบุผู้พูดให้กับบทพูด คือ การกำหนดการกระทำ (Action) ให้กับตัวละคร (Actor) แต่ละตัวที่ไม่สามารถสกัดได้ในขั้นตอนการสกัดชื่อตัวละคร (Actor Extraction) แต่ถูกกำกับในขั้นตอน (Actor Annotation) ซึ่งในขั้นตอนนี้การกระทำของตัวละครจะถูกกำหนดให้เป็น “none” (action = “none”) ซึ่งการกำหนดการกระทำให้กับตัวละครสามารถทำได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

โดยการวิเคราะห์โครงสร้างของประโยค ซึ่งในภาษาไทยยังไม่มีกฎไวยากรณ์ (Grammar) ที่แน่นอนในการแจกแจงประโยค (Parsing) และภาษาไทยไม่มีการแบ่งขอบเขตที่แน่นอนเหมือนภาษาอังกฤษที่ใช้จุด (Full stop) ในการบอกจุดสิ้นสุดของประโยค ทำให้การเขียนกฎไวยากรณ์สำหรับการแจกแจงประโยคนั้นมีความยุ่งยากและซับซ้อนเป็นอย่างมาก

ดังนั้นในงานวิจัยนี้จึงใช้ กฎฮิวริสติกส์อย่างง่าย (Simple Heuristic Rule) ในการกำหนดการกระทำให้กับตัวละครที่ถูกกำกับ ดังอัลกอริทึมที่แสดงในภาพที่ 3.8 คือ ทำการวิเคราะห์คำ (Word) ในวลี (Phrase) ตามลำดับไปทีละคำ ถ้าคำใดเป็นชื่อตัวละครที่มียังไม่มีการกำหนดการกระทำ (w) จะถูกเก็บค่าไว้ในตัวแปร *cur_actor* = w และตัวแปร *best_action* = “” แล้ว เมื่อคำที่เป็นชื่อตัวละครอีกครั้งจะทำการกำหนดการกระทำตามตัวแปร *best_action* ให้กับตัวละครที่เก็บอยู่ในตัวแปร *cur_actor* และกำหนดให้ตัวแปร *cur_actor* เก็บชื่อตัวละครใหม่ที่ยังไม่มีการกำหนดการกระทำ แต่ถ้าคำใดที่ทำหน้าที่เป็นคำกริยาจะทำการอัพเดทตัวแปร *best_action* โดยจะให้เก็บค่าความสำคัญของคำกริยาที่บ่งบอกอาการพูด (Verb Classes) ที่มีความสำคัญตามลำดับดังนี้ Speech Verb, Thinking Verb, Action Verb, Other Verb และ none แล้วถ้าเมื่อสิ้นสุดวลีแล้ว ตัวแปร *cur_actor* และตัวแปร *best_action* ยังมีค่าอยู่ให้กำหนดการกระของตัวละครในตัวแปร *cur_actor* ตามค่าที่เก็บใน ตัวแปร *best_action*

```
Actor's Action Assignment
Input: word token of a phrase
Output: association of (actor_i, action_i) for every actor_i in the phrase

#set a pair of current actor and best action
cur_actor = null
best_action = null
Loop until no word in the phrase
  for each word w in the phrase:
    if w is actor with no action assigned :
      #check if there is already a pair found
      if cur_actor and best_action is not null
          action(cur_actor) = best_action
        clear values of cur_actor and best_action
      # then start with a new current actor
    cur_actor = w
    # if w is of type verb, set value of best_action
    else if  w is a {ST/TV/AV} verb
      if priority(w) > priority(best_action)
      best_action = w
End loop
action(cur_actor) = best_action
clear values of cur_actor and best_action
```

ภาพที่ 3.8 แสดงตัวอัลกอริทึมในการกำหนดการกระทำให้กับตัวละคร

จากการสำรวจตำแหน่งของผู้พูดที่ปรากฏในวลีต่างๆ กับบทพูดจากหนังสือนิทาน 40 เรื่อง พบว่า ตำแหน่งของตัวละครส่วนใหญ่จะปรากฏในวลีที่อยู่ติดกับบทพูด (Adjacent Phrase) และอยู่พารากราฟเดียวกับบทพูด (Same_paragraph) จำนวน 653 ตำแหน่ง และวลีที่อยู่ติดกับบทพูดที่มีตัวละครที่เป็นผู้พูดปรากฏอยู่ (Adjacent Phrase) ส่วนใหญ่จะตามหลังบทพูด (Following Phrase) จากการสำรวจพบเป็นจำนวน 472 วลี แสดงดังตารางที่ 3.3

เอกสารนี้สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ 3.3 แสดงความถี่ความสัมพันธ์ของตำแหน่งวลีที่แสดงผู้พูดกับความใกล้ชิดของวลีนั้นกับบทพูด

	Same_paragraph		Position of Phrase	
	Yes	No	Before	After
Adjacent_phrase	653	48	229	472
Non-Adjacent_phrase	21	3	21	3
Total	674	51	250	475

ดังนั้นจึงเกิดแนวความคิดในการระบุผู้พูดโดยใช้กฎฮิวริสติก (Heuristic Rule) ดังอัลกอริทึมที่แสดงดังภาพที่ 3.9 คือ ทำการวิเคราะห์บทพูดแต่ละบทในเรื่อง โดยพิจารณาที่ละบท โดยใช้ตัวละครที่ปรากฏอยู่ในพารากราฟเดียวกันกับบทพูดเป็นตัวละครที่อาจจะเป็นผู้พูด (Candidate Speaker) แล้วทำตามขั้นตอนดังนี้

Speaker Assignment Algorithm

Input: *tagged text structure and candidate speakers*
Output: *association of (quote$_i$, speaker) for each quote$_i$*

set of (quote, speaker) pairs in a story
$Q = \{(q_1, s_1), (q_2, s_2) \ldots, (q_n, s_n)\}$

Loop for each quote q_i
 ListActorBefore = list of actors from previous phrase
 ListActorAfter = list of actors from following phrase
 if the paragraph with quote has a candidate(s)
  sort ListActorBefore and ListActorAfter
  s_i = the highest priority actor from both lists
 else #no candidate in the same paragraph
  #get actors from a previous paragraph
  ListActorBefore = previous paragraph list of actors
  sort ListActorBefore
  s_i = the highest priority actor in ListActorBefore
 endif
 if s_i is empty and a previous phrase is a quote
  then #assign second last speaker as a speaker
   $s_i = s_{i-2}$ when s_{i-2} not same as s_{i-1}
  else #everything else fails
   # assign last speaker as a speaker
   $s_i = s_{i-1}$
 endif
End loop

ภาพที่ 3.9 แสดงตัวอัลกอริทึมในการระบุผู้พูดให้กับบทพูด

ถ้ามีตัวละครที่อยู่พารากราฟเดียวกันกับบทพูด แล้วทำการแบ่งตัวละครเหล่านั้นเป็น 2 กลุ่ม คือ ตัวละครที่ปรากฏก่อนหน้าบทพูด (ListActorBefore) และตัวละครที่ปรากฏตามหลังและอยู่ในวลีบรรยายที่อยู่ติดบทพูดนั้น (ListActorAfter) หลังจากนั้นทำการเรียงรายชื่อตัวละครแต่ละกลุ่ม โดยเรียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเอกสารนี้ทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตามลำดับความสำคัญ (Priority) ของการกระทำที่กำหนดให้ตัวละครตามชนิดของคำกริยาที่กำหนดให้กับตัวละคร คือ SV, TV, AV และOther ตามลำดับ และลำดับตามความใกล้ชิดของวลีที่ตัวละครปรากฏกับบทพูด (proximity to quote phrase) ทำการเปรียบเทียบตัวละครตัวแรกจากทั้งสองกลุ่ม ทำการเลือกตัวละครที่มีความสำคัญของการกระทำที่มากกว่า แต่ถ้าความสำคัญของตัวละครตัวแรกจากทั้งสองกลุ่มเท่ากัน อัลกอริทึมจากเลือกจากตัวละครที่ปรากฏตามหลังบทพูดหรือเลือกจาก ListActorAfter นั้นเอง

แต่ถ้าในพารากราฟที่กำลังพิจารณาบทพูดไม่มีตัวละครปรากฏอยู่ อัลกอริทึมจะทำการตรวจสอบวลีที่อยู่ก่อนบทพูดว่าเป็น บทพูดหรือบทบรรยายทั่วไป ถ้าวลีก่อนถ้าบทพูดที่กำลังพิจารณาอยู่เป็นบทพูด อัลกอริทึมจะเลือกผู้พูดล่าสุดลำดับที่สองให้เป็นผู้พูดสำหรับบทพูดที่กำลังพิจารณาอยู่ โดยงานวิจัยนี้ตั้งสมมุติฐานที่ว่า บทพูดนั้นเป็นการสนทนากันระหว่างตัวละคร (Alternative Speaker) หรือเมื่อพบว่าวลีที่อยู่ก่อนบทพูดเป็นบทบรรยายทั่วไป อัลกอริทึมจะใช้ตัวละครที่ได้จากพารากราฟก่อนหน้า (previous paragraph) เป็นผู้สมัครสำหรับการเลือกให้เป็นผู้พูด โดยจะทำการเรียงลำดับความสำคัญ (Priority) ของการกระทำตามชนิดของคำกริยา (Verb Classes) ที่ได้กำหนดให้กับละคร และลำดับตามความใกล้ชิดของวลีที่ตัวละครปรากฏกับบทพูด (proximity to quote phrase) ตัวละครตัวแรกที่ได้จากการเรียงลำดับจะถูกกำหนดให้เป็นผู้พูดสำหรับบทพูดที่กำลังพิจารณา

สำหรับในทุกกรณีที่นอกเหนือจากที่กล่าวมาอัลกอริทึมจะทำการกำหนดผู้พุดให้กับบทพูดจากตัวละครที่เป็นคนพูดล่าสุดให้กับบทพูดที่กำลังพิจารณา

สำหรับผลลัพธ์ของงานวิจัยนี้จะอยู่ในรูปแบบโครงสร้างทางภาษาของภาษา XML ที่มีการกำกับตัวละครให้กับบทพูดเรียบร้อยแล้ว และการทดลองและผลการทดลองของงานดำเนินงานวิจัยนี้ จะกล่าวถึงในบทที่ 4 ผลการวิจัย และสรุปผลการดำเนินงานวิจัย จะกล่าวในบทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บทที่ 4
ผลการวิจัย

บทนี้จะกล่าวถึงการวัดประสิทธิภาพของอัลกอริทึมที่ได้ออกแบบไว้ เพื่อตรวจสอบความแม่นยำ และความถูกต้องของการสกัดคำที่เป็นชื่อตัวละครที่มีบทพูด ซึ่งจะถูกใช้เป็นตัวเลือกสำหรับการระบุผู้พูดให้กับบทพูด และตรวจสอบความถูกต้องของการระบุผู้พูดให้กับบทพูด โดยการวัดประสิทธิภาพจากหนังสือนิทานภาษาไทยสำหรับเด็กจำนวน 40 เรื่อง ที่ประกอบด้วยพารากราฟจำนวน 1086 พารากราฟ บทพูด 825 บท วลีส่วนบรรยายที่อยู่ติดกับบทพูด 1020 วลี จากวลีส่วนบรรยายทั้งหมด 3078 วลี ที่มีจำนวนคำทั้งหมด 24,839 คำ

ในการวัดประสิทธิภาพของการคำที่เป็นชื่อตัวละครที่มีบทพูด จากหนังสือนิทานภาษาไทย 40 เรื่อง ในงานวิจัยนี้จะใช้การวัดค่าความถูกต้อง (Precision) และการวัดค่าครบถ้วนในการสกัด (Recall) ดังสมการที่ 4.1 และ 4.2 สำหรับการวัดประสิทธิภาพของการระบุผู้พูดให้กับบทพูด งานวิจัยนี้จะวัดความถูกต้องจากความถูกต้องของการระบุผู้พูดให้กับบทพูดที่ถูกต้อง หรือค่าความถูกต้อง (Precision) นั้นเอง โดยเปรียบเทียบกับผลลัพธ์ที่ได้จากผู้ใช้งาน งานวิจัยนี้ไม่วัดค่าความครบถ้วน (Recall) ของการระบุผู้พูดให้กับบทพูด เนื่องจากผลลัพธ์ที่ได้มีค่าเท่ากับค่าความถูกต้อง (Precision) เพราะ บทพูดที่ถูกเลือกโดยระบบ ก็คือจำนวนบทพูดทั้งหมดจากหนังสือนิทานภาษาไทยสำหรับเด็ก 40 เรื่อง 825 บทพูดนั้นเอง

$$Precision = \frac{RelevantRetrievedWordBySystem}{AllRetrievedWordBySystem} \quad (4.1)$$

$$Recall = \frac{RelevantRetrievedWordBySystem}{AllRelevantWordInAllStory} \quad (4.2)$$

โดยที่

RelevantRetrievedWordBySystem คือ จำนวนคำที่หมายถึงชื่อตัวละครที่ระบบสกัดได้ถูกต้อง
AllRetrievedWordBySystem คือ จำนวนคำที่หมายถึงชื่อตัวละครที่ระบบสกัดออกมาทั้งหมด
AllRelevantWordInAllStory คือ จำนวนคำที่หมายถึงชื่อตัวละครทั้งหมดจากหนังสือนิทาน

การทดลองเพื่อวัดประสิทธิภาพของงานวิจัยนี้แบ่งเป็น 3 การทดลอง คือ การวัดประสิทธิภาพการระบุผู้พูดเมื่อใช้ชื่อตัวละครที่ได้จากการสกัดจากวลีที่อยู่ติดกับบทพูดเท่านั้น, การวัดประสิทธิภาพการระบุผู้พูดเมื่อใช้ชื่อตัวละครที่ได้จากการทดลองที่ 1 รวมกับการกำกับชื่อตัวละครในวลีที่อยู่ติดกับบทพูดด้วยรายชื่อที่ระบบสร้างขึ้น และการวัดประสิทธิภาพการระบุผู้พูดเมื่อใช้ชื่อตัวละครจากการทดลองที่ 2 รวมกับการกำกับชื่อตัวละครที่เหลือทั้งหมดด้วยรายชื่อตัวละครที่ระบบสร้างขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

4.1 การวัดประสิทธิภาพการสกัดคำที่เป็นตัวละครในวลีที่อยู่ติดกับบทพูดด้วยกฎ และการระบุผู้พูดด้วยตัวละครที่สกัดได้

สำหรับการผลการทดลองในส่วนนี้ เป็นการวัดประสิทธิของกฎที่ใช้ในการสกัดคำที่เป็นชื่อตัวละครด้วยกฎเพียงอย่างเดียว และการใช้ชื่อตัวละครในการระบุผู้พูด ได้ผลการทดลองดังตารางที่ 4.1 ระบบสามารถสกัดคำที่เป็นชื่อตัวละครในวลีที่อยู่ติดกับบทพูดได้ 586 คำ จากคำที่เป็นชื่อตัวละครในวลีที่อยู่ติดกับบทพูดจำนวน 831 คำ ซึ่งมีความถูกต้อง 546 คำ คิดเป็นค่าความถูกต้อง 93.17% และค่าความครบถ้วน 65.70% หลังจากนั้นใช้ชื่อตัวละครที่สกัดได้เป็นตัวเลือกสำหรับการระบุผู้พูด ผลการระบุพูดจากหนังสือนิทานภาษาไทยสำหรับเด็กจำนวน 40 เรื่อง ที่มีบทพูดทั้งหมด 825 บทพูด อัลกอริทึมที่ได้คิดค้นสามารถระบุได้ถูกต้อง 683 บทพูด หรือคิดเป็น 82.79%

King Mongkut's Institute of Technology Ladkrabang

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ 4.1 การสกัดคำที่เป็นตัวละครในวลีที่อยู่ติดกับบทพูด และการระบุผู้พูดจากตัวละครนั้น

เรื่องที่	จำนวนคำที่เป็นชื่อตัวละคร	จำนวนบทพูด	จำนวนชื่อตัวละครที่ถูกสกัดโดยระบบ			Precision (%)	Recall (%)	จำนวนการระบุผู้พูด		ความถูกต้อง (%)
			ถูก	ผิด	รวม			ถูก	ผิด	
1	64	68	48	2	50	96.00	75.00	62	6	91.18
2	20	22	14	1	15	93.33	70.00	21	1	95.45
3	42	36	24	0	24	100.00	57.14	31	5	86.11
4	47	45	27	3	30	90.00	57.45	39	6	86.67
5	34	39	26	4	30	86.67	76.47	38	1	97.44
6	63	60	43	0	43	100.00	68.25	57	3	95.00
7	10	8	7	0	7	100.00	70.00	8	0	100.00
8	42	35	24	0	24	100.00	57.14	34	1	97.14
9	8	17	5	1	6	83.33	62.50	13	4	76.47
10	32	46	24	3	27	88.89	75.00	34	12	73.91
11	7	14	7	3	10	70.00	100.00	7	7	50.00
12	40	41	27	2	29	93.10	67.50	36	5	87.80
13	23	21	15	0	15	100.00	65.22	19	2	90.48
14	25	27	17	0	17	100.00	68.00	24	3	88.89
15	9	14	5	1	6	83.33	55.56	5	9	35.71
16	19	22	16	0	16	100.00	84.21	16	6	72.73
17	7	6	3	0	3	100.00	42.86	5	1	83.33
18	11	9	7	0	7	100.00	63.64	7	2	77.78
19	18	12	9	1	10	90.00	50.00	7	5	58.33
20	26	22	15	1	16	93.75	57.69	13	9	59.09

เอกสารนี้สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ ดัดแปลง เผยแพร่ หรือแสวงหาประโยชน์ทางการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น โดยไม่ได้รับอนุญาตเป็นลายลักษณ์อักษรจากสถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

เรื่องที่	จำนวนคำที่เป็นชื่อตัวละคร	จำนวนบทพูด	จำนวนชื่อตัวละครที่ถูกสกัดโดยระบบ			Precision (%)	Recall (%)	จำนวนการระบุผู้พูด		ความถูกต้อง (%)
			ถูก	ผิด	รวม			ถูก	ผิด	
21	64	58	43	0	43	100.00	67.19	43	15	74.14
22	31	23	22	1	23	95.65	70.97	20	3	86.96
23	8	4	4	0	4	100.00	50.00	4	0	100.00
24	1	1	1	2	3	33.33	100.00	1	0	100.00
25	19	15	14	0	14	100.00	73.68	14	1	93.33
26	2	3	2	0	2	100.00	100.00	3	0	100.00
27	2	2	2	0	2	100.00	100.00	2	0	100.00
28	2	1	1	0	1	100.00	50.00	1	0	100.00
29	28	25	15	0	15	100.00	53.57	25	0	100.00
30	6	9	5	2	7	71.43	83.33	6	3	66.67
31	21	28	17	4	21	80.95	80.95	21	7	75.00
32	7	4	2	0	2	100.00	28.57	2	2	50.00
33	9	11	8	0	8	100.00	88.89	10	1	90.91
34	17	8	6	0	6	100.00	35.29	4	4	50.00
35	12	15	9	3	12	75.00	75.00	13	2	86.67
36	21	12	8	2	10	80.00	38.10	7	5	58.33
37	2	1	1	0	1	100.00	50.00	1	0	100.00
38	5	3	3	0	3	100.00	60.00	3	0	100.00
39	10	21	10	2	12	83.33	100.00	15	6	71.43
40	17	17	10	2	12	83.33	58.82	12	5	70.59
รวม	**831**	**825**	**546**	**40**	**586**	**93.17**	**65.70**	**683**	**142**	**82.79**

เอกสารนี้สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเอกสารนี้ทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use. Forbidden to modify the content, and cite the document when use.

4.2 การวัดประสิทธิภาพการสกัดคำที่เป็นตัวละครด้วยกฎ และการกำกับชื่อตัวละครจากรายชื่อที่สร้างขึ้นในวลีที่อยู่ติดกับบทพูด และการระบุผู้พูดด้วยตัวละครที่สกัดและกำกับได้

สำหรับการผลการทดลองในส่วนนี้ เป็นการวัดประสิทธิของกฎที่ใช้ในการสกัดคำที่เป็นชื่อตัวละครร่วมกับการใช้รายชื่อที่สกัดได้ในการกำกับชื่อตัวละครในวลีที่อยู่ติดกับบทพูด แล้วใช้ชื่อตัวละครเหล่านั้นในการระบุผู้พูด ได้ผลการทดลองดังตารางที่ 4.2 ระบบสามารถสกัดคำที่เป็นชื่อตัวละครในวลีที่อยู่ติดกับบทพูดได้เพิ่มเติมจากการทดลองที่ 1 เป็น 858 คำ จากคำที่เป็นชื่อตัวละครในวลีที่อยู่ติดกับบทพูดจำนวน 831 คำ ซึ่งมีความถูกต้อง 807 คำ คิดเป็นค่าความถูกต้อง 94.06% และค่าความครบถ้วน 97.11% หลังจากนั้นใช้ชื่อตัวละครที่สกัดได้เป็นตัวเลือกสำหรับการระบุผู้พูด ผลการระบุพูดจากหนังสือนิทานภาษาไทยสำหรับเด็กจำนวน 40 เรื่อง ที่มีบทพูดทั้งหมด 825 บทพูด อัลกอริทึมที่ได้คิดค้นสามารถระบุได้ถูกต้อง 703 บทพูด หรือคิดเป็น 85.21%

ข้อสังเกตสำหรับการทดลองนี้พบว่าจำนวนชื่อตัวละครที่ระบบสามารถสกัดได้ทั้งหมดนั้นจำนวน 858 คำ ซึ่งมีค่ามากกว่าชื่อตัวละครที่มีจำนวน 831 คำ เนื่องจากว่าในขั้นตอนการสกัดชื่อตัวละครด้วยกฎเพื่อสร้างรายชื่อตัวละครนั้น มีการสกัดชื่อตัวละครที่ไม่ถูกต้องออกมาด้วย ทำให้รายชื่อตัวละครที่ใช้กำกับตัวละครมีคำที่ไม่ใช่ตัวละครอยู่ด้วย ซึ่งส่งผลให้การกำกับชื่อตัวละครนั้นกำกับคำที่ไม่ใช่ชื่อตัวละครออกมาด้วย และทำให้การสกัดชื่อตัวละครที่ได้มากกว่าชื่อตัวละครที่มีทั้งหมด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ 4.2 การสกัดคำที่เป็นตัวละครร่วมกับการกำกับชื่อตัวละครในวลีที่อยู่ติดกับบทพูด และการระบุผู้พูดจากตัวละครนั้น

เรื่องที่	จำนวนคำที่เป็นชื่อตัวละคร	จำนวนบทพูด	จำนวนชื่อตัวละครที่ถูกสกัดโดยระบบ			Precision (%)	Recall (%)	จำนวนการระบุผู้พูด		ความถูกต้อง (%)
			ถูก	ผิด	รวม			ถูก	ผิด	
1	64	68	63	2	65	96.92	98.44	62	6	91.18
2	20	22	20	1	21	95.24	100.00	21	1	95.45
3	42	36	42	0	42	100.00	100.00	32	4	88.89
4	47	45	47	5	52	90.38	100.00	39	6	86.67
5	34	39	33	4	37	89.19	97.06	38	1	97.44
6	63	60	63	0	63	100.00	100.00	58	2	96.67
7	10	8	10	0	10	100.00	100.00	8	0	100.00
8	42	35	42	0	42	100.00	100.00	34	1	97.14
9	8	17	6	1	7	85.71	75.00	12	5	70.59
10	32	46	30	4	34	88.24	93.75	37	9	80.43
11	7	14	7	5	12	58.33	100.00	7	7	50.00
12	40	41	40	2	42	95.24	100.00	38	3	92.68
13	23	21	23	0	23	100.00	100.00	20	1	95.24
14	25	27	25	0	25	100.00	100.00	25	2	92.59
15	9	14	6	1	7	85.71	66.67	5	9	35.71
16	19	22	19	0	19	100.00	100.00	16	6	72.73
17	7	6	7	0	7	100.00	100.00	5	1	83.33
18	11	9	11	0	11	100.00	100.00	7	2	77.78
19	18	12	17	4	21	80.95	94.44	8	4	66.67
20	26	22	25	1	26	96.15	96.15	15	7	68.18

This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

เรื่องที่	จำนวนคำที่เป็นชื่อตัวละคร	จำนวนบทพูด	จำนวนชื่อตัวละครที่ถูกสกัดโดยระบบ			Precision (%)	Recall (%)	จำนวนการระบุผู้พูด		ความถูกต้อง (%)
			ถูก	ผิด	รวม			ถูก	ผิด	
21	64	58	63	0	63	100.00	98.44	45	13	77.59
22	31	23	31	1	32	96.88	100.00	21	2	91.30
23	8	4	8	0	8	100.00	100.00	4	0	100.00
24	1	1	1	2	3	33.33	100.00	1	0	100.00
25	19	15	19	0	19	100.00	100.00	14	1	93.33
26	2	3	2	0	2	100.00	100.00	3	0	100.00
27	2	2	2	0	2	100.00	100.00	2	0	100.00
28	2	1	2	0	2	100.00	100.00	1	0	100.00
29	28	25	28	0	28	100.00	100.00	25	0	100.00
30	6	9	6	2	8	75.00	100.00	6	3	66.67
31	21	28	21	4	25	84.00	100.00	24	4	85.71
32	7	4	4	0	4	100.00	57.14	2	2	50.00
33	9	11	9	0	9	100.00	100.00	10	1	90.91
34	17	8	12	0	12	100.00	70.59	5	3	62.50
35	12	15	12	4	16	75.00	100.00	13	2	86.67
36	21	12	19	3	22	86.36	90.48	7	5	58.33
37	2	1	2	0	2	100.00	100.00	1	0	100.00
38	5	3	5	0	5	100.00	100.00	3	0	100.00
39	10	21	10	3	13	76.92	100.00	15	6	71.43
40	17	17	15	2	17	88.24	88.24	14	3	82.35
รวม	831	825	807	51	858	94.06	97.11	703	122	85.21

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารนี้ทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use. Forbidden to modify the content, and cite the document when use.

4.3 การวัดประสิทธิภาพการสกัดคำที่เป็นตัวละครด้วยกฎ และการกำกับชื่อตัวละครจากรายชื่อที่สร้างขึ้นสำหรับทุกวลี และการระบุผู้พูดด้วยตัวละครสกัดและกำกับได้ทั้งหมด

สำหรับการผลการทดลองในส่วนนี้ เป็นการวัดประสิทธิของกฎที่ใช้ในการสกัดคำที่เป็นชื่อตัวละครร่วมกับการใช้รายชื่อที่สกัดได้ในการกำกับชื่อตัวละครสำหรับทุกวลีการบรรยายภายในเรื่อง แล้วใช้ชื่อตัวละครเหล่านั้นในการระบุผู้พูด ได้ผลการทดลองดังตารางที่ 4.3 ระบบสามารถสกัดคำที่เป็นชื่อตัวละครทั้งหมดในเรื่องได้ 1569 คำ จากคำที่เป็นชื่อตัวละครทั้งหมดจำนวน 1512 คำ ซึ่งมีความถูกต้อง 1469 คำ คิดเป็นค่าความถูกต้อง 93.63% และค่าความครบถ้วน 97.16% หลังจากนั้นใช้ชื่อตัวละครที่สกัดได้เป็นตัวเลือกสำหรับการระบุผู้พูด ผลการระบุพูดจากหนังสือนิทานภาษาไทยสำหรับเด็กจำนวน 40 เรื่อง ที่มีบทพูดทั้งหมด 825 บทพูด อัลกอริทึมที่ได้คิดค้นสามารถระบุได้ถูกต้อง 706 บทพูด หรือคิดเป็น 85.58%

ข้อสังเกตสำหรับการทดลองนี้พบว่าจำนวนชื่อตัวละครที่ระบบสามารถสกัดได้ทั้งหมดนั้นจำนวน 1569 คำ ซึ่งมีค่ามากกว่าชื่อตัวละครที่มีจำนวน 1512 คำ เช่นเดียวกับการทดลองที่ 2 เนื่องจากว่าในขั้นตอนการสกัดชื่อตัวละครด้วยกฎเพื่อสร้างรายชื่อตัวละครนั้น มีการสกัดชื่อตัวละครที่ไม่ถูกต้องออกมาด้วย ทำให้รายชื่อตัวละครที่ใช้กำกับตัวละครมีคำที่ไม่ใช่ตัวละครอยู่ด้วย ซึ่งส่งผลให้การกำกับชื่อตัวละครนั้นกำกับคำที่ไม่ใช่ชื่อตัวละครออกมาด้วย และทำให้การสกัดชื่อตัวละครที่ได้มากกว่าชื่อตัวละครที่มีทั้งหมด

King Mongkut's Institute of Technology Ladkrabang

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ 4.3 การสกัดคำที่เป็นตัวละครร่วมกับการกำกับชื่อตัวละครในเรื่องทั้งหมด และการระบุผู้พูดจากตัวละครนั้น

เรื่องที่	จำนวนคำที่เป็นชื่อตัวละคร	จำนวนบทพูด	จำนวนชื่อตัวละครที่ถูกสกัดโดยระบบ			Precision (%)	Recall (%)	จำนวนการระบุผู้พูด		ความถูกต้อง (%)
			ถูก	ผิด	รวม			ถูก	ผิด	
1	101	68	100	2	102	98.04	99.01	62	6	91.18
2	25	22	25	1	26	96.15	100.00	21	1	95.45
3	59	36	59	0	59	100.00	100.00	32	4	88.89
4	83	45	83	5	88	94.32	100.00	39	6	86.67
5	56	39	55	5	60	91.67	98.21	37	2	94.87
6	96	60	96	0	96	100.00	100.00	58	2	96.67
7	16	8	16	0	16	100.00	100.00	8	0	100.00
8	65	35	65	0	65	100.00	100.00	33	2	94.29
9	9	17	7	1	8	87.50	77.78	12	5	70.59
10	42	46	37	4	41	90.24	88.10	37	9	80.43
11	23	14	19	20	39	48.72	82.61	8	6	57.14
12	87	41	87	3	90	96.67	100.00	38	3	92.68
13	34	21	34	0	34	100.00	100.00	20	1	95.24
14	40	27	40	0	40	100.00	100.00	25	2	92.59
15	29	14	22	1	23	95.65	75.86	7	7	50.00
16	30	22	30	0	30	100.00	100.00	17	5	77.27
17	16	6	16	0	16	100.00	100.00	5	1	83.33
18	41	9	41	0	41	100.00	100.00	7	2	77.78
19	35	12	34	14	48	70.83	97.14	8	4	66.67
20	37	22	36	1	37	97.30	97.30	15	7	68.18

This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

เรื่องที่	จำนวนคำที่เป็นชื่อตัวละคร	จำนวนบทพูด	จำนวนชื่อตัวละครที่ถูกสกัดโดยระบบ			Precision (%)	Recall (%)	จำนวนการระบุผู้พูด		ความถูกต้อง (%)
			ถูก	ผิด	รวม			ถูก	ผิด	
21	75	58	74	0	74	100.00	98.67	45	13	77.59
22	43	23	43	2	45	95.56	100.00	22	1	95.65
23	27	4	27	0	27	100.00	100.00	4	0	100.00
24	3	1	3	2	5	60.00	100.00	1	0	100.00
25	63	15	63	0	63	100.00	100.00	14	1	93.33
26	15	3	15	0	15	100.00	100.00	3	0	100.00
27	6	2	6	0	6	100.00	100.00	2	0	100.00
28	5	1	5	0	5	100.00	100.00	1	0	100.00
29	50	25	50	0	50	100.00	100.00	25	0	100.00
30	16	9	16	6	22	72.73	100.00	6	3	66.67
31	47	28	46	8	54	85.19	97.87	24	4	85.71
32	12	4	7	0	7	100.00	58.33	2	2	50.00
33	18	11	18	0	18	100.00	100.00	10	1	90.91
34	35	8	25	0	25	100.00	71.43	5	3	62.50
35	29	15	29	10	39	74.36	100.00	13	2	86.67
36	32	12	30	5	35	85.71	93.75	7	5	58.33
37	14	1	14	0	14	100.00	100.00	1	0	100.00
38	38	3	38	0	38	100.00	100.00	3	0	100.00
39	22	21	22	5	27	81.48	100.00	15	6	71.43
40	38	17	36	5	41	87.80	94.74	14	3	82.35
รวม	**1512**	**825**	**1469**	**100**	**1569**	**93.63**	**97.16**	**706**	**119**	**85.58**

Forbidden to modify the content, and cite the document when use.
This material is reserved for educational use only, not allowed for commercial use.

บทที่ 5
สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

งานวิจัยนี้นำเสนอการระบุผู้พูดจากหนังสือนิทานภาษาไทย โดยไม่ใช้รายชื่อตัวละครที่สร้างโดยผู้ใช้งาน เช่น (Glass & Bangay, 2006) (Glass และ Bangay, 2007) ที่นำเสนอการระบุผู้พูด โดยทั้งสองงานจำเป็นต้องสร้างรายชื่อตัวละครโดยผู้งานไว้ก่อน เพื่อใช้ในการระบุผู้พูด โดยงานวิจัยนี้นำเสนอการสร้างรายชื่อตัวละครโดยอัตโนมัติ ซึ่งใช้ในการกำกับชื่อตัวละคร เพื่อใช้เป็นตัวเลือกสำหรับการระบุผู้พูด และการระบุผู้พูดโดยใช้กฎฮิวริสติก (Heuristics) โดยไม่มีการเรียนรู้ข้อมูล (Training Data) ทำให้ผลลัพธ์ของการระบุผู้พูดไม่ขึ้นอยู่กับข้อมูลชุดฝึกสอน (Training set) แต่งานวิจัยนี้ยังไม่มีจัดกลุ่มตัวละครที่เป็นตัวละครเดียวกัน และแยกแยะตัวละครที่มีชื่อเดียวกัน กล่าวคือ การแก้ปัญหาการใช้คำสรรพนามแทนชื่อตัวละคร การใช้คำนามเรียกแทนชื่อตัวละคร และการระบุตัวละครที่ใช้ชื่อเดียวกันแต่หมายถึงตัวละครต่างกัน

สำหรับนิทานที่ใช้ในการทดลองเป็นนิทานภาษาไทยสำหรับทั้งหมด 40 เรื่อง มีบทพูด 825 บท โดยการแบ่งการทดลองเป็น 3 การทดลอง คือ การระบุผู้พูดโดยใช้ตัวละครจากชื่อตัวละครที่สกัดได้ด้วยกฎเพียงอย่างเดียว, การระบุผู้พูดโดยใช้ตัวละครจากชื่อตัวละครที่สกัด และการกำกับชื่อตัวละครในวลีที่อยู่ติดกับบทพูดด้วยรายชื่อตัวละครที่สร้างขึ้น และการระบุผู้พูดโดยใช้ตัวละครจากชื่อตัวละครที่สกัด และการกำกับชื่อตัวละคร ในส่วนที่เหลือจากการทดลองที่ 2 ด้วยรายชื่อตัวละครที่สร้างขึ้น ตามลำดับ ได้ผลการทดลองตามบทที่ 4 ที่ได้กล่าวไปแล้วคือ ความถูกต้องของการระบุผู้เป็น 82.79%, 85.21% และ 85.58% ตามลำดับ เป็นผลลัพธ์ที่ค่อนข้างดี สำหรับความถูกต้องของการสกัดคำที่เป็นตัวละครที่มีบทพูดได้ 93.17%, 94.06% และ 93.63% ส่วนค่าความครบถ้วนเป็น 65.70%, 97.11% และ 97.16% ตามลำดับ จากจำนวนของคำที่เป็นผู้พูดที่ถูกสกัดออกมา 586, 858 และ 1569 คำ ตามลำดับ จะสังเกตได้ว่าการเพิ่มขึ้นของตัวเลือกที่จะใช้ในการระบุผู้พูดจากการทดลองที่ 2 กับ 3 ทำให้การระบุผู้พูดถูกต้องเพิ่มขึ้นเพียงเล็กน้อย ซึ่งการทดลองที่ 2 ใช้ตัวละครในวลีที่อยู่ติดกับบทพูดแต่การทดลองที่ 3 ใช้ชื่อตัวละครในส่วนอื่นๆ ของเรื่องด้วย และการทดลองที่ 1 ใช้เพียงชื่อตัวละครเพียงบางส่วนของวลีที่อยู่ติดกับบทพูด ดังนั้นจึงสรุปได้ว่าผู้พูดโดยส่วนใหญ่จะอยู่ในวลีที่อยู่ติดกับบทพูดไม่ว่าอยู่ก่อนหรือหลังบทพูด ซึ่งสนับสนุนการใช้กฎที่สกัดชื่อตัวละครที่ปรากฏในวลีที่อยู่ติดกับบทพูด ทำให้การระบุผู้พูดได้ผลลัพธ์ที่ดี

5.2 ข้อเสนอแนะ

เนื่องจากการสกัดชื่อตัวละครของงานวิจัยนี้ ไม่มีการจัดกลุ่มของตัวละครที่เป็นตัวละครเดียวกัน และระบุตัวละครคนละตัวแต่ที่ใช้ชื่อเรียกเดียวกัน ดังนั้นจึงเป็นงานวิจัยในอนาคตที่จะต้องแก้ปัญหาการใช้สรรพนาม คำนาม และชื่อต่างๆ ที่ใช้เรียกแทนตัวละครเดียวกัน โดยจะต้องสามารถระบุให้เป็นตัวละครเดียวกันให้ถูกต้อง และการแก้ปัญหาการที่ตัวละครมีชื่อเรียกเดียวกันแต่เป็นคนละตัว เช่น ตัวละครที่ชื่อ “คนตัดไม้” ซึ่งจากเนื้อมีการกล่าวถึงคนตัดไม้คนแรกเมื่อต้นแรก และท้ายเรื่องได้กล่าวถึงคนตัดไม้อีกคนหนึ่ง แต่ใช้คำเรียกที่เหมือนกันคือ “คนตัดไม้”

อย่างไรก็ตามงานวิจัยนี้ส่วนใหญ่ใช้กฎฮิวริสติก (Heuristics) ในการประมวลผลที่ได้ผลดีในระดับหนึ่ง ซึ่งการนำเทคนิคของการเรียนรู้ของเครื่องจักรมาประยุกต์จากกฎที่ได้คิดค้น อาจทำให้ผลลัพธ์ในการระบุผู้พูดได้ผลดีเพิ่มขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

บรรณานุกรม

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python* (1st ed.). USA: O'Reilly Media, Inc.

Glass, K., & Bangay, S. (2006). Hierarchical Rule Generalisation for Speaker Identification in Fiction Books. In *proceedings of SAICSIT'06* (pp. 31-40). South African: South African Institute for Computer Scientists and Information Technologists.

Glass, K., & Bangay, S. (2007). A Naïve, Salience-Based Method for Speaker Identification in Fiction Books. In *proceedings of the 18th International Symposium of the Pattern Recognition Association of South Africa* (pp. 1-6). Pietermaritzburg, South Africa: PRASA.

Kong, F., GuoDong, Z., & Qiaoming, Z. (2009). Employing the Centering Theory in Pronoun Resolution from the Semantic Perspective. In *proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing* (pp. 987-996). Singapore: Association for Computational Linguistics Stroudsburg, PA, USA.

NAiST Lab, K. U. (2011). *Jitar model and Jitar 20100224*. Retrieved August 08, 2013, from http://naist.cpe.ku.ac.th/pkg/

Sasidhar, B., Yohan, P. M., Vinaya Babu, A., & Govardhan, A. (2011). A Survey on Named Entity Recognition in Indian Languages with particular reference to Telugu. In *IJCSI International Journal of Computer Science Issues*, (pp. 438-443).

Satayamas, V. (2012). *Part-of-speech tagger for Thai Language*. Retrieved August 08, 2013, from http://veer66.wordpress.com/tag/pos/

Soon, M. W., Ng, T. H., & Lim, Y. ,. (2001). A Machine Learning Approach to Coreference Resolution of Noun Phrases. *Computational Linguistics, 27*(4), 521-544.

Sutheebanjard, P., & Premchaiswadi, W. (2009). Thai Personal Named Entity Extraction without using Word Segmentation or POS Tagging. In *Eighth International Symposium on Natural Language Processing*, (pp. 221-226). Bangkok.

Zhang, J. Y., Black, A. W., & Sproat, R. (2003). Identifying Speaker in Children's Stories for Speech Synthesis. In *proceedings of EUROSPEECH 2003*, (pp. 2041-2044). Geneva, Switzerland.

ณัฐธิดา เตชะนภารักษ์, สุภรณี กัลป์ยาณกุล, และ อรณี นิลศรีไพรวัลลย์. (2011). *การจำแนกข้อความประเภทนวนิยายตามตัวละครแบบกึ่งอัตโนมัติ* (เอกสารทางวิชาการ หมายเลข 13p33c002). ศูนย์เทคโนโลยีอีเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ. ค้นเมื่อวันที่ 8 สิงหาคม 2556, จาก http://thailang.nectec.or.th/halloffame/images/stories/best/download/13p33c002.pdf

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use. Forbidden to modify the content, and cite the document when use.

วรศักดิ์ ตั้งกุลทวีทรพย์, พีระศักดิ์ รัตนมณี, และ ธนพล จิดาพิทักษ์. (2010). *การแบ่งคำภาษาไทยด้วยเทคนิคไฮบริด* (เอกสารทางวิชาการ หมายเลข 34S001). ศูนย์เทคโนโลยีอีเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ. ค้นเมื่อวันที่ 8 สิงหาคม 2556, จาก http://thailang.nectec.or.th/halloffame/images/stories/best/download/best2010_12p34s001.pdf

อัศนีย์ ก่อตระกูล. (2007). *การพัฒนาระบบสกัดข้อสนเทศและความรู้จากเอกสารไร้โครงสร้างภาษาไทย* (เอกสารทางวิชาการ หมายเลข NT-B-22-14-12-46-06). กรุงเทพมหานคร: ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์.

King Mongkut's Institute of Technology Ladkrabang

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ภาคผนวก ก
คำกริยาที่แสดงอาการพูดจากหนังสือนิทานสำหรับเด็ก 40 เรื่อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ ก.1 คำกริยาที่แสดงอาการพูด และตำแหน่งวลีที่พบคำกริยานั้น และตำแหน่งของชื่อตัวละครที่กระทำกริยานั้นกับคำกริยา

คำกริยา	ตำแหน่งของวลีที่พบคำกริยากับบทพูด						ตำแหน่งของตัวละที่กระทำกริยาที่พบ		
	ติดหรือไม่ติดกับบทพูด		ก่อนหรือหลังบทพูด		วลีอยู่พารากราฟเดียวกันกับบทพูด				
	ไม่ติด	ติด	ก่อน	หลัง	ไม่ใช่	ใช่	ไม่ติดกับคำกริยา	ระบุไม่ได้	ติดกับคำกริยา
ตอบ	2	57	46	13	2	57	2	2	55
ตะกุกตะกัก	0	2	1	1	0	2	0	0	2
ตะโกน	0	13	10	3	0	13	1	0	12
ตะคอก	0	1	0	1	0	1	1	0	0
ตั้งข้อเสนอ	0	2	1	1	0	2	0	0	2
ตัดสินใจ	0	2	0	2	2	0	2	0	0
ตำหนิ	0	2	1	1	0	2	0	0	2
เตือน	0	3	2	1	0	3	0	0	3
ถอนหายใจ	0	5	3	2	0	5	1	0	4
ถาม	2	46	40	8	4	44	2	0	46
ท่องคาถา	0	1	1	0	0	1	0	0	1
ทัก	0	6	5	1	1	5	0	0	6
ทักทาย	1	1	1	1	0	2	0	0	2
นึกถึง	0	1	0	1	1	0	1	0	0
นึกในใจ	0	1	1	0	0	1	0	0	1
แนะนำ	0	2	1	1	0	2	0	0	2
บ่น	0	2	2	0	0	2	0	0	2
บอก	2	47	39	10	0	49	4	0	45
ประกาศ	0	3	2	1	1	2	1	0	2

Forbidden to modify the content, and cite the document when use.
This material is reserved for educational use only, not allowed for commercial use.

ตารางที่ ก.1 คำกริยาที่แสดงอาการพูด และตำแหน่งวลีที่พบคำกริยานั้น และตำแหน่งของชื่อตัวละครที่กระทำกริยานั้นกับคำกริยา (ต่อ)

คำกริยา	ตำแหน่งของวลีที่พบคำกริยากับบทพูด						ตำแหน่งของตัวละที่กระทำกริยาที่พบ		
	ติดหรือไม่ติดกับบทพูด		ก่อนหรือหลังบทพูด		วลีอยู่พารากราฟเดียวกันกับบทพูด				
	ไม่ติด	ติด	ก่อน	หลัง	ไม่ใช่	ใช่	ไม่ติดกับคำกริยา	ระบุไม่ได้	ติดกับคำกริยา
ประหลาดใจ	0	2	1	1	1	1	2	0	0
ปลอบ	0	3	1	2	1	2	1	0	2
แปลกใจ	0	5	3	2	2	3	2	0	3
ผงกหัว	0	1	1	0	0	1	0	0	1
ฝันเฟื่อง	1	1	1	1	0	2	2	0	0
พึมพำ	1	11	8	4	0	12	0	0	12
พูด	8	195	119	84	15	188	19	7	177
ไม่เห็นด้วย	0	1	1	0	0	1	0	0	1
ยกมือกุมขมับ	0	1	0	1	1	0	1	0	0
ยอมรับ	0	2	1	1	0	2	0	0	2
ยิ้ม	0	10	6	4	0	10	2	0	8
ยืนกราน	0	3	2	1	0	3	0	0	3
ยื่นข้อเสนอ	0	1	0	1	1	0	0	0	1
เยาะเย้ย	0	2	1	1	0	2	0	0	2
แยกเขี้ยวคำราม	0	1	1	0	0	1	0	0	1
แย้ง	0	3	2	1	0	3	0	0	3
ร้อง	1	59	39	21	4	56	4	2	54
ร้องไห้	0	3	1	2	1	2	0	0	3
รับคำ	0	1	1	0	0	1	0	0	1

Forbidden to modify the content, and cite the document when use.
This material is reserved for educational use only, not allowed for commercial use.
เอกสารนี้สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเอกสารนี้ทุกครั้งที่มีการใช้งาน ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งไม่อนุญาตให้นำไปใช้ในรูปแบบอื่นใดที่ยังไม่ได้รับอนุญาตเป็นลายลักษณ์อักษรจากผู้เขียนเท่านั้น

ตารางที่ ก.1 คำกริยาที่แสดงอาการพูด และตำแหน่งวลีที่พบคำกริยานั้น และตำแหน่งของชื่อตัวละครที่กระทำกริยานั้นกับคำกริยา (ต่อ)

คำกริยา	ตำแหน่งของวลีที่พบคำกริยากับบทพูด						ตำแหน่งของตัวละที่กระทำกริยาที่พบ		
	ติดหรือไม่ติดกับบทพูด		ก่อนหรือหลังบทพูด		วลีอยู่พารากราฟเดียวกันกับบทพูด				
	ไม่ติด	ติด	ก่อน	หลัง	ไม่ใช่	ใช่	ไม่ติดกับคำกริยา	ระบุไม่ได้	ติดกับคำกริยา
รับรอง	0	2	1	1	0	2	0	0	2
ร่ายมนต์	0	3	1	2	2	1	3	0	0
รำพึง	0	1	1	0	0	1	0	0	1
เร่ง	0	1	1	0	0	1	0	0	1
เริ่มต้น	0	1	1	0	0	1	0	0	1
ล้อ	1	1	1	1	0	2	0	0	2
ล้อเลียน	0	4	2	2	0	4	0	0	4
เล่า	0	7	4	3	2	5	3	0	4
เลียริมฝีปาก	0	2	1	1	0	2	0	0	2
ว่า	0	17	12	5	1	16	0	1	16
โวยวาย	0	1	1	0	0	1	0	0	1
ส่งสัญญาณ	0	1	1	0	0	1	0	0	1
สงสัย	0	1	1	0	0	1	0	0	1
ส่งเสียงดัง	0	1	1	0	1	0	0	0	1
ส่งเสียงเตือน	1	3	3	1	2	2	0	0	4
ส่ายหัว	0	1	0	1	0	1	0	0	1
เสนอ	0	4	3	1	0	4	0	0	4
เสริม	0	1	0	1	0	1	0	0	1
แสดงความคิดเห็น	0	2	1	1	0	2	0	0	2

This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ ก.1 คำกริยาที่แสดงอาการพูด และตำแหน่งวลีที่พบคำกริยานั้น และตำแหน่งของชื่อตัวละครที่กระทำกริยานั้นกับคำกริยา (ต่อ)

คำกริยา	ตำแหน่งของวลีที่พบคำกริยากับบทพูด						ตำแหน่งของตัวละที่กระทำกริยาที่พบ		
	ติดหรือไม่ติดกับบทพูด		ก่อนหรือหลังบทพูด		วลีอยู่พารากราฟเดียวกันกับบทพูด				
	ไม่ติด	ติด	ก่อน	หลัง	ไม่ใช่	ใช่	ไม่ติดกับคำกริยา	ระบุไม่ได้	ติดกับคำกริยา
แสดงความดีใจ	0	1	0	1	0	1	0	0	1
หอบแฮก	0	3	2	1	0	3	0	0	3
หัวเราะ	1	14	9	6	2	13	2	0	13
เห็นด้วย	0	2	2	0	0	2	0	0	2
ให้ความเห็น	0	5	3	2	1	4	4	0	1
อธิบาย	0	8	5	3	1	7	0	0	8
ออกความเห็น	0	2	2	0	0	2	0	0	2
อ้าง	0	1	1	0	0	1	1	0	0
อ่าน	0	2	1	1	0	2	0	0	2
อ้าปากกว้าง	0	1	1	0	0	1	0	0	1
อุทาน	0	7	5	2	0	7	1	0	6
เอ่ยขึ้น	0	1	1	0	0	1	0	0	1
โอ้อวด	0	2	1	1	0	2	0	0	2
ไอแค้กๆ	0	2	1	1	0	2	0	0	2
รวม	**24**	**701**	**475**	**250**	**51**	**674**	**64**	**12**	**649**

Forbidden to modify the content, and cite the document when use.
This material is reserved for educational use only, not allowed for commercial use.

ภาคผนวก ข
โครงสร้างของนามวลีจากการสำรวจนิทานภาษาไทย 40 เรื่อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ ข.1 โครงสร้างนามวลีที่เป็นชื่อตัวละครจากการสำรวจนิทานภาษาไทยสำหรับเด็ก 40 เรื่อง

ที่	โครงสร้างนามวลี	ตัวอย่างของนามวลี
1	ncn ncn adj vt adj cl adj	กระรอก สี น้ำตาล อม แดง ตัว หนึ่ง
2	ncn prel prev prev vt ncn vpost	กบ ที่ ยัง พอ มี แรง อยู่
3	ncn prep ntit npn adj cl det	ครอบครัว ของ เจ้า ค้างคาว น้อย ตัว นี้
4	ncn adj cl prel vt ncn	ชาย ชรา คน ที่ ถือ ไม้เท้า
5	ncn ncn adj adj cl adj	หนังสือ ประวัติศาสตร์ เก่า คร่ำคร่า เล่ม หนึ่ง
6	ncn ncn adj cl norm nnum	หนู สี ขาว ตัว ที่ สอง
7	ncn ncn adj ncn ncn npn	ลูก สาว สุด ที่รัก ชื่อ ลูอิส
8	ncn ncn ncn adj cl adj	สุนัข บ้าน สี เข้ม ตัว หนึ่ง
9	ncn ncn ncn adj vt adj	หนู ขน สี เทา อม ฟ้า
10	ncn prep ncn vt ncn ncn	หนู ใน โรงงาน เก็บ เมล็ด ข้าว
11	ncn vt ncn adj cl adj	คน ตัด ไม้ อีก คน หนึ่ง
12	ntit ncn ncn adj cl det	เจ้า แมว ขี้ ร้าย ตัว นั้น
13	ntit ncn ncn ncn ncn adj	เจ้า นก เค้า แมว จอม โหด
14	ncn ncn ncn cl adj	นก เค้า แมว ตัว หนึ่ง
15	ncn prep ncn vt ncn	หนู ใน โรง รีด นม
16	ntit ncn adj cl adj	คุณ ยาย ใจดี คน หนึ่ง
17	ncn adj adj prep pper	ปาก บาง ยาว ของ มัน
18	ncn adj cl norm nnum	หนู ขาว ตัว ที่ สอง
19	ncn adj prel vt ncn	ชาย ชรา ที่ ถือ ไม้เท้า
20	ncn ncn adj cl adj	เจ้าของ ปราสาท อีก หลัง หนึ่ง
21	ncn ncn adj prep pper	กระโปรง สี ชมพู ของ เธอ
22	ncn ncn prep ncn ncn	ลูก สาว ของ เจ้าของ ฟาร์ม
23	ncn ncn prep ntit ncn	ใย ทองคำ ของ เจ้า แมงมุม
24	ncn prep ncn cl det	ขา ของ คน เหล่า นั้น
25	ncn prep ncn conjc nnum	สีข้าง ของ หนู ทั้ง สอง
26	ncn prep ncn ncn npn	พ่อ กับ แม่ นก สาลิกา
27	ncn vi adj ncn adj	เสื้อ กัน หนาว สี แดง
28	ncn vt ncn adj cl	คน ตัด ไม้ อีก คน
29	ntit ncn ncn cl adj	เจ้า หนังสือ นิทาน เล่ม น้อย
30	ncn prep ncn ncn	ภรรยา ของ เจ้าของ ฟาร์ม
31	ncn adj cl adj	กวี หนุ่ม คน หนึ่ง
32	ncn adj nnum cl	หนู ฉลาด สาม ตัว
33	ncn ncn cl adj	ต้น สน ต้น หนึ่ง
34	ncn npn cl adj	สุนัข จิ้งจอก ตัว หนึ่ง
35	pref2 vt	ผู้ ก่อตั้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ ข.1 โครงสร้างนามวลีที่เป็นชื่อตัวละครจากการสำรวจนิทานภาษาไทยสำหรับเด็ก 40 เรื่อง (ต่อ)

ที่	โครงสร้างนามวลี	ตัวอย่างของนามวลี
36	ntit ncn ncn ncn	เจ้า นก เค้า แมว
37	ncn adj prep pper	เพื่อน ใหม่ ของ มัน
38	ncn ncn adj cl	สมาชิก หนู ทุก ตัว
39	ncn prel adj adj	กบ ที่ อาวุโส ที่ สุด
40	ntit ncn cl det	เจ้า แมงมุม ตัว นี้
41	honm ncn cl adj	พระ ราชา องค์ แรก
42	ncn adj ncn ncn	หงส์ ประหลาด สี ทอง
43	ncn adj vt ncn	คน ที่ ถือ ไม้เท้า
44	ncn cl norm nnum	กบ ตัว ที่ สาม
45	ncn ncn adj adj	บุรุษ ไปรษณีย์ หนุ่ม น้อย
46	ncn ncn adv ncn	สถานี รถ ประจำ ทาง
47	ncn ncn ncn adj	หนู ขน สี ทอง
48	ncn ncn ncn npn	บุรุษ ไปรษณีย์ ชื่อ เฟลิกซ์
49	ncn ncn nct det	เจ้าของ ปราสาท เหล่า นี้
50	ncn ncn nnum cl	ผึ้ง งาน สาม ตัว
51	ncn ncn prep pper	ใย ทองคำ ของ มัน
52	ncn npn adj adj	ต้น โอ๊ก เก่า แก่
53	ncn npn nnum cl	นก นางแอ่น สอง ตัว
54	ncn prep ncn npn	พ่อ ของ นก สาลิกา
55	ncn prep nct pper	รัง ของ พวก มัน
56	ncn prep ntit ncn	ใย ของ เจ้า แมงมุม
57	ncn punc prep pper	เพื่อน ๆ ของ มัน
58	ncn vi prep ncn	หนังสือ เกี่ยว กับ สุขภาพ
59	ncn vt cl adj	คน รับใช้ คน หนึ่ง
60	nct ncn adj det	พวก แมลง แถว นั้น
61	nct ncn prep ncn	พวก นกยูง ใน สวน
62	ncn prep pref3 ncn	ไก่ ของ ชาว บ้าน
63	nct pper adj cl	พวก มัน ทั้ง ฝูง
64	npn ncn prep ncn	อึ หนู ใน น้ำผึ้ง
65	npn ncn prep pper	จูเลียน เพื่อน ของ เธอ
66	ntit ncn npn adj	เจ้า สุนัข จิ้งจอก เจ้าเล่ห์
67	ntit ncn prep ncn	คุณ ปู่ ของ เธอ
68	ntit ncn vt ncn	คุณ ป้า ขาย ข้าวแกง
69	ntit npn ncn adj	เจ้า สะพานลอย จอม เกเร
70	ntit npn ncn ncn	เจ้า หงส์ ขน ทอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ ข.1 โครงสร้างนามวลีที่เป็นชื่อตัวละครจากการสำรวจนิทานภาษาไทยสำหรับเด็ก 40 เรื่อง (ต่อ)

ที่	โครงสร้างนามวลี	ตัวอย่างชื่อตัวละคร
71	pref1 vt prep pper	ความ เศร้า ของ เขา
72	pref2 vt ncn adj	ผู้ นำ ตัว ใหม่
73	pref3 ncn cl adj	ชาว ไร่ คน หนึ่ง
74	ncn ncn adj	หนู จำนวน มาก
75	ncn cl adj	นก ตัว หนึ่ง
76	ncn cl det	นก ตัว นั้น
77	ncn ncn ncn	คุณ ไปรษณีย์ นก
78	ncn prep pper	ภรรยา ของ เขา
79	ncn vt ncn	ฟาร์ม เลี้ยง ม้า, คน ตัด ไม้
80	ncn adj cl	หนังสือ ทุก เล่ม
81	ncn prep ncn	ราชา แห่ง นก, หนู ใน หนอง
82	ntit ncn npn	คุณ ยาย บรรณารักษ์
83	ncn npn adj	แมลง ทับ ซรา
84	nct ncn adj	พวก ม้า ทั้งหลาย
85	ncn nnum cl	นกเขา สอง ตัว, เด็ก ห้า ขวบ
86	ncn cl indet	ม้า ตัว อื่น
87	ncn ncn indet	เป็ด ตัว อื่นๆ
88	ncn ncn npn	ดอก ป๊อป ปี้, นักบิน วัย กลางคน
89	ncn pref2 ncn	กบ ผู้ เฒ่า
90	ncn prep npn	พ่อ ของ ลีน่า
91	npn cl adj	ชาย คน หนึ่ง
92	ntit ncn adj	เจ้า นก น้อย
93	ntit ncn ncn	เจ้า หนังสือ นิทาน
94	ntit npn adj	เจ้า ค้างคาว น้อย
95	ncn adj ncn	ผึ้ง ทุก ตัว
96	ncn adj nnum	เด็ก ทั้ง สอง
97	ncn adj npn	ผี น้อย ลูแตง
98	ncn conjc nnum	หนู ทั้ง สอง
99	ncn nct adj	คน กลุ่ม หนึ่ง
100	ncn nct det	คน พวก นั้น
101	ncn pref3 ncn	ภรรยา ชาว นา
102	ncn prel adj	กบ ที่ อาวุโส ที่สุด
103	ncn vt vt	คน รับ ใช้
104	nct ncn npn	เหล่า กองเชียร์ เต่า
105	nnum ncn adj	สอง พี่น้อง ฝาแฝด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ตารางที่ ข.1 โครงสร้างนามวลีที่เป็นชื่อตัวละครจากการสำรวจนิทานภาษาไทยสำหรับเด็ก 40 เรื่อง (ต่อ)

ที่	โครงสร้างนามวลี	ตัวอย่างชื่อตัวละคร
106	npn ncn adj	ราชินี ตัว น้อย
107	npn ncn ncn	หงส์ ขน ทอง
108	npn pref2 adj	หงส์ ผู้ เย่อหยิ่ง
109	ntit nct det	เจ้า พวก นี้
110	ntit npn adv	เจ้า หนู น้อย
111	pper nnum cl	เขา สอง คน
112	pref2 ncn adj	ผู้ เคราะห์ ร้าย
113	pref2 ncn npn	นัก ดนตรี ตาบอด
114	pref2 prev adj	ผู้ ยิ่ง ใหญ่
115	pref2 vcs ncn	ผู้ เป็น พ่อ
116	pref2 vt ncn	ผู้ นำ ข่าว
117	pind advm4 cl	ใคร สัก คน
118	pind cl adl	ใคร คน หนึ่ง
119	ncn ncn	นก ไปรษณีย์, หนู นา, หนู บ้าน
120	ncn npn	นก กระสา
121	ncn adj	สะพานลอย เกเร
122	ntit ncn	เจ้า ด่าง
123	ntit npn	เจ้า สะพานลอย
124	nct ncn	พวก นกยูง
125	pref3 ncn	ชาว บ้าน
126	npn adj	ลูอิส น้อย
127	adj cl	ทุก ตัว
128	nct npn	พวก เต่า
129	nct pper	พวก มัน
130	honm ncn	พระ ราชา
131	honm npn	พระ บาท
132	ncn det	ดอกไม้ นั้น
133	ncn indet	ม้า อื่นๆ
134	ncn prfx	ตัว เอง
135	npn ncn	ราชินี ผึ้ง
136	pref1 vi	ความ กระตือรือร้น
137	pref1 vt	ความ โกรธ
138	pref2 ncn	นัก วิทยาศาสตร์
139	ncn	นกยูง, นักบิน, กระต่าย
140	npn	ลีน่า, ซานตาคลอส, มาร์ติน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

ประวัตินักวิจัย

ชื่อ - นามสกุล (ภาษาไทย) รศ.ดร. พรฤดี เนติโสภากุล
ชื่อ - นามสกุล (ภาษาอังกฤษ) Assoc. Prof. Ponrudee Netisopakul, Ph.D.
สถานที่ติดต่อ คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ไปรษณีย์อิเล็กทรอนิกส์ ponrudee@it.kmitl.ac.th

ประวัติการศึกษา

- Ph.D. (Computer Information and Science), Case Western Reserve University, Cleveland, OH, USA.
- M.S. (Computer Information Science), University of Delaware, Newark, DE, USA.
- M.S. (Computer Science), University of Southern California, Los Angeles, CA, USA.
- B.S. with Honor (Statistics), Chulalongkorn University, Bangkok, THAILAND.

ประสบการณ์งานวิจัยที่เกี่ยวข้อง

Kuptabuth, S., Netisopakul, P., "On Factors Affect Document Clustering: Comparison of Summary versus Full Documents", 6th International Joint Conference on Computer Science and Software Engineering, May 13-15, 2009, Phuket, Thailand.

Lertlilitrungroj, W., Netisopakul, P., "Simulation Modeling for Tollway Collection Decision Support System", 6th International Joint Conference on Computer Science and Software Engineering, May 13-15, 2009, Phuket, Thailand.

Netisopakul, P., Saapajit, W., "Pre-diagnosis Doctor Simulation using Case-Based Techniques", 2009 World Congress on Computer Science and Information Engineering, March 31-April 2, 2009, Los Angeles, USA.

Netisopakul, P., Lertvikul, S., "Development of Vendor Managed Inventory using Web Service", 2008 International Conference Global Research in Business and Economics, Sept 17-19, 2008, Orlando, USA.51

Netisopakul, P., "Software Engineering: Theory, Principles, and Practices (in Thai)", Translated from the original text book by Pressman, Top Publishing Co. Ltd., 2006

Netisopakul, P., Siriumpunkul, N., "Educational Service Web Database Prototype", Third International Conference on Intelligent Computing, August 21-24, 2007, China.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.

Netisopakul, P., "Web Metrics Support System (WMSS): Case Study at Faculty of Architecture, Chiang Mai University", Hawaii International Conference on Business, May 25-28, 2006, Honolulu, Hawaii.

Leenawong, C., Netisopakul, P., "Multiobjective Optimization Models for Production Planning at the Dairy Plant of Thailand's Royal Chitralada Projects", International Congress on Logistics and SCM System, April 30-May 6 2006, Kaohsiung, Taiwan.

Netisopakul, P., Leenawong, C., "Application of Nearest Neighbor Algorithm in E-Tourism Advisory System", International Technical Conference on Circuits/Systems, Computers and Communications (ITCCSCC),July 4-7, 2005, Korea.

Leenawong, C., Wattanasiripong, N., Netisopakul, P., "Interaction-based Algorithm for Replacing Components in the Multiple Complex System Model", International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC), July 4-7, 2005, Korea.

Leenawong, C. and Netisopakul, P. "Modeling and Optimization Approaches for Team Building Problem" Asia Pacific Industrial Engineering and Management Systems (APIEMS) 12-15 December 2004, USA. Pp. 34.14.1 - 34.14.12.

Netisopakul P. and Leelapat W. "User Adaptive Web Search Engine Architecture" Proceeding of International Conference on Computing, Communications and Control Technologies, August 2004, USA.Vol. VII pp. 92-97.

Punjataewakupt S. and Netisopakul P. "Knowledge Modules in Educational Software" Proceeding of International Conference on Computing, Communication and Control Technologies, August 2004, USA.Vol. VII pp. 103-106.52

Netisopakul P. "Visualizing Dynamic Objects in Object-Oriented Program" Proceedings of the 7th World Multiconference on Systemics, Cybernetics and Informatics, July 2003. USA. Vol. XIII pp. 321-325.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
This material is reserved for educational use only, not allowed for commercial use.
Forbidden to modify the content, and cite the document when use.