

การกรอกรำหยาบในข้อความภาษาไทยโดยไฟไนท์สเตทแมชชีน

วุฒิชัย วิเชียรไชย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

วิทยาศาสตร์มหาบัณฑิต (สถิติประยุกต์)

คณะสถิติประยุกต์

สถาบันบัณฑิตพัฒนบริหารศาสตร์

2549

การกรอกคำหยาบในข้อความภาษาไทยโดยไฟไนท์สเตทแมชชีน
วุฒิชัย วิเชียรไชย
คณะสถิติประยุกต์

คณะกรรมการสอบวิทยานิพนธ์ ได้พิจารณาแล้วเห็นสมควรอนุมัติให้เป็นส่วนหนึ่งของ
การศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต (สถิติประยุกต์)

ผู้ช่วยศาสตราจารย์ ประธานกรรมการ
(ดร. ระวีวรรณ เอื้อพันธ์วิริยะกุล)

รองศาสตราจารย์ กรรมการ
(ดร. วิจิต หล่อจิระชุนห์กุล)

ผู้ช่วยศาสตราจารย์ กรรมการ
(ดร. โอม ศรีนิล)

อาจารย์ คณบดี
(ดร. เลอสรณ์ โปสุวรรณ)

วันที่ เดือน พ.ศ.

บทคัดย่อ

ชื่อวิทยานิพนธ์	การกรองคำหยาบในข้อความภาษาไทยโดยไฟไนต์สเตตแมชชีน
ชื่อผู้เขียน	วุฒิชัย วิเชียรไชย
ชื่อปริญญา	วิทยาศาสตรมหาบัณฑิต (สถิติประยุกต์)
ปีการศึกษา	2549

งานวิจัยนี้ เป็นงานวิจัยที่เกี่ยวกับการพัฒนาโปรแกรมกรองคำหยาบในข้อความภาษาไทย โดยแบ่งขั้นตอนการทำงานของโปรแกรมออกเป็นสองขั้นตอนใหญ่ๆ คือ ขั้นตอนแรกเป็นการตัดคำ ขั้นตอนที่สองเป็นการค้นหาและแทนที่คำหยาบด้วยเครื่องหมายดอกจัน ซึ่งนิยามคำหยาบที่ใช้ นำมาจากพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2542

การตัดคำในข้อความภาษาไทย ในงานวิจัยนี้ใช้วิธีตัดคำด้วยพจนานุกรม แบบเทียบข้อความ (String Matching) และเทียบคำที่ยาวที่สุด (Longest Matching) และใช้ตัวแบบทางคณิตศาสตร์ไฟไนต์สเตตแมชชีนด้วยข้อมูลออก (Finite State Machines with Outputs) ในการพัฒนาโปรแกรมค้นหาและแทนที่คำหยาบในข้อความภาษาไทยด้วยเครื่องหมายดอกจัน

งานวิจัยนี้ได้ทดสอบประสิทธิภาพ โดยเอกสารจากหนังสือพิมพ์ไทยรัฐ มติชน ข่าวสด และคมชัดลึก จำนวนหนึ่งร้อยคอลัมน์ และข้อความจากเว็บบอร์ดจำนวนเจ็ดร้อยเจ็ดข้อคิดเห็นในแหล่งข้อมูลอินเทอร์เน็ต พบว่าประสิทธิภาพการแทนที่คำหยาบจริงด้วยเครื่องหมายดอกจันมีความถูกต้อง 99.14%

อย่างไรก็ตามยังมีคำหยาบที่ไม่สามารถแทนที่ด้วยเครื่องหมายดอกจัน เนื่องจากการตัดคำในข้อความภาษาไทยยังไม่ถูกต้อง 100% จึงได้เสนอแนวทางปรับปรุงโดยการเพิ่มคำศัพท์ที่ไม่ปรากฏในพจนานุกรมและเพิ่มข้อความพิเศษที่ไม่ปรากฏในพจนานุกรมข้อความพิเศษ เพื่อเพิ่มประสิทธิภาพในการตัดคำและกรองคำหยาบให้สูงขึ้น

ABSTRACT

Title of Thesis	Impolite Word Filtering in Thai Text by Finite State Machines
Author	Mr. Vuttichai Vicheanchai
Degree	Master of Science (Applied Statistics)
Year	2006

In this research, we present an implementation of impolite word filtering in Thai text program. The program mainly works two steps; word segmentation and searching and replacing impolite words by asterisk. The word segmentation step is completed by using string matching and longest matching method and gives the output as single word collection. Then the process of searching and replacing impolite words works by using finite state machines with outputs. The collection of impolite words used in the program refers to Royal Thai Dictionary of Buddhist Era 2542.

The testing data are taken from 100 articles of Thai newspaper; Thairath, Matichon, Khaosod and Komchadluek and 707 comments from Thai webboards. The accuracy of the program is 99.14%

However, there are some impolite words that are not properly detected and replaced owing to word segmentation phase that could not deliver 100% of correctness. The technique of word segmentation, therefore, can be potentially improved by supplementing some of undefined words and combination words that possibly exist in Thai text to the dictionary.

กิตติกรรมประกาศ

วิทยานิพนธ์เรื่องการรณรงค์หาขายในข้อความภาษาไทยโดยไฟไนท์สเตทแมชชีน สำเร็จได้เนื่องจากบุคคลหลายท่านได้กรุณาช่วยเหลือให้ข้อมูล ข้อเสนอแนะ คำปรึกษาแนะนำ ความคิดเห็นและกำลังใจ

ผู้เขียนขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร. ระวีวรรณ เอื้อพันธ์วิริยะกุล รองศาสตราจารย์ ดร. วิจิต หล่อจิระชุมหัทกุล และผู้ช่วยศาสตราจารย์ ดร. โอม ศรีนิล ที่ได้ให้คำชี้แนะแนวทางและตรวจสอบวิทยานิพนธ์ทุกขั้นตอน

ขอขอบพระคุณอาจารย์ทุกท่านของคณะสถิติประยุกต์ ที่ได้ถ่ายทอดความรู้ให้แก่ผู้ศึกษา และขอขอบคุณเจ้าหน้าที่คณะสถิติประยุกต์ ที่ได้ให้ความช่วยเหลือประสานงานการติดต่อเป็นอย่างดีด้วยอัธยาศัยไมตรีที่อบอุ่นเป็นกันเอง

ขอขอบคุณนักศึกษาร่วมรุ่นที่ได้ให้ความช่วยเหลือ และประสานงานตลอดช่วงเวลาที่ได้ศึกษาอยู่ที่คณะสถิติประยุกต์

ท้ายสุด ขอกราบขอบพระคุณบิดา มารดา และขอบคุณพี่น้องที่ได้ช่วยส่งเสริมสนับสนุน และเป็นกำลังใจตลอดมาให้แก่ผู้จัดทำวิทยานิพนธ์

วุฒิชัย วิเชียรไชย

เมษายน 2549

สารบัญ

	หน้า
บทคัดย่อ	(3)
ABSTRACT	(4)
กิตติกรรมประกาศ	(5)
สารบัญ	(6)
สารบัญตาราง	(8)
สารบัญภาพ	(10)
บทที่ 1 บทนำ	1
1.1 ความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของการวิจัย	2
1.3 ขอบเขตของการวิจัย	2
1.4 กรอบแนวความคิด	3
1.5 ประโยชน์ที่จะได้รับจากการวิจัย	3
บทที่ 2 การทบทวนวรรณกรรม	4
2.1 งานวิจัยที่เกี่ยวข้องกับการตัดคำ	4
2.2 สรุป	24
บทที่ 3 นิยามและทฤษฎีที่ใช้ในงานวิจัย	25
3.1 นิยาม	25
3.2 ทฤษฎีที่ใช้ในงานวิจัย	26
บทที่ 4 ขั้นตอนการวิจัย	36
4.1 ขั้นตอนการประมวลผล	36
4.2 ขั้นตอนการเตรียมเอกสาร	46
4.3 ขั้นตอนการทดสอบประสิทธิภาพ	46

บทที่ 5 ผลการทดสอบ สรุปผลการวิจัยและข้อเสนอแนะ	49
5.1 ข้อมูลที่ใช้ในการทดสอบ	49
5.2 ผลการทดสอบประสิทธิภาพ	49
5.3 สรุปผลการวิจัย	55
5.4 ข้อเสนอแนะ	56
บรรณานุกรม	58
ภาคผนวก	60
ภาคผนวก ก รหัสเลขฐานสองของตัวอักษรในภาษาไทยที่ใช้ในรหัสต้นฉบับ	61
ภาคผนวก ข คำหยาบที่มีความหมายตามนิยามในงานวิจัย	64
ภาคผนวก ค ตารางแสดงข้อความในเอกสารพิเศษที่ใช้ในการตัดคำโดยวิธีเทียบข้อความ	73
ภาคผนวก ง ตัวอย่างข้อความจากเว็บบอร์ดที่ใช้ทดสอบประสิทธิภาพของโปรแกรมในงานวิจัย	76
ภาคผนวก จ ผลการตัดคำและการแทนที่คำหยาบด้วยเครื่องหมายดอกจันของตัวอย่างข้อความจากเว็บบอร์ดที่ใช้ทดสอบประสิทธิภาพของโปรแกรมในงานวิจัย	80
ภาคผนวก ฉ ตัวอย่างข้อความจากข่าวหนังสือพิมพ์ที่ใช้ทดสอบประสิทธิภาพของโปรแกรมในงานวิจัย	88
ภาคผนวก ช ผลการตัดคำและการแทนที่คำหยาบด้วยเครื่องหมายดอกจันของตัวอย่างข้อความจากข่าวหนังสือพิมพ์ที่ใช้ทดสอบประสิทธิภาพของโปรแกรมในงานวิจัย	97
ประวัติผู้เขียน	118

สารบัญตาราง

ตารางที่	หน้า	
2.1	แสดงการตัดคำแบบเทียบคำที่ยาวที่สุด	10
2.2	แสดงการเปรียบเทียบการตัดคำ แบบคุณลักษณะของคำบริบทที่เรียนรู้ โดยระบบบริปเปอร์ แบบคุณลักษณะของคำบริบทที่เรียนรู้โดยโครงข่าย วินโนว์ แบบลักษณะสิ่งที่เกิดร่วมกัน โดยมีลำดับของคำที่เขียนต่อกัน ทั้งข้างหน้าและข้างหลังที่เรียนรู้โดยระบบบริปเปอร์ และแบบลักษณะสิ่ง ที่เกิดร่วมกัน โดยมีลำดับของคำที่เขียนต่อกันทั้ง ข้างหน้าและข้างหลังที่เรียนรู้ โดยโครงข่ายวินโนว์	22
2.3	แสดงการเปรียบเทียบการตัดคำกำกวมโดยระบบ FOIL และแบบจำลองไตรแกรม	24
3.1	แสดงตัวอย่างข้อความที่เกิดจากการประสมของคำศัพท์ในพจนานุกรม ซึ่งถ้าตัดคำแบบเทียบคำที่ยาวที่สุดแล้วจะทำให้ได้ผลลัพธ์ไม่ถูกต้อง และผลลัพธ์ที่จัดเก็บไว้เมื่อเทียบข้อความพบ	27
3.2	แสดงสถานะอธิบายไฟไนท์สเตตแมชชีนด้วยข้อมูลออกจากไดอะแกรมสถานะ ในภาพที่ 3.1	31
3.3	แสดงตารางคอนฟิซันเมตริก ในการทดสอบความถูกต้องของตัวแบบ	35
4.1	แสดงผลการตัดคำแบบเทียบคำที่ยาวที่สุดของข้อความ “เขาไปตลาด”	43
5.1	แสดงคอนฟิซันเมตริก ในการทดสอบประสิทธิภาพของโปรแกรมค้นหา และแทนที่คำหายในข้อความภาษาไทยด้วยเครื่องหมายดอกจัน	50
5.2	แสดงคอนฟิซันเมตริก ในการทดสอบประสิทธิภาพของโปรแกรมค้นหา และแทนที่คำหายในข้อความภาษาไทยด้วยเครื่องหมายดอกจันจาก ข้อคิดเห็นในเว็บบอร์ด	51
5.3	แสดงคอนฟิซันเมตริกในการทดสอบประสิทธิภาพของโปรแกรมค้นหา และแทนที่คำหายในข้อความภาษาไทยด้วยเครื่องหมายดอกจันจาก ข้อความข่าวหนังสือพิมพ์	52

5.4 แสดงคอนฟิซชั่นเมตริก ในการทดสอบประสิทธิภาพของโปรแกรมค้นหา และแทนที่คำหาบในข้อความภาษาไทยด้วยเครื่องหมายดอกจันจาก การรวมคอนฟิซชั่นเมตริกในตารางที่ 5.2 และตารางที่ 5.3	53
---	----

สารบัญภาพ

ภาพที่	หน้า
2.1 แสดงการค้นหาไอเดนติไฟเออร์ในคอมพิวเตอร์โดยทรานซิชันไดอะแกรม	7
2.2 แสดงการเก็บค่าในพจนานุกรมบนโครงสร้างข้อมูลแบบทรี	11
3.1 แสดงไดอะแกรมสถานะอธิบายไฟในท์สเตทแมชชีนด้วยข้อมูลออก	30
3.2 แสดงค่า Balance Factor ของแต่ละ โหนดในต้นไม้	32
3.3 แสดงต้นไม้เอวีแอล	33
3.4 แสดงการเพิ่มโหนดในต้นไม้เอวีแอลในภาพที่ 3.4 ทำให้ต้นไม้ไม่เป็นต้นไม้เอวีแอล	33
3.5 แสดงการปรับต้นไม้ให้เป็นต้นไม้เอวีแอล โดยการหมุนครั้งเดียวทางซ้าย	33
3.6 แสดงการปรับต้นไม้ให้เป็นต้นไม้เอวีแอลโดยวิธีการหมุนสองครั้ง โดยเริ่มจากภาพที่ 3.6 a) 3.6 b) 3.6 c) และ 3.6 d) ตามลำดับ	34
4.1 แสดงการทำงานของระบบการกรองคำหายบในข้อความภาษาไทย	37
4.2 แสดงการสร้างโครงสร้างของต้นไม้เอวีแอลจากคำว่า “กก” “กระบอก” และ “กระรอก”	39
4.3 แสดงการปรับต้นไม้ที่ไม่เป็นคุณสมบัติของต้นไม้เอวีแอลจากภาพที่ 4.2 ให้เป็นต้นไม้ เอวีแอลโดยการหมุนครั้งเดียวทางซ้าย	40
4.4 แสดงโหนดของต้นไม้เอวีแอลของข้อความเอกสารพิเศษ	41
4.5 แสดงการตัดคำโดยวิธีเทียบข้อความและเทียบคำที่ยาวที่สุด	44
4.6 แสดงไดอะแกรมสถานะในการอธิบายการค้นหาและแทนที่คำหายบ ในข้อความภาษาไทยด้วยเครื่องหมายดอกจัน	45

บทที่ 1

บทนำ

ในบทนี้จะกล่าวถึงความสำคัญของปัญหา วัตถุประสงค์ของการวิจัย ขอบเขตของการวิจัย กรอบแนวความคิดและประโยชน์ที่จะได้รับการวิจัย โดยมีรายละเอียดดังต่อไปนี้

1.1 ความสำคัญของปัญหา

ภาษาไทยเป็นเอกลักษณ์ของชาติไทย และมีวัฒนธรรมในการใช้ภาษาเป็นแบบแผนที่ดีงาม คนไทยจึงมีวัฒนธรรมในการใช้ถ้อยคำภาษาที่สุภาพ และเหมาะสมกับกาลเทศะ ฉะนั้นในการใช้ภาษาหยาบคายในสื่ออินเทอร์เน็ตจึงถือว่าขาดความเหมาะสมอย่างยิ่ง นอกจากนี้คำหยาบที่เกิดขึ้นบนเว็บไซต์ (Website) ขององค์กรต่างๆก่อให้เกิดผลเสียต่อภาพลักษณ์ขององค์กรทำให้ผู้ใช้บริการที่ได้อ่านเกิดความรู้สึกในทางลบต่อเว็บไซต์ขององค์กรนั้น รวมทั้งแสดงให้เห็นถึงการดำเนินงานที่ขาดประสิทธิภาพของเว็บมาสเตอร์ (Webmaster) ซึ่งไม่สามารถสกัดคำหยาบบนเว็บไซต์และอาจส่งผลให้เยาวชนที่อ่านคำหยาบเหล่านั้นจดจำนำไปใช้ในทางที่ผิด ทำให้เกิดผลเสียต่อตนเองและบุคคลอื่น ดังนั้นหากมีการใช้คำหยาบในข้อความภาษาไทยมากขึ้นจะทำให้ส่งผลเสียต่อวัฒนธรรมในการใช้ภาษาไทยเป็นอย่างยิ่ง

ด้วยเหตุนี้เพื่อเป็นการป้องกันปัญหาดังกล่าวไม่ให้เกิดขึ้น จึงจำเป็นที่จะต้องมีการตรวจสอบ แต่เนื่องจากข้อมูลข่าวสารที่ปรากฏในสื่ออินเทอร์เน็ตแต่ละวันนั้น มีจำนวนมาก ทำให้เป็นเรื่องที่ค่อนข้างลำบากในการที่จะให้มนุษย์เป็นผู้ที่ทำการตรวจสอบคำหยาบที่เกิดขึ้น เพราะมนุษย์มีข้อจำกัดในด้านการทำงานหลายอย่าง เช่น ไม่สามารถที่จะทำงานติดต่อกันเป็นเวลานานได้ แต่การตรวจสอบเป็นงานที่ต้องใช้ความรอบคอบและใช้เวลานาน ซึ่งหากเราสามารถสร้างเครื่องมือที่มีประสิทธิภาพช่วยในการตรวจสอบคำหยาบได้ จะทำให้สามารถลดภาระการทำงานของเว็บมาสเตอร์และแก้ไขปัญหาของคำหยาบได้

จากการศึกษาในปัจจุบัน ได้มีรูปแบบการสกัดคำที่ไม่ต้องการให้ปรากฏในเว็บไซต์สามรูปแบบใหญ่ๆ ดังนี้ รูปแบบที่หนึ่งเป็นการสกัดโดยผู้ดูแลระบบ รูปแบบที่สองเป็นการสกัดโดยให้ผู้ใช้โหวตกระทู้หรือข้อความที่ไม่ต้องการให้ปรากฏบนเว็บบอร์ด จากนั้นผู้ดูแลระบบจะทำการตรวจสอบข้อความที่มีการโหวตอีกครั้งหนึ่ง ซึ่งถ้าเห็นว่าไม่เหมาะสมจริงก็จะทำการลบหรือกระทำอย่างใดอย่างหนึ่งตามที่เห็นสมควร และรูปแบบที่สามเป็นการสกัดโดยให้เจ้าของเว็บบอร์ดป้อนคำสั่งฐานข้อมูล (Database) ในระบบเพื่อสกัดคำเหล่านั้นไม่ให้ปรากฏในเว็บบอร์ดของตนเอง ซึ่งวิธีการสกัดคำทั้งสามแบบนี้ยังไม่มีการจัดทำเป็นบทความหรืองานวิจัยขึ้นมา ดังนั้นวิทยานิพนธ์ฉบับนี้มีวัตถุประสงค์ที่จะสร้าง โปรแกรมค้นหาและแทนที่คำหยาบในข้อความภาษาไทยด้วยเครื่องหมายดอกจัน

1.2 วัตถุประสงค์ของการวิจัย

พัฒนาโปรแกรมค้นหาและแทนที่คำหยาบในข้อความภาษาไทยด้วยเครื่องหมายดอกจัน โดยต้องพัฒนาโปรแกรมตัดคำก่อน เพื่อนำคำที่ได้ไปประมวลผลในโปรแกรมค้นหาและแทนที่คำหยาบด้วยเครื่องหมายดอกจัน

1.3 ขอบเขตของการวิจัย

1. พัฒนาโปรแกรมโดยใช้ภาษาจาวา (Java) ซึ่งเป็นภาษาระดับสูง (High Level Language) ที่มีความยืดหยุ่นสูงบนระบบปฏิบัติการวินโดวส์เอ็กซ์พี (Windows XP)
2. กำหนดคำหยาบโดยนิยามจากพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2542
3. พัฒนาโปรแกรมตัดคำในข้อความภาษาไทย โดยใช้คำศัพท์จากพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ.2542 รวมทั้งชื่อจังหวัดในประเทศไทย ชื่อประเทศ ชื่อทวีปและชื่อดินแดน และกรองคำหยาบในข้อความภาษาไทย ซึ่งการกรองคำหยาบในงานวิจัยนี้จะทำการค้นหาและแทนที่คำหยาบในข้อความภาษาไทยด้วยเครื่องหมายดอกจัน โดยใช้ตัวแบบทางคณิตศาสตร์ไฟไนท์สเตทแมชชีนด้วยข้อมูลออก (Finite State Machines with Outputs) เป็นแนวทางในการพัฒนา

4. การทดสอบประสิทธิภาพ โปรแกรมค้นหาและแทนที่คำหายาในข้อความภาษาไทย ด้วยเครื่องหมายดอกจัน และสรุปผลการทดสอบโดยใช้เอกสารจากหนังสือพิมพ์และข้อความจาก เว็บบอร์ดในสื่ออินเทอร์เน็ตที่ไม่มีรูปภาพ บทกลอน หรือสมการใดๆ ทั้งสิ้น

1.4 กรอบแนวความคิด

1. กำหนดนิยามของคำหายาและคำปกติ
2. ศึกษาขั้นตอนวิธีการตัดคำในข้อความภาษาไทย
3. ออกแบบและพัฒนาโปรแกรมตัดคำในข้อความภาษาไทย
4. ศึกษาและรวบรวมคำหายาในพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2542 ที่มี ความหมายตามนิยามของคำหายาที่กำหนดไว้
5. ออกแบบ และพัฒนาโปรแกรมค้นหาและแทนที่คำหายาในข้อความภาษาไทยด้วย เครื่องหมายดอกจัน โดยอธิบายด้วยตัวแบบทางคณิตศาสตร์ไฟไนท์สเตทแมชชีนด้วยข้อมูลออก
6. ทดสอบประสิทธิภาพ โปรแกรมค้นหาและแทนที่คำหายาในข้อความภาษาไทยด้วย เครื่องหมายดอกจัน
7. สรุปผลการวิจัยและข้อเสนอแนะ

1.5 ประโยชน์ที่จะได้รับการวิจัย

1. สามารถนำโปรแกรมไปใช้กับเว็บไซต์ เพื่อไม่ให้ปรากฏคำหายาในข้อความภาษาไทย บนเว็บไซต์ที่เปิดโอกาสให้ผู้ใช้เขียนแสดงข้อคิดเห็นต่างๆ
 2. ลดคำหายาในสื่ออินเทอร์เน็ต เพื่อไม่ให้เยาวชนได้อ่านแล้วจดจำนำคำหายาไปพูด หรือเขียนให้เป็นผลเสียกับผู้อื่น
 3. เพิ่มความน่าเชื่อถือต่อเว็บไซต์ขององค์กรที่นำโปรแกรมไปใช้ ทำให้ผู้ใช้สังเกตเห็นถึงความเอาใจใส่และประสิทธิภาพการทำงานของเว็บมาสเตอร์ในการสกัดคำหายา
 4. ได้แนวทางการปรับปรุงขั้นตอนวิธีการกรองคำหายาในข้อความภาษาไทย
- จากรายละเอียดที่กล่าวมาแล้วข้างต้น ทำให้เห็นความสำคัญของปัญหาของคำหายาใน ภาษาไทยที่ปรากฏในเว็บไซต์ต่างๆ ในบทต่อไปจะกล่าวถึงการศึกษางานวิจัยที่เกี่ยวข้องในการ วิจัยครั้งนี้

บทที่ 2

การทบทวนวรรณกรรม

ในบทนี้จะกล่าวถึงการศึกษางานวิจัยที่เกี่ยวข้อง ซึ่งประกอบด้วยงานวิจัยในเรื่องของการตัดคำ โดยมีลำดับตามปีที่ทำการศึกษา โครงสร้างข้อมูลแบบต้นไม้เอวี่แอล และข้อสรุปในการพิจารณาเลือกใช้วิธีการ โครงสร้างข้อมูลเพื่อนำมาประยุกต์ใช้ในงานวิจัยฉบับนี้ ซึ่งมีรายละเอียดดังต่อไปนี้

2.1 งานวิจัยที่เกี่ยวข้องกับการตัดคำ

การเขียนประโยคภาษาไทยเป็นการเขียนติดต่อกัน โดยไม่มีเครื่องหมายวรรคตอนเป็นส่วนใหญ่และมีการแบ่งวรรคตอนด้วยช่องว่างบ้างเป็นบางครั้ง จึงทำให้มีความแตกต่างจากประโยคภาษาอังกฤษที่มีการแบ่งแยกคำแต่ละคำด้วยช่องว่าง และมีเครื่องหมายในการจบประโยคอย่างชัดเจน จึงทำให้การตัดคำในข้อความภาษาไทยมีความยุ่งยากกว่าภาษาอังกฤษ ลักษณะเช่นนี้ทำให้การประมวลผลภาษาไทยด้วยคอมพิวเตอร์มีความจำเป็นที่จะหาวิธีการให้คอมพิวเตอร์มีความสามารถที่จะเรียนรู้ขอบเขตของคำ เพื่อสามารถตัดคำได้ถูกต้องตามหลักภาษาศาสตร์ในการประมวลผลคำภาษาไทย เนื่องจากงานวิจัยฉบับนี้มีความจำเป็นต้องทำการตัดคำในภาษาไทยเพื่อนำคำที่ได้ไปประมวลผลในการค้นหาคำหาย โดยงานวิจัยเกี่ยวกับการตัดคำในภาษาไทยได้มีการพัฒนาติดต่อกันมาเป็นเวลานาน ทำให้มีงานวิจัยเกิดขึ้นมากมายหลายวิธี ซึ่งสามารถแบ่งงานวิจัยการตัดคำออกตามหลักวิธีที่ใช้ในการตัดคำได้เป็นดังนี้ หนึ่งวิธีการสร้างกฎ (Rule Base Approach) สอง วิธีการสร้างอัลกอริทึม (Algorithm Approach) สามวิธีการใช้พจนานุกรม (Dictionary Base Approach) และสี่วิธีการใช้คลังข้อความ (Corpus Base Approach)

2.1.1 วิธีการสร้างกฎ

การตัดคำโดยวิธีการสร้างกฎมีงานวิจัยที่ได้ทำการศึกษาอยู่สองชิ้นคือ

2.1.1.1 งานวิจัยของ Yupin Thairatananond (1981) เป็นงานวิจัยการตัดพยางค์โดยการใช้กฎที่สร้างขึ้นจากหลักไวยากรณ์ภาษาไทย และมีการจัดเก็บพยางค์ต่างๆที่เป็นช้อยกเว้นไว้ในแฟ้มข้อมูล เนื่องจากมีบางพยางค์ไม่เป็นไปตามกฎที่สร้างไว้ ลักษณะของกฎมีการพิจารณาจากลักษณะของอักษรที่ปรากฏในพยางค์หรือคำ ซึ่งทำให้สามารถจัดและแบ่งหมวดหมู่ตัวอักษรภาษาไทยได้เป็นห้ากลุ่มใหญ่ๆ ดังต่อไปนี้คือ

- 1) กลุ่มพยัญชนะ (Consonant)
 - (1) พยัญชนะที่อยู่ข้างหน้าพยางค์เสมอ
 - (2) พยัญชนะที่ส่วนใหญ่จะอยู่หน้าพยางค์
 - (3) พยัญชนะที่เป็นตัวสะกด
 - (4) พยัญชนะที่เป็นสระ
 - (5) อื่นๆ
- 2) กลุ่มสระ (Vowel)
 - (1) สระที่ไม่ต้องมีตัวสะกด
 - (2) สระที่อยู่หน้าพยางค์เสมอ
 - (3) สระที่ส่วนใหญ่จะมีตัวสะกดร่วมด้วย
 - (4) สระที่มีหรือไม่มีตัวสะกดร่วมด้วย
- 3) กลุ่มวรรณยุกต์ (Tone mark)
- 4) กลุ่มตัวเลข (Numeral)
- 5) กลุ่มอักขระพิเศษ (Special character)

วิธีการนี้มีขั้นตอนการทำงานตัดพยางค์จากขวามาซ้าย โดยใช้กฎต่างๆที่สร้างขึ้นมาจากลักษณะของตัวอักษรที่ได้กล่าวไปแล้ว และกฎต่างๆที่สร้างขึ้นจะจัดเก็บไว้ในรหัสต้นฉบับ (Source Code) ของโปรแกรมซึ่งทำให้การเพิ่มหรือแก้ไขกฎไม่สามารถทำได้สะดวก จากการทดสอบปรากฏว่าผลลัพธ์ที่ได้จากการตัดพยางค์ด้วยวิธีการนี้ ได้ผลความถูกต้องไม่น้อยกว่า 85%

2.1.1.2 งานวิจัยของ Surin Charnyapornpong (1983) เป็นงานวิจัยเกี่ยวกับการตัดพยางค์ภาษาไทยโดยใช้กฎที่สร้างขึ้นจากหลักไวยากรณ์ของภาษาไทย และได้ทำการวิเคราะห์ลักษณะต่างๆ ของพยางค์ภาษาไทย ซึ่งจะต่างจากกฎที่แยกเป็นหมวดพยัญชนะ สระ วรรณยุกต์ กลุ่มตัวเลข และกลุ่มอักขระพิเศษอย่างใดอย่างหนึ่งโดยอิสระ (Yupin Thairatananond, 1981) แต่

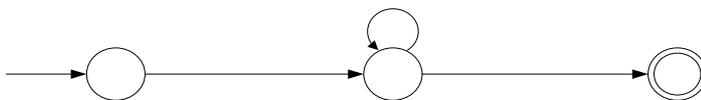
2.1.2 วิธีการสร้างอัลกอริทึม

การตัดคำโดยวิธีการสร้างอัลกอริทึมมีงานวิจัยที่ได้ทำการศึกษาอยู่หนึ่งชิ้นคือ

งานวิจัยของวิจิต หล่อจิระชุนท์กุล และเจริญ คุณินทร์พันธุ์ (2524) เป็นงานวิจัยในการสร้างอัลกอริทึมการแยกอักขระไทยที่เรียงติดกันออกเป็นโทเคน (Token) โดยโทเคนคือ กลุ่มอักขระที่มีจำนวนพยางค์น้อยที่สุด เมื่อเขียนแยกกันอยู่คนละบรรทัดแล้ว ยังคงอยู่ในรูปที่อ่านได้ เช่น คำว่า “จักขุ” สามารถแยกออกเป็นสองโทเคนคือ จัก-ขุ คำว่า “ประจักษ์” สามารถแบ่งออกเป็นสองโทเคนคือ ประ-จักษ์ เป็นต้น

จากการศึกษาภาษาไทยนั้นมีอักขระทั้งหมดเจ็ดสิบสองตัว ซึ่งมีโครงสร้างทั้งหมดสี่ระดับ ดังนี้ ระดับที่หนึ่ง เป็นระดับของอักขระประเภทวรรณยุกต์ทั้งหมดสี่ตัวคือ เอก โท ตรี จัตวา และการันต์ ระดับที่สอง เป็นระดับที่อยู่ต่ำกว่าระดับที่หนึ่ง อักขระที่อาจจะอยู่ในนี้ มีทั้งหมด เจ็ดตัว ซึ่งเป็นสระทั้งสี่คือ ไม้หันอากาศ สระอิ สระอี สระอือ ไม้ไต่คู้ และหยาดน้ำค้าง ระดับที่สาม เป็นระดับที่อยู่ต่ำกว่าระดับที่สอง และเป็นระดับหลักของภาษาไทย พยัญชนะทั้งหมดสี่สิบสี่ตัว จะอยู่ในระดับนี้ รวมทั้ง ฤ ฦ สระอะ สระอา สระอำ สระเอ สระแอ สระโอ สระออ สระโอ ๆ และ ๆ และระดับที่สี่ เป็นระดับต่ำสุด อักขระในระดับนี้มีสามตัวเป็นสระทั้งสี่คือ สระอุ อุ และพินทุ

การพัฒนาอัลกอริทึมแยกอักขระภาษาไทยออกเป็นโทเคน ไม่ได้อาศัยหลักไวยากรณ์ไทยมาใช้ในการตัดคำ แต่ใช้การสังเกตลักษณะของคำและพยางค์ในภาษาไทยเป็นสำคัญ ซึ่งจะสร้างเป็นพรีซิเดนซ์เมตริก (Precedence Matrix) ขึ้นมาเพื่อให้อัลกอริทึมมีขนาดเล็กแต่จะมีความซับซ้อนมาก แนวทางที่ใช้ในการพัฒนาอัลกอริทึมเป็นลักษณะของการเขียนเลกซิคอล โพรเซสเซอร์ (Lexical Processor) เพื่อแยกอักขระไทยซึ่งเรียงติดกัน (Linear String) ออกเป็นโทเคนในเลกซิคอลโพรเซสเซอร์ ดังนั้นจึงจำเป็นต้องหาเดลิมิเตอร์ (Delimiter) เพื่อค้นหาไอนเดนติไฟเออร์ (Identifier) ซึ่งอาจเป็นตัวอักษรหรือตัวเลขหรืออักขระพิเศษ (Special Character) เช่น ; = . เป็นต้น ในทำนองเดียวกันการค้นหาไอนเดนติไฟเออร์ในคอมไพเลอร์ (Compiler) สามารถสรุปได้โดยทรานซิชันไดอะแกรม (Transition Diagram) ง่าย ๆ ดังภาพต่อไปนี้



ภาพที่ 2.1 แสดงการค้นหาไอนเดนติไฟเออร์ในคอมไพเลอร์โดยทรานซิชันไดอะแกรม

ทรานซิชันไดอะแกรมหรือไดอะแกรมสถานะ (State Diagram) ในทฤษฎีนิรนัยอัตโนมัติ
 สเตทอโตมาตา (Deterministic Finite State Automata) ซึ่งเป็นตัวแบบทางคณิตศาสตร์ไฟไนท์
 สเตทแมชชีนอีกแบบหนึ่งซึ่งมีลักษณะแน่นอน โดยแต่ละคู่ระหว่างสถานะกับข้อมูลเข้าจะเปลี่ยน
 สถานะไปยังสถานะต่อไปได้เพียงสถานะเดียวเท่านั้น

ดังนั้นทรานซิชันไดอะแกรมของงานวิจัยนี้ จะแบ่งสถานะทั้งหมดออกเป็นสามสถานะคือ
 0 1 และ 2 ข้อมูลเข้าคือ อักขระ ฟังก์ชันการเปลี่ยนแปลงสถานะคือ กฎที่สร้างขึ้นเป็นลักษณะของ
 การเขียนเลขซิกอลโพรเซสเซอร์ สถานะเริ่มต้นคือ 0 และสถานะสุดท้ายคือ 2 โดยสามารถอธิบาย
 ได้ดังนี้ อักขระตัวเดียวอาจเป็นเดลิมีเตอร์ในสภาวะหนึ่ง แต่ไม่เป็นเดลิมีเตอร์ในอีกสภาวะหนึ่งก็ได้
 เช่น การป้อนข้อมูลเป็น "...ห..." จะได้ว่า "ห" เป็น เดลิมีเตอร์ เพราะมี "ง" เป็นอักขระนำหน้า
 แต่ถ้าข้อมูลเป็น "...เห..." "ห" จะไม่เป็นเดลิมีเตอร์แล้ว เพราะมี "เ" เป็นอักขระนำหน้า เป็นต้น
 ดังนั้นการแยกอักขระไทยออกเป็นโทเคนจึงมีความสลับซับซ้อนมาก ด้วยเหตุนี้จึงจำเป็นต้อง
 จะต้องการทราบความเป็นเดลิมีเตอร์ของอักขระภายใต้สภาพต่างๆ ในที่นี้จะพิจารณาจากอักขระ
 ตาม k ตำแหน่ง ซึ่ง $k = 1, 2$ หรือ 3 ทั้งนี้เพราะความยากลำบากของการกำหนดความเป็นเดลิ
 มีเตอร์ ตัวอย่างการกำหนดความเป็นเดลิมีเตอร์ เช่น ถ้า "ค" เป็นอักขระตามของ "ก" "ค" จะยัง
 ไม่เป็นเดลิมีเตอร์ แต่ถ้ามีไม้หันอากาศเป็นอักขระตามในระดับที่สองของพยัญชนะ "ก" ด้วย
 "ค" จะกลายเป็นเดลิมีเตอร์ได้แก่ "กค" เป็นต้น อีกประการหนึ่งหากอักขระตามในระดับที่สองเป็น
 ไม้หันอากาศตามหลังอักขระในระดับที่สามที่ไม่ใช่กลุ่มอักขระในเซต {กคตททภมยรลวษสท}
 แล้วอักขระในระดับที่สามที่มีไม้หันอากาศตามนั้นจะเป็นเดลิมีเตอร์ เช่น "...ประจักษ์..." ซึ่ง "จ"
 ไม่ได้อยู่ในกลุ่มอักขระของเซตดังกล่าวและมีไม้หันอากาศตามหลัง "จ" จึงเป็นเดลิมีเตอร์ เป็นต้น
 แต่อักขระที่อยู่ในกลุ่มอักขระของเซตดังกล่าว ก็อาจจะเป็นเดลิมีเตอร์ได้ เช่น ถ้าอักขระตำแหน่ง
 ปัจจุบันเป็น "ก" และอักขระตามเป็นไม้หันอากาศ ยกตัวอย่าง "...ปลากัด..." เป็นต้น จากการศึกษา
 งานวิจัยนี้สรุปได้ว่า ประสิทธิภาพในการแยกอักขระไทยที่เรียงติดกันออกเป็นโทเคนอยู่ในร้อยละ
 80 อย่างไรก็ตามประสิทธิภาพอาจปรับปรุงให้สูงขึ้นได้โดยการสร้างตารางของคำที่พบบ่อย ซึ่งไม่
 สามารถแยกออกได้โดยอัลกอริทึมนี้ เช่น คำว่า "เป็น" อัลกอริทึมไม่สามารถแยกคำนี้ออกเป็นโท
 เคน ได้ ทั้งนี้จากกฎที่ว่า "น" เมื่ออยู่ติดกับกลุ่มอักขระ ในเซต {กขกงจชดตทนบปภมยรลวษ
 หพอ} จะทำให้ไม่อาจแยกคำว่า "เป็น" ออกมาเป็นโทเคน ได้ เช่น ข้อความ "เป็นการ" "เป็นงาน"
 "เป็นนาน" "เป็นบาป" ฯลฯ เหล่านี้ซึ่งมีสองโทเคน แต่อัลกอริทึมไม่อาจแยกออกมาได้ หากมี
 ตารางดังกล่าวซึ่งมีคำ "เป็น" อยู่ ข้อความเหล่านี้ก็จะถูกแยกออกมาได้เป็นสองโทเคน ซึ่งเป็นการ
 เพิ่มประสิทธิภาพของอัลกอริทึมอย่างมีนัยสำคัญ ทั้งนี้เพราะเป็นคำที่พบบ่อยในภาษาไทย
 ด้วยเหตุนี้คำในตารางดังกล่าวควรเป็นคำที่พบบ่อยและไม่มีข้อยกเว้นใดๆ เกี่ยวกับความเป็น

เดลิมีเตอร์ของอักขระที่ติดตามมาเพื่อความสะดวกของอัลกอริทึม

2.1.3 วิธีการใช้พจนานุกรม

การตัดคำโดยวิธีการใช้พจนานุกรมมีงานวิจัยที่ได้ทำการศึกษายู่สี่ชิ้นคือ

2.1.3.1 งานวิจัยของยีน กูว์รเวรธ และวิวรรธน์ อิมอาร์มณ (2529) เป็นงานวิจัยการแบ่งพยางค์ด้วยพจนานุกรม ซึ่งถือได้ว่าเป็นงานวิจัยแรกของการตัดคำที่มีการนำพจนานุกรมมาใช้โดยจะจัดเก็บพยางค์ต่างๆ ไว้ในพจนานุกรม และมีการนำกฎไวยากรณ์เข้ามาช่วยในกรณีที่ไม่มีพจนานุกรม หลักการทำงานของวิธีการตัดพยางค์ด้วยพจนานุกรมนี้ ก็คือจะทำการตรวจสอบสายอักขระ (String) ที่เข้ามาจากซ้ายไปขวากับพยางค์ที่ได้เก็บไว้ในพจนานุกรม ในกรณีที่ทำการตรวจสอบแล้วปรากฏว่าพบคำที่มีพยางค์มากกว่าหนึ่งพยางค์ในพจนานุกรม ก็จะทำการเลือกแบ่งพยางค์โดยเลือกพยางค์ที่ยาวที่สุด แล้วทำเครื่องหมายจุดย้อนกลับกับพยางค์ที่เหลือ แล้วทำต่อไปเรื่อยๆจนสิ้นสุดสายอักขระ แต่ถ้าในกรณีที่เลือกพยางค์ที่ยาวที่สุดแล้ว ทำให้เกิดคำที่ไม่พบในพจนานุกรมก็จะทำการย้อนรอย (Backtracking) กลับไปยังจุดย้อนกลับล่าสุดแล้วเลือกพยางค์ที่มีความยาวรองลงมาแทน ซึ่งวิธีการนี้จะเรียกว่าการตัดคำแบบเทียบคำที่ยาวที่สุด (Longest Matching)

งานวิจัยนี้ใช้กับไมโครคอมพิวเตอร์ ที่ทำงานภายใต้ระบบปฏิบัติการเอ็มเอสดอสและจัดการพจนานุกรมแบบกำหนดระเบียบคงที่ ทำให้การตรวจสอบจะกระทำได้ทันทีอย่างรวดเร็ว แม้ในขณะกดคีย์ป้อนข้อมูล ความเร็วของการตรวจสอบจะช้าลงบ้างบางจังหวะเมื่อต้องเข้าไปตรวจสอบกับกฎ เพราะหากคำในพจนานุกรมไม่พบ แต่โอกาสนี้จะเกิดขึ้นน้อยกว่า 5% สำหรับเอกสารที่ใช้ในชีวิตประจำวัน ความถูกต้องของการแบ่งแยกพยางค์ด้วยวิธีนี้ สามารถทำถูกต้องได้สูงกว่า 99% แต่อย่างไรก็ตามจากการทดลองพบว่า ถ้าให้เอกสารเป็นเอกสารที่มีคำศัพท์วิชาการมากขึ้น ความถูกต้องจะลดลง เปอร์เซนต์ที่ได้ด้วยวิธีนี้เป็นที่น่าพอใจมาก เมื่อเทียบกับวิธีใช้กฎเกณฑ์แต่เพียงอย่างเดียว

ข้อเสียของวิธีนี้ที่เห็นได้ชัดคือ ต้องเสียพื้นที่ในหน่วยความจำสำหรับการเก็บพจนานุกรม แต่ปัญหานี้แก้ไขได้โดยง่าย เพราะไมโครคอมพิวเตอร์ในปัจจุบันใช้หน่วยความจำขนาดใหญ่อยู่แล้ว และการเลือกพจนานุกรมในงานวิจัยนี้มีความสำคัญ โดยพจนานุกรมที่ใช้ที่นี่ นับว่าเป็นวิธีการคัดเลือกคำที่มีโอกาสใช้มากมาไว้ในพจนานุกรม ซึ่งเป็นวิธีการที่ได้ประสิทธิภาพในการทำงานของระบบ แต่กรรมวิธีการเลือกพจนานุกรมเป็นเรื่องที่ต้องเสียเวลามาก พจนานุกรมที่ใช้เป็นการจัดเก็บคำหรือพยางค์ต่างๆ ที่ใช้ในชีวิตประจำวันจากเอกสารและสิ่งตีพิมพ์ต่างๆ ทั้งสิ้นประมาณหนึ่งร้อยสามสิบหกรายการ โดยแยกประเภทของเอกสารให้คลุมกับตัวอย่างของคำที่ใช้ใน

ชีวิตประจำวันมากที่สุด ข้อมูลที่ได้จะทำการจัดเก็บและแจกแจงความถี่ มีค่าประมาณห้าพันสี่ร้อย คำ (ถ้ามีคำซ้ำให้เก็บเพียงหนึ่งคำ) จากเอกสารและสิ่งตีพิมพ์ข้างต้นซึ่งมีค่าทั้งสิ้นประมาณหนึ่งแสนสี่หมื่นคำ ซึ่งหากพิจารณาถึงโครงสร้างของพจนานุกรมนี้แล้ว จะมีขนาดเพียงประมาณห้าสิบลบิกิโลไบต์ (ไม่รวมค่าความถี่)

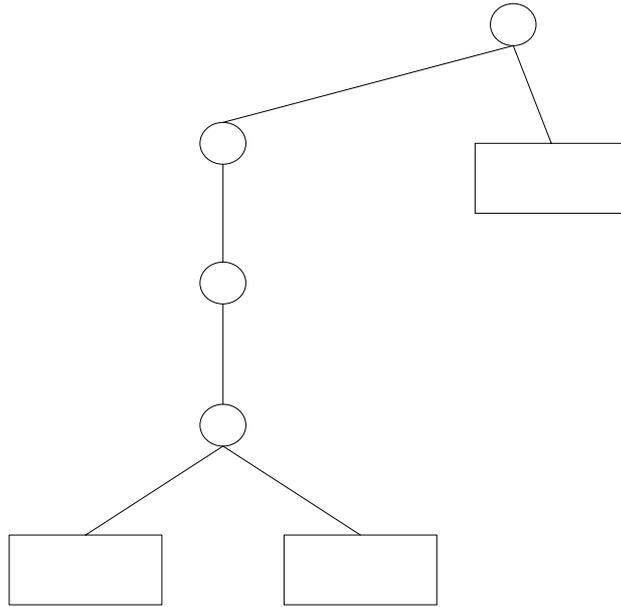
2.1.3.2 งานวิจัยของสัมพันธ ะธีรนรมย์ (2534) เป็นงานวิจัยการแบ่งคำไทยด้วยพจนานุกรม โดยมีเป้าหมายเพื่อที่จะเพิ่มประสิทธิภาพในด้านความเร็วของขั้นตอนวิธีในการตัดคำและการลดขนาดของพจนานุกรม เนื่องจากเมื่อนำพจนานุกรมเข้ามาใช้ในการตัดคำแล้วจะทำให้ความถูกต้องในการตัดคำเพิ่มมากกว่าการตัดคำโดยใช้กฎอย่างเดียว ดังนั้นงานวิจัยนี้จึงไม่ได้เน้นการเพิ่มประสิทธิภาพในด้านความถูกต้องมากนัก เพราะถือว่าการตัดคำโดยใช้พจนานุกรมให้ค่าความถูกต้องที่สูงอยู่แล้ว ขั้นตอนวิธีการตัดคำในงานวิจัยนี้จะมีลักษณะคล้ายกับงานวิจัยการแบ่งพยางค์โดยใช้พจนานุกรม (ยีน ภู่วรรณ และวิวรรธน์ อิมอรณณ์, 2529) คือใช้ขั้นตอนวิธีแบบเทียบคำที่ยาวที่สุดที่กล่าวไปแล้ว แต่ในงานวิจัยนี้จะทำการเก็บคำแทนพยางค์ในพจนานุกรม ตัวอย่างการตัดคำโดยเทียบคำที่ยาวที่สุด แสดงดังตารางที่ 2.1 ดังนี้

ตารางที่ 2.1 แสดงการตัดคำแบบเทียบคำที่ยาวที่สุด

ประโยค	คำที่เป็นไปได้	คำที่ถูกเลือก
เขาหาคะดาศ	เขา	เขา
หาคะดาศ	หา หาก	หาก (เลือกคำที่ยาวที่สุด)
ระดาศ	- (ไม่มีในพจนานุกรม)	(ย้อนรอย)
หาคะดาศ	หา หาก	หา (เลือกคำที่ยาวรองลงมา)
กะดาศ	กะดาศ	กะดาศ

จากตารางที่ 2.1 จะแสดงการตัดคำแบบเทียบคำที่ยาวที่สุด โดยประโยคที่นำมาตัดคำคือ “เขาหาคะดาศ” สามารถตัดคำได้เป็น “เขา หา กะดาศ” ส่วนโครงสร้างของพจนานุกรมที่นำมาใช้ในงานวิจัยนี้คือ โครงสร้างข้อมูลแบบทรี (Trie) โดยทรีเป็นดิจิทัลทรี (Digital Tree) ชนิดหนึ่ง สามารถมีลูกได้มากกว่าสองตามจำนวนอักขระทั้งหมดที่ใช้ในการจัดเก็บ โดยลูกจะอยู่ระดับถัดลงไปและจะเรียงตามลำดับจากซ้ายไปขวาของคำในอักขระ ซึ่งจากการนำโครงสร้างข้อมูลแบบทรีเข้ามาใช้สามารถช่วยลดขนาดของพจนานุกรมได้ และนอกจากนี้โครงสร้างแบบทรียังสามารถสืบค้นหาคำศัพท์ได้รวดเร็วโดยใช้เวลาเท่ากับ $O(n)$ โดยที่ n เป็นความยาวของคำที่

ต้องการค้นหา และทริยังสนับสนุนการค้นหาอักขระเต็มหน้า (Prefix) ด้วย และสามารถเพิ่มหรือแก้ไขคำศัพท์ได้อย่างสะดวกและรวดเร็ว ตัวอย่างโครงสร้างข้อมูลแบบทริยังแสดงในภาพที่ 2.2 ถ้าคำในพจนานุกรมมีเพียงสามคำคือ กระจอก กระจอก และหน้าไม้



ภาพที่ 2.2 แสดงการเก็บคำในพจนานุกรมบนโครงสร้างข้อมูลแบบทริยัง

สรุปจากงานวิจัยนี้ได้มีการนำโครงสร้างทริยังมาประยุกต์ใช้ เพื่อลดขนาดของพจนานุกรม โดยผลการเปรียบเทียบขนาดของพจนานุกรมจำนวนห้าพันสี่ร้อยคำ สามารถใช้เนื้อที่สองหมื่นเจ็ดพันเก้าร้อยเจ็ดสิบห้าไบต์ ซึ่งมีขนาดน้อยกว่างานวิจัยการแบ่งพยางค์ด้วยพจนานุกรม (ยืน กู่วรรณ และวิวรรณ อิมอรณ, 2529) ซึ่งใช้เนื้อที่ประมาณสามหมื่นสองพันสี่ร้อยแปดสิบสองไบต์ อีกทั้งในส่วนด้านความซับซ้อนของขั้นตอนวิธีการสืบค้นก็ลดลงด้วยเนื่องมาจากลักษณะทางโครงสร้างของทริยัง

2.1.3.3 งานวิจัยของวิรัช ศรีเลิศวานิช (2535) ได้มีการพัฒนาการตัดคำที่เรียกว่า แมกซิ มอลแมทชิง (Maximal Matching) ซึ่งขั้นตอนวิธีนี้จะสามารถแก้ไขความบกพร่องของการตัดคำแบบเทียบคำที่ยาวที่สุดได้ โดยจุดบกพร่องที่กล่าวนี้คือวิธีการตัดคำแบบเทียบคำที่ยาวที่สุดจะเลือก คำที่ยาวเกินไปตั้งแต่ครั้งแรก ทำให้ข้อความที่ตามมาเกิดข้อผิดพลาดได้ ตัวอย่างเช่นประโยค “ยาว หลาก ว่าๆ” จะตัดคำได้เป็น “ยาว หลาก ว่าๆ” โดยที่ถูกต้องควรจะเป็น “ยาว หลาก ว่าๆ” หลัก

ของการตัดคำแบบแม็กซิมอลเม็ทซึ่งมีขั้นตอนดังนี้คือ ทำการตัดคำที่เป็นไปได้ทุกๆแบบก่อน หลังจากนั้นก็ให้เลือกประโยคที่มีจำนวนคำน้อยที่สุด ตัวอย่างเช่นประโยค “ไปหามเหสี” สามารถตัดคำได้เป็น “ไป หาม เห สี” และ “ไป หา มเหสี” ซึ่งมีจำนวนคำสี่คำและสามคำตามลำดับ แต่เมื่อพิจารณาจำนวนคำแล้วก็จะเลือกประโยค “ไป หา มเหสี” เพราะเป็นประโยคที่มีจำนวนคำน้อยที่สุดจากประโยคที่สามารถตัดคำที่เป็นไปได้ทั้งหมด แต่ถ้าเกิดกรณีที่มีประโยคที่สามารถตัดคำที่เป็นไปได้มีจำนวนคำน้อยที่สุดเท่ากันมากกว่าหนึ่งแบบ ก็จะเลือกตัดคำโดยวิธีเทียบคำที่ยาวที่สุดแทน ตัวอย่างเช่นประโยค “ปลา นอน ตาก ลม” สามารถตัดคำได้สองแบบคือ “ปลา นอน ตาก ลม” กับ “ปลา นอน ตา ลม” ซึ่งมีจำนวนคำที่ตัดได้เท่ากันทั้งสองแบบ ดังนั้นจึงใช้วิธีการตัดคำแบบเทียบคำที่ยาวที่สุด ทำให้ได้ผลลัพธ์เป็น “ปลา นอน ตาก ลม” สรุปได้ว่าวิธีการนี้สามารถช่วยแก้ปัญหาของการตัดคำแบบเทียบคำที่ยาวที่สุดได้ เพราะการตัดคำแบบเทียบคำที่ยาวที่สุดจะเลือกคำที่ยาวที่สุดเมื่อเจอข้อความที่กำลังจะพิจารณา โดยไม่มีการพิจารณาถึงข้อความถัดไป ซึ่งมีลักษณะเหมือนการใช้กริดดี้อัลกอริทึม (Greedy Algorithm) ที่พิจารณาเฉพาะบริเวณที่ใกล้เท่านั้น แต่วิธีการตัดคำให้แม็กซิมอลเม็ทซึ่งจะเป็นวิธีการตัดคำโดยพิจารณาข้อความทั้งหมดแทน แต่อย่างไรก็ตามเนื่องจากวิธีการนี้ใช้เฉพาะพจนานุกรมในการตัดคำเท่านั้น จำเป็นต้องมีการนำโครงสร้างทางไวยากรณ์หรือความสัมพันธ์ทางความหมายเข้ามาใช้ในการพิจารณาด้วย

2.1.3.4 งานวิจัยของสมนึก สินธุปวน (2546) จุดประสงค์ของงานวิจัยนี้คือ การตัดคำโดยวิธีการวิเคราะห์การกระจายคำที่ใช้หลักไวยากรณ์ของภาษาไทยและพจนานุกรม ด้วยโปรแกรมเชิงเจเนติก หลักการโปรแกรมเชิงเจเนติก มีพื้นฐานการเลียนแบบกฎของธรรมชาติในการสืบทอดเผ่าพันธุ์ของสิ่งมีชีวิต ในหลักการของการสืบทอดนี้ลูกที่กำเนิดจะได้รับการถ่ายทอดคุณลักษณะจากพ่อและแม่ด้วย ทำให้การผสมพันธุ์เช่นนี้มีต่อเนื่องจากรุ่นหนึ่งไปยังอีกรุ่นหนึ่ง ซึ่งจะได้ประชากรใหม่ที่มีคุณลักษณะที่ดีขึ้นเรื่อยๆ นอกจากนี้หลักการโปรแกรมเชิงเจเนติก ยังสามารถนำมาใช้สร้างต้นไม้ให้ได้ค่าความเหมาะสมที่ดีที่สุด โดยค่าความเหมาะสมดังกล่าววัดได้จากค่าผลรวมความผิดพลาดของผลลัพธ์ที่ได้จากโปรแกรม โดยตัวโปรแกรมจะถูกพัฒนาโดยการข้ามพันธุ์ (Crossover) การกลายพันธุ์ (Mutation) และการคัดลอกพันธุ์ (Reproduction) ดังนั้นในงานวิจัยนี้นำโปรแกรมเชิงเจเนติกมาใช้ เพื่อช่วยตรวจสอบความถูกต้องทางด้านไวยากรณ์ในการตัดคำ โดยการหาต้นไม้การวิเคราะห์การกระจายคำของประโยคภาษาไทยที่ให้ค่าความเหมาะสมสูงที่สุดในแง่การตัดคำที่ถูกต้องและสอดคล้องกับไวยากรณ์มากที่สุด ในที่นี้ต้นไม้การวิเคราะห์การกระจายคำของประโยคภาษาไทยก็คือ ประชากรซึ่งประกอบไปด้วย โหนดฟังก์ชันและโหนดเทอร์มินัล โดยโหนดฟังก์ชันนั้น ได้แก่กฎในไวยากรณ์ภาษาไทยซึ่งมีทั้งหมดหนึ่งร้อยห้ากฎ ส่วนโหนดเทอร์มินัลนั้น ได้แก่ชนิดของคำตามไวยากรณ์ภาษาไทยซึ่งมีทั้งหมดสามสิบเอ็ดชนิด

การข้ามพันธุ์จะกระทำระหว่างต้นไม้การวิเคราะห์กระจายค่า จากต้นไม้สองต้นที่ถูกเลือก (Selection Criteria) ให้เป็นต้นไม้ผู้ปกครองเพื่อทำการข้ามพันธุ์ให้ได้ต้นไม้ลูกใหม่สองต้น โดยที่จุดข้ามพันธุ์จะต้องเป็นโหนดจากต้นไม้ที่ถูกเลือก ซึ่งมีเงื่อนไขสองข้อคือ

ข้อหนึ่ง ถ้าจุดข้ามพันธุ์ที่สุ่มได้เป็นโหนดฟังก์ชัน จุดข้ามพันธุ์ทั้งสองนั้นต้องเป็นโหนดฟังก์ชันเดียวกันและมีสายอักขระเหมือนกัน โดยจะต้องเริ่มและจบให้เหมือนกันด้วย มิฉะนั้นการแลกเปลี่ยนต้นไม้ย่อยทั้งสองจะกระทำไม่ได้ เนื่องจากต้นไม้ผลลัพธ์จะมีโครงสร้างที่ไม่ถูกต้อง

ข้อสอง ถ้าจุดข้ามพันธุ์เป็นโหนดที่เป็นเทอร์มินัล จะต้องมิชนิคของค่าที่เหมือนกันและสายอักขระเหมือนกัน มิฉะนั้นจะทำการแลกเปลี่ยนกันไม่ได้

ขั้นตอนการข้ามพันธุ์คือ ทำการสุ่มเลือกโหนดของต้นไม้ผู้ปกครองสองต้นแล้วตรวจสอบเงื่อนไข ถ้าเป็นจริงจึงทำการข้ามพันธุ์ ถ้าไม่เป็นจริงจะไม่ทำการข้ามพันธุ์แต่จะทำการคัดลอกผู้ปกครองทั้งสองไปยังประชากรรุ่นใหม่ มิฉะนั้นการแลกเปลี่ยนต้นไม้ทั้งสองจะกระทำไม่ได้

การกลายพันธุ์สามารถทำได้ทั้งในส่วนที่เป็นโหนดฟังก์ชัน และส่วนของโหนดเทอร์มินัล ค่าที่ได้จากการตัดออกจากประโยคจากการทำการกลายพันธุ์จึงอาจมีวิธีการที่แตกต่างกันดังนี้

วิธีการกลายพันธุ์ในส่วนของเทอร์มินัล สามารถทำได้ดังนี้

หนึ่ง ถ้าโหนดเทอร์มินัลเป็นใบ (Leaf) แรกสุดของต้นไม้ ตามลำดับของการค้นหาแบบทางลึก (Depth First Search) การกลายพันธุ์ทำได้โดยการสุ่มตัดค่าจากสายตัวอักขระบางส่วนทางขวาสุดไปให้โหนดเทอร์มินัลถัดไปที่อยู่ในระดับเดียวกันของต้นไม้

สอง ถ้าโหนดเทอร์มินัลเป็นใบท้ายสุดของต้นไม้ การกลายพันธุ์ทำได้โดยการสุ่มตัดค่าจากสายอักขระบางส่วนทางซ้ายสุดไปให้กับโหนดเทอร์มินัลก่อนหน้า ที่อยู่ในระดับเดียวกันของต้นไม้

สาม ถ้าโหนดเทอร์มินัลไม่เป็นใบแรกสุดและใบท้ายสุดของต้นไม้ การสุ่มเพื่อทำการตัดค่าจากสายอักขระ จะสามารถเกิดขึ้นได้ทั้งทางด้านซ้ายมือหรือทางด้านขวามือ ซึ่งถ้าสุ่มได้ทางด้านซ้ายมือก็จะสุ่มตัดค่าจากสายอักขระบางส่วนทางซ้ายสุดไปให้กับโหนดเทอร์มินัลก่อนหน้า ที่อยู่ในระดับเดียวกันของต้นไม้ แต่ถ้าสุ่มได้ทางด้านขวามือก็จะสุ่มตัดค่าจากสายตัวอักขระบางส่วนทางขวาสุดไปให้โหนดเทอร์มินัลถัดไปที่อยู่ในระดับเดียวกันของต้นไม้

วิธีการกลายพันธุ์ในส่วนของโหนดฟังก์ชัน สามารถทำได้โดยการสร้างต้นไม้แบบเติบโตให้เป็นต้นไม้ย่อยต้นใหม่ โดยเริ่มจากให้โหนดฟังก์ชันนั้นเป็นราก (Root) แต่สายอักขระภายใต้ต้นไม้ย่อยต้นใหม่จะต้องมีอักขระเริ่มต้นและจบลงเหมือนต้นไม้ย่อยเดิม

ขั้นตอนการกลายพันธุ์

หนึ่ง สุ่มเลือกโหนดที่จะทำการกลายพันธุ์จากจำนวนโหนดทั้งหมดในต้นไม้ ถ้าเป็น

โหนดฟังก์ชันให้ทำวิธีการกลายพันธุ์ในส่วนของโหนดฟังก์ชัน ถ้าโหนดที่สุ่มได้เป็นโหนดเทอร์มินัลก็ให้ทำการกลายพันธุ์ในส่วนของโหนดที่เป็นเทอร์มินัล

สอง ทำการกลายพันธุ์จนได้บุตรตามจำนวนที่กำหนดไว้

สาม ทำการคัดเลือกพันธุ์ที่มีค่าความเหมาะสมที่ดีที่สุดจากบุตรในข้อสอง แล้วแทนบุตรที่ได้คัดเลือกไปเป็นประชากรรุ่นใหม่

ขั้นตอนการทำงานของโปรแกรมเชิงเจนเนติกในงานวิจัยนี้ สามารถแบ่งเป็นขั้นตอนดังต่อไปนี้

ขั้นตอนที่หนึ่ง เมื่อมีประโยคเข้ามาให้ทำการสร้างประชากรเริ่มต้น เป็นการสร้างต้นไม้วิเคราะห์การกระจายคำจำนวนสองหมื่นต้น เพื่อนำมาเป็นประชากรเริ่มต้น โดยสามารถกระทำได้โดยการสร้างต้นไม้แบบเติบโต โดยเริ่มจากรากต้นไม้คือ โหนดฟังก์ชันเป็นกฎไวยากรณ์ของประโยคภาษาไทย เมื่อเจอโหนดฟังก์ชันก็จะสุ่มเลือกกฎที่เกี่ยวข้องกับฟังก์ชันแล้วสร้างโหนดต้นไม้ในระดับถัดไป ซึ่งจะทำในลักษณะค้นหาแบบทางลึก (Depth First Search) จนได้ใบทั้งหมดที่เป็นเทอร์มินัล โดยในงานวิจัยนี้ได้กำหนดค่าความสูงไม่เกินเจ็ด เพื่อป้องกันการสร้างต้นไม้ที่มีความสูงมากจนเกินไปสำหรับการตัดคำให้แก้ไขสามารถทำได้สองวิธีคือ วิธีที่หนึ่ง จะตัดคำแบบสุ่มในประโยคจากซ้ายไปขวา แล้วใส่ค่าที่ได้ให้แก้ไขที่สร้างขึ้นทีละใบตามลำดับในลักษณะค้นหาแบบทางลึก ยกตัวอย่างเช่น ถ้ามีอักษรที่ยังไม่ได้ตัดคำเหลือหนึ่งตัว ก็จะสุ่มเลขศูนย์ถึงเลขหนึ่ง เพื่อเป็นค่าความยาวของคำที่จะตัดออกไปให้ใบนั้นๆ ดังนั้นใบบางใบอาจไม่ได้รับคำที่ตัดเลย กรณีค่าความยาวของคำที่สุ่มมีค่าเป็นศูนย์ และมีแนวโน้มที่คำที่มีความยาวจะปรากฏบนใบในลำดับต้น ดังนั้นจึงเป็นวิธีที่ไม่ค่อยเหมาะสมกับประโยคที่มีจำนวนอักขระค่อนข้างมาก และวิธีที่สอง จะทำการสร้างต้นไม้ให้เสร็จก่อนจึงทำการตัดคำ แล้วใส่ค่านั้นกับใบในลักษณะยูนิฟอร์ม (Uniform) คือมีจำนวนอักขระในแต่ละใบเท่าๆกัน ถ้ากรณีที่มีจำนวนเทอร์มินัลมีมากกว่าจำนวนตัวอักขระให้ทำการสุ่มเลือกเทอร์มินัลจำนวนเท่ากับจำนวนตัวอักขระ แล้วกำหนดตัวอักขระหนึ่งตัวให้แก่เทอร์มินัลที่สุ่มได้จากซ้ายไปขวา ดังนั้นแต่ละเทอร์มินัลที่สุ่มได้จะมีค่าที่มีความยาวเป็นหนึ่งตัวอักขระและเทอร์มินัลที่ไม่ถูกสุ่มจะไม่มีตัวอักขระเลย ถ้ากรณีที่มีจำนวนเทอร์มินัลมีจำนวนน้อยกว่าตัวอักขระหลังจากแบ่งตัวอักขระให้กับเทอร์มินัลด้วยจำนวนที่เท่าๆกันแล้ว ให้ทำการสุ่มเลือกเทอร์มินัลให้เท่ากับจำนวนตัวอักขระที่เหลือ เพื่อเพิ่มจำนวนตัวอักขระให้พอดีกับเทอร์มินัล ตัวอย่างเช่น มีเทอร์มินัลจำนวนห้าเทอร์มินัล และมีจำนวนตัวอักขระทั้งหมดสิบสี่ตัว ดังนั้นแต่ละเทอร์มินัลจะมีตัวอักขระได้สองตัว ซึ่งจะเหลืออีกสี่ตัวอักขระ จึงทำการสุ่มเทอร์มินัลทั้งหมดสี่ครั้ง แล้วทำการเพิ่มตัวอักขระที่เหลือต่อท้ายตัวอักขระสุดท้ายแก่เทอร์มินัลที่สุ่มได้ และถ้ากรณีที่มีจำนวนเทอร์มินัลสามารถแบ่งได้เท่ากับจำนวนตัวอักขระ ก็จะทำการแบ่งจำนวนอักขระให้แต่ละเทอร์มินัลเท่าๆกัน

ขั้นตอนที่สอง ทำการคำนวณค่าเหมาะสมที่สุดจากประชากรในขั้นตอนที่หนึ่ง ค่าความเหมาะสมจะแบ่งออกเป็นสองส่วนคือ ค่าความเหมาะสมของโครงสร้างต้นไม้การวิเคราะห์กระจายค่า และค่าความเหมาะสมของการตัดค่า การคำนวณค่าความเหมาะสมของโครงสร้างสามารถทำได้ในแต่ละโหนด โดยจะทำการกำหนดคะแนนค่าความเหมาะสมให้กับโหนดเทอร์มินัลก่อน ซึ่งจะดูจากชนิดของค่าในโหนดเทอร์มินัลว่าตรงกับค่าในพจนานุกรมหรือไม่ ถ้าชนิดของค่าในโหนดเทอร์มินัลตรงกับค่าในพจนานุกรมก็จะให้คะแนนเป็น 1.0 แต่ถ้าชนิดของค่าในโหนดเทอร์มินัลไม่ตรงกับค่าในพจนานุกรมก็จะให้คะแนนเป็น 0.0 จากนั้นก็จะทำการคำนวณค่าความเหมาะสมของโหนดฟังก์ชัน โดยจะคำนวณจากค่าเฉลี่ยของค่าความเหมาะสมของโครงสร้างต้นไม้ย่อยของโหนดฟังก์ชันนั้น ให้ทำการคำนวณค่าความเหมาะสมเช่นนี้ขึ้นไปจนถึงราก ดังนั้นค่าความเหมาะสมที่ได้ก็จะอยู่ที่โหนดรากนั่นเอง การคำนวณค่าความเหมาะสมของการตัดค่า สามารถทำได้โดยพิจารณาจากค่าที่เทอร์มินัลที่มีการสะกดที่ถูกต้องและตรงกับชนิดของเทอร์มินัลหรือไม่ การให้คะแนนตามความถูกต้องของค่าในเทอร์มินัล i สามารถทำได้โดยสมการที่ 1 ดังต่อไปนี้

$$f_i = c_i * l_i \quad (1)$$

โดยที่ f_i คือ ค่าความถูกต้องของค่าในเทอร์มินัล i

c_i คือ ปัจจัย (Factor) ซึ่งถ้าค่าในเทอร์มินัลสะกดผิดจะมีค่าเป็น 0 ถ้าสะกดถูกแต่ไม่ตรงกับชนิดของเทอร์มินัลจะมีค่าเป็น 0.5 แต่ถ้าสะกดถูกและตรงกับชนิดของเทอร์มินัลจะมีค่าเป็น 1.0

l_i คือ ความยาวของตัวอักษรในค่าที่อยู่ในเทอร์มินัล

เมื่อคำนวณค่าความถูกต้องของค่าในแต่ละเทอร์มินัลเรียบร้อยแล้ว ก็จะทำการคำนวณหาค่าความเหมาะสมของการตัดค่า โดยหาได้จากค่าเฉลี่ยความถูกต้องของค่าในทุกเทอร์มินัลดังสมการที่ 2

$$f = \sum_{i=1}^p f_i / l \quad (2)$$

โดยที่ p คือ จำนวนโหนดเทอร์มินัลของต้นไม้

l คือ ความยาวของตัวอักษรของประโยค

ดังนั้นเมื่อได้ค่าความเหมาะสมของโครงสร้างและการตัดค่าแล้ว ก็จะทำการคำนวณค่าความเหมาะสมรวมดังสมการที่ 3

$$f = w_s * f_s + w_t * f_t \quad (3)$$

โดยที่ f คือ ค่าความเหมาะสมรวม

f_s คือ ค่าความเหมาะสมของโครงสร้างต้นไม้การวิเคราะห์กระจายค่า

f_t คือ ค่าความเหมาะสมของการตัดค่า

w_s คือ น้ำหนักของค่าความเหมาะสมของโครงสร้าง

w_t คือ น้ำหนักของค่าความเหมาะสมของการตัดค่า

การกำหนดค่าน้ำหนัก w_s และ w_t จะเริ่มให้น้ำหนักของ w_t มากกว่า w_s ทั้งนี้เพื่อให้ต้นไม้ส่วนใหญ่มีการตัดค่าที่โหนดเทอร์มินัลให้ถูกต้องก่อน แล้วจึงทำการปรับปรุงโครงสร้างของต้นไม้ให้ถูกต้อง โดยเริ่มต้นจะกำหนดค่า w_s ให้เท่ากับ 0.0 และ w_t เท่ากับ 1.0

ขั้นตอนที่สาม ทำการเลือกว่าจะทำการข้ามพันธุ์หรือการกลายพันธุ์หรือการคัดลอกพันธุ์ โดยถ้าค่าความเหมาะสมมีค่าใกล้เคียงหรือเท่ากับ 0.1 จะทำการข้ามพันธุ์ ถ้าค่าความเหมาะสมมีค่าใกล้เคียงหรือเท่ากับ 0.9 จะทำการกลายพันธุ์ และถ้าค่าความเหมาะสมมีค่าใกล้เคียงหรือเท่ากับ 0.0 จะทำการคัดลอกพันธุ์ ถ้าการคัดลอกพันธุ์ถูกเลือกจะทำการเลือกหนึ่งประชากรจากประชากรปัจจุบันและทำการคัดลอกพันธุ์ไปสู่ประชากรใหม่ ถ้าการข้ามพันธุ์ถูกเลือกจะทำการเลือกสองประชากรจากประชากรปัจจุบัน ทำการข้ามพันธุ์และสร้างลูกของทั้งคู่แล้วนำลูกของทั้งคู่ไปใส่ประชากรใหม่ และถ้าการกลายพันธุ์ถูกเลือกจะทำการเลือกหนึ่งประชากรจากประชากรปัจจุบัน แล้วทำการกลายพันธุ์และนำลูกที่ได้ไปใส่ในประชากรใหม่ งานวิจัยนี้จะใช้หลักการเลือกประชากรแบบการแข่งขัน (Tournament Selection) ในช่วงต้นของการทำงาน โดยจะพยายามให้ต้นไม้หรือประชากรที่มีค่าความเหมาะสมต่ำมีโอกาสถูกเลือกเข้าสู่รุ่นต่อไปได้เท่ากับประชากรที่มีค่าความเหมาะสมสูง ทั้งนี้เพราะประชากรที่มีคะแนนต่ำอาจจะมีบางส่วนของต้นไม้ที่มีส่วนประกอบที่ดี

ขั้นตอนที่สี่ ตรวจสอบจำนวนต้นไม้หรือประชากรรุ่นใหม่เท่ากับประชากรปัจจุบันหรือไม่ ถ้าไม่เท่า ให้ทำขั้นตอนที่ 2 และ 3 ซ้ำ แต่ถ้าเท่ากันให้แทนประชากรปัจจุบันด้วยประชากรรุ่นใหม่

ขั้นตอนที่ห้า ตรวจสอบค่าความเหมาะสมรวมว่ามีการเปลี่ยนแปลงหรือไม่ ถ้ามีการเปลี่ยนแปลงให้เริ่มนับจำนวนรอบเป็นหนึ่งใหม่ แต่ถ้าไม่มีการเปลี่ยนแปลงให้บวกจำนวนรอบของการทำงานเพิ่มขึ้นอีกหนึ่งรอบ หลังจากนั้นให้ทำการตรวจสอบดูว่าจำนวนรอบการทำงานเท่ากับหนึ่งร้อยรอบการทำงานหรือไม่ถ้าจำนวนรอบการทำงานเท่ากับหนึ่งร้อยรอบให้ไปทำขั้นตอนที่หก แต่ถ้าจำนวนรอบการทำงานน้อยกว่าหนึ่งร้อยรอบให้กลับไปทำขั้นตอนที่สองถึงห้าซ้ำ

ขั้นตอนที่หก ทำการกำหนดค่า w_s ให้เท่ากับ 0.2 และ w_t ให้เท่ากับ 0.8 และให้เริ่มนับ

จำนวนรอบการทำงานเป็นหนึ่งใหม่

ขั้นตอนที่เจ็ด ทำเหมือนกับขั้นตอนที่สอง

ขั้นตอนที่แปด ทำเหมือนกับขั้นตอนที่สาม

ขั้นตอนที่เก้า ตรวจสอบจำนวนต้นไม้หรือประชากรรุ่นใหม่เท่ากับประชากรปัจจุบันหรือไม่ ถ้าไม่เท่าให้ทำขั้นตอนที่เจ็ดและแปดซ้ำ แต่ถ้าเท่ากันให้แทนประชากรปัจจุบันด้วยประชากรรุ่นใหม่

ขั้นตอนที่สิบ ตรวจสอบค่าความเหมาะสมรวมว่ามีการเปลี่ยนแปลงหรือไม่ ถ้ามีการเปลี่ยนแปลงให้เริ่มนับจำนวนรอบเป็นหนึ่งใหม่ แต่ถ้าไม่มีการเปลี่ยนแปลงให้บวกจำนวนรอบของการทำงานเพิ่มขึ้นอีกหนึ่งรอบ หลังจากนั้นให้ทำการตรวจสอบว่าจำนวนรอบการทำงานเท่ากับหนึ่งร้อยรอบการทำงานหรือไม่ ถ้าจำนวนรอบการทำงานเท่ากับหนึ่งร้อยรอบก็ให้ถือว่าได้ค่าความเหมาะสมที่ดีที่สุดแล้ว ก็จะหยุดการทำงาน แต่ถ้าจำนวนรอบการทำงานน้อยกว่าหนึ่งร้อยรอบให้กลับไปทำขั้นตอนที่เจ็ดถึงสิบซ้ำ

สรุปผลการทดลองของงานวิจัยนี้สามารถทำการตัดคำ และวิเคราะห์คำในประโยคทางไวยากรณ์ ของประโยคมีความถูกต้องถึง 100% จากการทดลองกับประโยคภาษาไทยหนึ่งร้อยหกสิบประโยค โดยจะให้ค่าความเหมาะสมในส่วนโครงสร้างของต้นไม้วิเคราะห์คำในประโยคทางไวยากรณ์และส่วนของการตัดคำ แล้วนำทั้งสองส่วนมารวมกัน ซึ่งในช่วงแรกจะให้น้ำหนักความสำคัญของการตัดคำมากกว่าโครงสร้าง แล้วค่อยๆลดความสำคัญลงเพื่อให้ความสำคัญของโครงสร้างเพิ่มขึ้นจะช่วยให้การตัดคำในประโยคภาษาไทยทำได้ถูกต้อง และได้โครงสร้างของต้นไม้ที่ถูกต้องตามหลักไวยากรณ์ด้วย แต่ในส่วนของการวิเคราะห์คำในประโยคทางไวยากรณ์ในประโยคภาษาไทยยังค่อนข้างช้าโดยเฉพาะเมื่อมีจำนวนประชากรเพิ่มมากขึ้น อีกทั้งประโยคที่ใช้ในการทดลองมีน้อยเกินไป จึงควรที่จะเพิ่มเติมประโยคภาษาไทยให้มากยิ่งขึ้น เพราะจะช่วยประเมินความสามารถในการวิเคราะห์คำในประโยคทางไวยากรณ์โดยโปรแกรมเชิงเจนเนติกได้ดียิ่งขึ้น หรืออาจมีการเพิ่มกฎไวยากรณ์และจำนวนคำในพจนานุกรมให้มากยิ่งขึ้น เพื่อให้ครอบคลุมประโยคภาษาไทยอย่างสมบูรณ์

2.1.4 วิธีการใช้คลังข้อความ

การตัดคำโดยวิธีการใช้คลังข้อความมีงานวิจัยที่ได้ทำการศึกษาอยู่สี่ชิ้นคือ

2.1.4.1 งานวิจัยของ Asanee Kawtrakul, Chalathip Thumkano and Surapol Seriburi (1995) ในงานวิจัยนี้ได้นำเรื่องสถิติเข้ามาใช้ในการแก้ปัญหาการตัดคำ และการกำหนดหน้าที่ของคำหรือประเภทย่อยของคำ โดยมีการนำแบบจำลองไตรแกรมเข้ามาช่วยในการตัดคำ การตัดคำโดยใช้แบบจำลองไตรแกรม คือการตัดคำโดยมีการนำเอาค่าสถิติ ซึ่งพิจารณาจากความต่อเนื่องของหน้าที่คำหรือประเภทย่อยของคำ ส่วนวิธีการเลือกแบบการตัดคำที่ดีที่สุดนั้นทำโดยหาประโยชน์ที่มีความน่าจะเป็นมากที่สุดโดยการหาความน่าจะเป็นของแต่ละประโยคสามารถคำนวณตามสมการที่ 4 ดังนี้

$$\begin{aligned} P(W) &= \prod_{i=1}^n P(w_i | n) \\ &= \prod_{i=1}^n P(w_i | w_{i-1}, w_{i-2}) \end{aligned} \quad (4)$$

จากสมการที่ 4 คือการคำนวณหาความน่าจะเป็นของแต่ละประโยค โดย W เป็นประโยคที่ได้ตัดคำแล้ว และประโยค W ประกอบไปด้วยคำต่างๆซึ่ง $W = w_1 w_2 \dots w_n$ โดย w_i เป็นคำศัพท์ และการคำนวณหาความน่าจะเป็นของแต่ละประโยคจะมีข้อกำหนดว่า ความน่าจะเป็นของ w_i จะขึ้นกับ w_{i-1} และ w_{i-2} เท่านั้น แต่เนื่องจากการคำนวณนั้นจะต้องใช้คลังข้อความ (Corpus) ขนาดใหญ่มากและควรมีขนาดมากกว่า n^3 คำ โดยที่ n เป็นจำนวนคำที่เป็นไปได้ทั้งหมด สาเหตุที่ต้องใช้คลังข้อความขนาดใหญ่มากกว่า n^3 คำ เนื่องจากวิธีนี้จะต้องมีการนำค่าสถิติการเกิดของคำสามคำที่ติดกันมาใช้ในการคำนวณ ดังนั้นเพื่อให้มีสถิติของการเกิดของคำสามคำติดกันทุกๆแบบ อย่างน้อยที่สุดจะต้องใช้จำนวนคำเท่ากับ n^3 คำ ซึ่งในความเป็นจริงเราไม่สามารถหาคลังข้อความขนาดดังกล่าวได้ จึงทำให้มีการประมาณจากสมการที่ 4 เป็นสมการที่ 5 แทนดังนี้

$$\prod_{i=1}^n P(w_i | w_{i-1}, w_{i-2}) = \prod_{i=1}^n (\lambda_1 * P(w_i) + \lambda_2 * P(w_i | w_{i-1}) + \lambda_3 * P(w_i | w_{i-1}, w_{i-2})) \quad (5)$$

จากสมการที่ 5 นี้จะเป็นการแก้ปัญหาเรื่องจำนวนข้อมูลที่น่านำมาใช้ไม่เพียงพอ โดยจะมีการนำค่าความน่าจะเป็นของไบแกรม (Bigram) และยูนิแกรม (Unigram) เข้ามาช่วยในการคำนวณด้วย และค่า $\lambda_1 + \lambda_2 + \lambda_3$ จะต้องมีค่าเท่ากับหนึ่ง ซึ่งการคำนวณนั้นก็ใช้คลังข้อความเช่นกัน แต่

ถ้าในคลังนั้นไม่มีการเก็บค่าสถิติการเกิดของคำสองคำติดกันก็จะกำหนดค่าของ λ_2 เท่ากับศูนย์ และกรณีที่ไม่มีการเก็บค่าสถิติการเกิดของคำสามคำติดกันก็จะกำหนดค่าของ λ_3 เท่ากับศูนย์ เช่นกัน หลักการทำงานของกรวิจัยนี้จะทำการตัดคำจากประโยคที่เป็นข้อมูลเข้าที่เป็นไปได้หลายๆแบบ แล้วทำการคำนวณหาค่าความน่าจะเป็นจากแบบผลลัพธ์ต่างๆทั้งหมด แล้วนำแบบผลลัพธ์ที่มีค่าความน่าจะเป็นที่มากที่สุดมาเป็นผลลัพธ์ในการตัดคำของประโยคนั้น สรุปวิธีการนี้จะช่วยในการตัดคำได้ดี แต่จะใช้เวลาในการประมวลผลนาน เนื่องจากต้องสร้างแบบผลลัพธ์การตัดคำที่เป็นไปได้หลายๆแบบ แล้วคำนวณหาค่าความน่าจะเป็น ที่มีค่ามากที่สุดจากแบบผลลัพธ์ทั้งหมดของประโยคที่เป็นข้อมูลเข้า

2.1.4.2 งานวิจัยของ Surapant Meknavin, Paisarn Chareonponsawat and Boonserm

Kijsirikul (1997) ได้นำวิธีการทางสถิติเข้ามาช่วยในการแก้ไขปัญหาคำความกำกวม ซึ่งวิธีการทางสถิติที่นำมาใช้คือ การใช้ค่าสถิติที่เกิดจากลำดับของหน้าที่คำหรือประเภทย่อยของคำ หรืออาจกล่าวได้ว่าเป็นการนำเอาส่วนหนึ่งของไวยากรณ์มาใช้แก้ไขปัญหาคำความกำกวม งานวิจัยนี้ได้ตัดคำโดยใช้แบบจำลองไตรแกรมคือ การตัดคำโดยมีการนำเอาค่าสถิติซึ่งพิจารณาจากความต่อเนื่องของหน้าที่คำหรือประเภทย่อยของคำ ส่วนวิธีการเลือกแบบการตัดคำที่ดีที่สุดนั้นทำโดยหาประโยคที่มีความน่าจะเป็นมากที่สุดโดยการหาความน่าจะเป็นของแต่ละประโยคสามารถคำนวณตามสมการที่ 6

$$\begin{aligned} P(W_i) &= \sum_T P(W_i, T_i) \\ &= \sum_T \prod_{i=1}^n P(t_i / t_{i-1}, t_{i-2}) * P(w_i / t_i) \end{aligned} \quad (6)$$

จากสมการที่ 6 W_i เป็นประโยคที่ตัดคำแล้ว ซึ่งนำมาจากประโยคที่มีคะแนนที่ดีที่สุด N อันดับแรกจากวิธีการแม็กซิมอลแม็ชซิง (วิรัช ศรีเลิศล้ำวานิช, 2535) และ $W_i = w_1 w_2 \dots w_n$ โดย w_i เป็นคำที่ตัดได้ ส่วน $T_i = t_1 t_2 \dots t_n$ โดย t_i เป็นหน้าที่คำหรือประเภทย่อยของคำของ w_i และ $P(w_i / t_i)$ กับ $P(t_i / t_{i-1}, t_{i-2})$ สามารถคำนวณได้จากคลังข้อความ สรุปความหมายจากสมการนี้คือ การหาแบบการตัดคำที่ดีที่สุด โดยพิจารณาจากผลรวมความน่าจะเป็นของหน้าที่คำหรือประเภทย่อยของคำทุกแบบที่เป็นไปได้ของแต่ละประโยค และมีข้อกำหนดว่าค่าความน่าจะเป็นของการเกิดหน้าที่คำหรือประเภทย่อยของคำที่ตำแหน่งปัจจุบันจะขึ้นอยู่กับหน้าที่คำ หรือประเภทย่อยของคำสองคำก่อนหน้าเท่านั้น ทำให้วิธีการนี้เหมาะสมสำหรับงานที่ต้องการทราบขอบเขตคำเพียงอย่างเดียวเท่านั้น สรุปวิธีการนี้จะสามารถแก้ไขปัญหาคำความกำกวมได้ดี เนื่องจากมีการพิจารณาถึงหน้าที่ของคำหรือประเภทย่อยของคำเข้ามาประกอบด้วย แต่อย่างไรก็ตามถ้าในกรณีที่มีข้อความกำกวมมี

หน้าที่ของคำหรือประเภทย่อยของคำเหมือนกัน วิธีการนี้ก็ไม่สามารถที่จะแก้ไขปัญหาได้ดีนัก และข้อจำกัดอีกอย่างหนึ่งคือจะต้องทำการเก็บค่าสถิติจากคลังข้อความ โดยที่คลังข้อความที่ดีควรจะนำมาจากเอกสารหลายๆประเภท และจะต้องมีขนาดใหญ่พอสมควร ดังนั้นประสิทธิภาพในการตัดคำแบบนี้จะขึ้นอยู่กับคลังข้อความด้วย

2.1.4.3 งานวิจัยของไพศาล เจริญพรสวัสดิ์ (2541) เป็นงานวิจัยการตัดคำที่ได้แก้ไขปัญหาคำกำกวมและปัญหาคำศัพท์ที่ไม่ปรากฏในพจนานุกรมที่เป็นชื่อเฉพาะที่ประกอบด้วยชื่อคน ชื่อองค์กร หรือชื่อสถานที่เท่านั้น หลักการคือ นำคลังข้อความที่มีลักษณะทางไวยากรณ์ ซึ่งจะแบ่งออกเป็นสองลักษณะได้แก่ คุณลักษณะของคำบริบท (Context Word) และคุณลักษณะของสิ่งที่เกิดร่วมกัน (Collocation) โดยมีลำดับของคำที่เขียนต่อกันทั้งข้างหน้าและข้างหลังของคำกำกวมและชื่อเฉพาะนั้นๆมาพิจารณาเรียนรู้ด้วยระบบบริบเปอร์ (RIPPER) และโครงข่ายวินโนว์ (Winnow) ซึ่งภายในคลังข้อความ จะประกอบไปด้วย ข้อความหรือคำกำกวมที่มีคำบริบทที่อยู่รอบๆ ภายใน ± 10 คำ และคำศัพท์ที่ไม่ปรากฏในพจนานุกรม (Unknown Word Candidate) ที่มีคำบริบทที่อยู่รอบๆภายใน ± 10 คำ

คำกำกวม สามารถแบ่งออกเป็นสองประเภทใหญ่ๆคือ ประเภทที่หนึ่งเป็นคำกำกวมที่ไม่ได้ขึ้นกับบริบท (Context Independent) การตัดคำกำกวมแบบนี้ไม่จำเป็นต้องพึ่งบริบท เนื่องจากมีคำกำกวมคำเดียวที่เป็นไปได้ ซึ่งคำอื่นมีโอกาสเกิดขึ้นน้อยมาก เช่นข้อความ “ขนบนอก” สามารถตัดคำได้เป็น “ขนบนอก” กับ “ขนบนอก” ซึ่งคำกำกวม “ขนบนอก” มีโอกาสที่จะเกิดขึ้นมากกว่า “ขนบนอก” และประเภทที่สองเป็นคำกำกวมที่ขึ้นกับบริบท (Context Dependent) คำกำกวมที่เกิดจากการตัดคำแบบนี้มีความเป็นไปได้ทุกคำ ดังนั้นจึงต้องดูบริบทช่วยในการตัดคำ เช่น “ตากลม” สามารถตัดได้เป็น “ตากลม” กับ “ตากลม”

คำศัพท์ที่ไม่ปรากฏในพจนานุกรม สามารถแบ่งออกเป็นสองประเภทใหญ่ๆคือ ประเภทที่หนึ่งเป็นคำศัพท์ที่ไม่ปรากฏในพจนานุกรมอย่างชัดเจน (Explicit Unknown Word) คือภายในคำนั้นๆจะไม่มีส่วนใดส่วนหนึ่งหรือทุกส่วนเป็นคำที่พบอยู่ในพจนานุกรม ตัวอย่างเช่น “โลตัส” “คาร์ฟูร์” เป็นต้น และประเภทที่สองคำศัพท์ที่ไม่ปรากฏในพจนานุกรมอย่างซ่อนเร้น (Hidden Unknown Word) คือภายในคำนั้นๆจะมีส่วนใดส่วนหนึ่งหรือทุกส่วนเป็นคำที่พบอยู่ในพจนานุกรมแต่เมื่อประสมกันแล้วจะไม่พบในพจนานุกรม เช่น “สุมานี” “สมชาย” เป็นต้น

การเรียนรู้ด้วยระบบบริบเปอร์ เป็นรูปแบบของการเรียนรู้แบบกฎประพจน์ (Propositional Rule Learning) โดยจะทำการสร้างเซตของกฎ (Rule Set) ของแต่ละประเภทไว้ในข้อมูลการเรียนรู้ (Training Data) โดยกฎที่สร้างขึ้นจะอยู่ในรูปแบบเงื่อนไขที่เชื่อมด้วยและ (and) ดังนี้ if T_1 and T_2 and ... and T_n then class C_x โดยที่ T_1 and T_2 and ... and T_n คือกฎหรือเงื่อนไขต่างๆ ซึ่ง

มีรูปแบบที่เป็นไปได้ดังนี้ $A_n = v A_c \Rightarrow \emptyset A_c \Leftarrow \emptyset v \in A_s$ หรือรูปแบบตรงกันข้ามกับรูปแบบที่กล่าวมา โดยที่ A_n เป็นค่านามลักษณะทั่วไป (Nominal Attribute) v เป็นค่าที่มีในข้อมูลการเรียนรู้สำหรับ A_n A_c เป็นตัวแปรแบบต่อเนื่อง \emptyset เป็นค่าหนึ่งของ A_c ในข้อมูลการเรียนรู้ ส่วน A_s เป็นเซตของค่าคุณสมบัติ (Set-Values Attribute) v เป็นสมาชิกของ A_s และ C_x เป็นประเภทของสิ่งที่ต้องการเรียนรู้ในงานวิจัยนี้คือ คำกำกวมและคำที่ไม่ปรากฏในพจนานุกรมแต่ละคำ

การทำงานของรีปเปอร์คือ จะทำการสร้างเซตของกฎของแต่ละประเภทไว้ในข้อมูลการเรียนรู้ โดยจะทำการแบ่งข้อมูลการเรียนรู้ออกเป็นเซตการเติบโต (Growing Set) กับเซตการตัด (Pruning Set) แล้วทำการสร้างกฎจากเซตการเติบโต การสร้างกฎเริ่มแรกภายในกฎจะปราศจากเงื่อนไข แล้วทำการเพิ่มเงื่อนไขเข้าไปในกฎจากซ้ายไปขวาจนกระทั่งครอบคลุมข้อมูลการเรียนรู้ของเซตการเติบโต หลังจากกฎถูกสร้างขึ้น กฎนั้นจะถูกทำการตัด โดยจะตัดเงื่อนไขที่ละเงื่อนไขในทิศทางตรงข้าม ซึ่งจะเริ่มจากเงื่อนไขสุดท้ายของกฎ แล้วนำกฎที่ได้ไปหาข้อผิดพลาดในเซตการตัด ซึ่งจะตัดเงื่อนไขที่ทำให้ค่าผิดพลาดมากที่สุดภายในเซตการตัด หลังจากนั้นจะนำกฎที่ได้ไปใส่ไว้ในเซตของกฎ หลังจากทีกฎหนึ่งๆเพิ่มเข้าไปในเซตของกฎ ระบบจะทำการคำนวณหาค่าความยาวของคุณสมบัติของเซตของกฎและข้อมูลการเรียนรู้ ซึ่งจะทำการหยุดเพิ่มกฎ เมื่อความยาวของคุณสมบัติของเซตในกฎและข้อมูลการเรียนรู้มีค่ามากกว่าค่าความยาวของคุณสมบัติที่น้อยที่สุดที่คำนวณได้ ซึ่งได้ทำการกำหนดค่าความยาวของคุณสมบัติที่น้อยที่สุดมีค่าเท่ากับ 64 (Cohen, 1995) หรือจะหยุดเพิ่มกฎเมื่อกฎที่ได้จากการเรียนรู้ครอบคลุมข้อมูลการเรียนรู้ที่ต้องการทั้งหมด

การเรียนรู้ด้วยโครงข่ายวินโนว์ วินโนว์คือโครงข่ายที่ประกอบไปด้วยโหนดต่างๆ ที่เชื่อมต่อไปยังโหนดเป้าหมาย (Golding and Roth, 1996) แต่ละโหนดจะถูกเรียกว่าสเปเชียลลิสต์ (Specialist) ที่มีค่าคุณสมบัติ โดยจะถูกกำหนดค่าตั้งแต่ศูนย์ถึงหนึ่ง

ขั้นตอนการทำงานของโครงข่ายวินโนว์คือ

ขั้นตอนที่หนึ่ง เลือกทำนายโหนดสเปเชียลลิสต์จากเซตข้อมูลการเรียนรู้ แล้วตรวจสอบค่าคุณสมบัตินับค่าของโหนดเป้าหมายว่าเท่ากันหรือไม่ ถ้าเท่ากันจะทำการเพิ่มน้ำหนักให้กับโหนดสเปเชียลลิสต์นั้น แต่ถ้าไม่เท่ากันก็จะทำการลดน้ำหนักของโหนดสเปเชียลลิสต์นั้น

ขั้นตอนที่สอง ทำการรวมน้ำหนักของทุกโหนดสเปเชียลลิสต์ ที่เชื่อมไปยังโหนดเป้าหมายเดียวกันแล้วเลือกทำนายโหนดเป้าหมายที่มีน้ำหนักรวมมากที่สุดมาเป็นกฎในการแก้ไขปัญหา

ขั้นตอนการตัดคำในงานวิจัยนี้คือ

ขั้นตอนที่หนึ่ง ทำการตัดแบบเม็ทซ์มอลเม็ชซิ่ง (วิรัช ศรีเลิศล้ำวานิช, 2535) และใช้แบบจำลองไตรแกรม (Surapant Meknavin, Paisarn Chareonponsawat and Boonserm Kijisirikul, 1997)

ขั้นตอนที่สอง เลือกประโยคที่มีคะแนนดีที่สุด N ประโยคที่ได้ตัดคำแล้วจากข้อที่หนึ่ง
ขั้นตอนที่สาม นำกฎที่ได้จากการเรียนรู้ด้วยระบบปริปเปอร์หรือโครงข่ายวินโนว์ มาช่วย
ในการตัดสินใจเลือกประโยคที่ได้จากข้อที่สองมาเป็นผลลัพธ์

ในงานวิจัยนี้สรุปได้ว่า การนำคุณลักษณะเข้ามาใช้ในการแก้ไขปัญหาคำกำกวมใน
การตัดคำ โดยวิธีการเรียนรู้ด้วยระบบปริปเปอร์และโครงข่ายวินโนว์มีประสิทธิภาพดีมากโดยเฉพาะ
วิธีการเรียนรู้ด้วยโครงข่ายวินโนว์ ซึ่งสามารถแสดงการเปรียบเทียบตามตารางที่ 2.3 ดังนี้

ตารางที่ 2.2 แสดงการเปรียบเทียบการตัดคำ แบบคุณลักษณะของคำบริบทที่เรียนรู้โดยระบบ
ปริปเปอร์ แบบคุณลักษณะของคำบริบทที่เรียนรู้โดยโครงข่ายวินโนว์ แบบลักษณะ
สิ่งที่เกิดร่วมกัน โดยมีลำดับของคำที่เขียนต่อกันทั้งข้างหน้าและข้างหลังที่เรียนรู้โดย
ระบบปริปเปอร์ และแบบลักษณะสิ่งที่เกิดร่วมกัน โดยมีลำดับของคำที่เขียนต่อกันทั้ง
ข้างหน้าและข้างหลังที่เรียนรู้โดยโครงข่ายวินโนว์

วิธีการตัดคำกำกวม	ความถูกต้องของวิธีการตัดคำ (%)			
	ข้อมูลในการเรียนรู้		ข้อมูลที่ใช้ทดสอบ	
	ขึ้นกับ	ไม่ขึ้นกับ	ขึ้นกับ	ไม่ขึ้นกับ
	บริบท	บริบท	บริบท	บริบท
คุณลักษณะของคำบริบทที่เรียนรู้ โดยระบบปริปเปอร์	96.98	99.94	86.60	99.74
คุณลักษณะของคำบริบทที่เรียนรู้ โดยโครงข่ายวินโนว์	100.00	100.00	95.33	99.70
ลักษณะสิ่งที่เกิดร่วมกัน โดยมี ลำดับของคำที่เขียนต่อกันทั้ง ข้างหน้าและข้างหลังที่เรียนรู้ โดยระบบปริปเปอร์	93.28	98.52	89.00	91.27
ลักษณะสิ่งที่เกิดร่วมกัน โดยมี ลำดับของคำที่เขียนต่อกันทั้ง ข้างหน้าและข้างหลังที่เรียนรู้ โดยโครงข่ายวินโนว์	100.00	99.97	92.10	93.82

แต่ในการแก้ปัญหาปัญหาคำศัพท์ที่ไม่ปรากฏในพจนานุกรมนั้น โดยเฉพาะคำศัพท์ที่ไม่ปรากฏในพจนานุกรมแบบออนไลน์จะแก้ปัญหาคำศัพท์ได้ค่อนข้างต่ำ เนื่องจากการค้นหาคำศัพท์ประเภทนี้ทำได้ยาก ซึ่งจะต้องหาวิธีที่จะค้นหาบริเวณที่เกิดคำศัพท์ประเภทนี้ให้มีประสิทธิภาพสูงขึ้นไป

2.1.4.4 งานวิจัยของชมภูนุช คุปติวุฒิ (2542) เป็นงานวิจัยการตัดคำคำกวมในข้อความภาษาไทยด้วยโปรแกรมตรรกะเชิงอุปนัย โดยจะนำคุณลักษณะทางไวยากรณ์มาใช้ในการตัดคำคำกวม โดยลักษณะทางไวยากรณ์ที่นำมาใช้คือ คำบริบทและคำที่อยู่ก่อนหน้าและหลังคำคำกวมมาเป็นคุณลักษณะ ในการเรียนรู้แบบโปรแกรมตรรกะเชิงอุปนัยได้นำระบบ FOIL มาใช้เพื่อให้ได้อนุประโยคผลลัพธ์ของคุณลักษณะ หรือกฎของแต่ละคำคำกวมที่จะนำไปใช้ในการตัดคำคำกวม

กรรมวิธีการเรียนรู้คำคำกวมในงานวิจัยนี้ จะแบ่งข้อความออกเป็นสองประเภทคือ

ประเภทที่หนึ่ง แบบไม่ขึ้นอยู่กับบริบท (Context Independent) การตัดคำคำกวมแบบไม่จำเป็นต้องพึ่งบริบท เนื่องจากมีคำคำกวมคำเดียวที่เป็นไปได้ ซึ่งคำอื่นมีโอกาสเกิดขึ้นน้อยมาก เช่นข้อความ “กรอบ” สามารถตัดคำได้เป็น “กร อบ” กับ “กรอบ” ซึ่งคำคำกวม “กรอบ” มีโอกาสที่จะเกิดขึ้นมากกว่า “กร อบ” แต่อย่างไรก็ตามวิธีการตัดคำแบบแม็กซิมอลเอนซ์ซึ่งนั้นก็ยังคงตัดคำคำกวมแบบนี้ไม่ถูกต้อง 100% เช่นข้อความ “พอกกลางคืน” สามารถตัดได้เป็น “พอก กลางคืน” กับ “พอก กลาง คืน”

ประเภทที่สอง แบบขึ้นอยู่กับบริบท (Context Dependent) คำคำกวมที่เกิดจากการตัดคำแบบนี้มีความเป็นไปได้ทุกคำ ดังนั้นจึงต้องดูบริบทช่วยในการตัดคำ เช่น

“ตากลม” สามารถตัดได้เป็น “ตาก ลม” กับ “ตา กลม”

“ตัวเล็ก” สามารถตัดได้เป็น “ตัวเล็ก” กับ “ตัว เล็ก”

การเรียนรู้โดยใช้ระบบ FOIL เป็นการนำกลุ่มของอนุประโยคผลลัพธ์ที่ได้จากการเรียนรู้ของแต่ละคำคำกวมของกลุ่มข้อมูลสอน ซึ่งในงานวิจัยนี้เรียนรู้คำคำกวมทั้งหมดหนึ่งร้อยห้าสิบคำมาเป็นเงื่อนไขช่วยในการตัดคำในแต่ละประโยคของประโยคแบบต่างๆ ที่คะแนนดีที่สุด N ประโยค ซึ่งผ่านการตัดคำด้วยวิธีการตัดคำแบบจำลองไตรแกรม โดยให้คะแนนแต่ละประโยคของ N ประโยคที่ได้จากการตัดคำด้วยวิธีแบบจำลองไตรแกรมมีค่าเท่ากับ 0 และเมื่อตรวจสอบพบว่าในประโยคมีคำคำกวมและเป็นไปตามเงื่อนไขที่เรียนรู้ด้วย FOIL แล้วจะบวกคะแนนของประโยคเพิ่มขึ้น 1 สำหรับแต่ละคำคำกวมแล้วจะทำการเรียงลำดับคะแนนของแต่ละประโยคจากมากไปหาน้อย ประโยคที่เป็นคำตอบจะเป็นประโยคที่มีคะแนนมากที่สุด แต่ถ้าคะแนนของประโยคมีค่าเท่ากันจะใช้ค่าความน่าจะเป็นของประโยคเป็นตัวตัดสิน ซึ่งการวัดความถูกต้องในการตัดคำคำกวมเป็นการดูเฉพาะคำคำกวมในประโยคที่เป็นคำตอบเท่านั้น โดยถือว่าประโยคหนึ่งประโยคจะมีคำคำกวมเพียงคำเดียวเท่านั้น งานวิจัยนี้สรุปได้ว่าการนำอนุประโยคผลลัพธ์ หรือกฎของแต่ละ

คำคำถามที่ได้จากการเรียนรู้ด้วยโปรแกรมตรรกะเชิงอุปนัยโดยระบบ FOIL สามารถช่วยตัดคำคำถามได้ดีกว่าวิธีการตัดคำโดยแบบจำลองไตรแกรม ซึ่งสามารถแสดงการเปรียบเทียบตามตารางที่ 2.4

ตารางที่ 2.3 แสดงการเปรียบเทียบการตัดคำคำถามโดยระบบ FOIL และแบบจำลองไตรแกรม

วิธีการตัดคำคำถาม	ความถูกต้องของวิธีการตัดคำ (%)			
	ข้อมูลสอนเรียงตามลำดับ		ข้อมูลทดสอบเรียงตามลำดับ	
	ขึ้นกับ	ไม่ขึ้นกับ	ขึ้นกับ	ไม่ขึ้นกับ
	บริบท	บริบท	บริบท	บริบท
ระบบ FOIL	89.93	97.42	83.27	88.24
แบบจำลองไตรแกรม	87.19	92.27	80.54	82.35

ผลที่ได้จากการตัดคำจากการเรียนรู้ด้วยระบบ FOIL ของงานวิจัยนี้ให้ความถูกต้องไม่มากนัก เนื่องจากยังมีคำถามอื่นอีกที่ยังไม่ได้นำไปเรียนรู้ ดังนั้นถ้าจะให้ความถูกต้องจากการตัดคำคำถามมากขึ้นก็ควรจะทำการเรียนรู้กฎที่นิยามของคำถามทั้งหมด

2.2 สรุป

การตัดคำในภาษาไทยได้มีการพัฒนาติดต่อกันมาเป็นเวลานาน ทำให้มีงานวิจัยเกิดขึ้นมากมายหลายวิธี ซึ่งข้อสรุปที่ได้จากวิธีการต่างๆ ก็ให้ความถูกต้องค่อนข้างสูง แต่ในทางด้านเวลาในการประมวลผลและความซับซ้อนของอัลกอริทึมแล้วยังมีความแตกต่างกัน ดังนั้นในการเลือกนำไปใช้งานจึงต้องทำการวิเคราะห์ข้อดีข้อเสีย ในตัดสินใจเลือกให้เหมาะสมกับงานที่จะนำไปใช้มากที่สุด ในงานวิจัยฉบับนี้มีความจำเป็นอย่างยิ่งที่ต้องอาศัยโปรแกรมตัดคำในข้อความภาษาไทยที่มีความถูกต้องสูง ใช้เวลาน้อยในการประมวลผล ประหยัดเนื้อที่ในการจัดเก็บและสามารถเพิ่มแก้ไขได้อย่างสะดวกรวดเร็ว รวมทั้งในการค้นหาและแทนที่คำหายขาดด้วยเครื่องหมายดอกจันนั้นจะต้องอาศัยตัวแบบ ที่สามารถเป็นแนวทางในการนำไปพัฒนาโปรแกรมค้นหาและแทนที่คำหายขาดด้วยเครื่องหมายดอกจันได้ และสามารถอธิบายการทำงานให้เกิดความเข้าใจได้โดยง่าย ในบทต่อไปจะกล่าวถึงนิยามและทฤษฎีที่ใช้ในงานวิจัยครั้งนี้เพื่อเป็นพื้นฐานในการดำเนินการขั้นต่อไป

บทที่ 3

นิยามและทฤษฎีที่ใช้ในงานวิจัย

ในบทนี้จะกล่าวถึงการกำหนดนิยามของคำและข้อความ รวมทั้งทฤษฎีที่นำมาใช้ในงานวิจัยฉบับนี้ โดยมีรายละเอียดดังต่อไปนี้

3.1 นิยาม

จากการศึกษานิยามของคำได้มีบุคคลหลายท่านได้ให้นิยามของคำซึ่งแบ่งออกเป็นประเภทต่างๆหลายประเภท ยกตัวอย่างเช่น คำบ่งบอกคือ คำที่ปรากฏติดกับชื่อเฉพาะ (อมรทิพย์ กวินปณิธาน, 2546) คำมูลคือ หน่วยคำที่เล็กที่สุดที่มีความหมาย (อุปกิต ศิลปสาร, 2533) “คำนาม” คือ คำที่ใช้เรียกชื่อของคน สัตว์ สิ่งของ สถานที่ สภาพ และกริยาอาการต่างๆในภาษา (อุปกิต ศิลปสาร, 2533) และคำโดดคือ คำพยางค์เดียวและจะมีเพียงรูปเดียวไม่มีการเปลี่ยนแปลงรูปคำหรือเติมวิภัติปัจจัย (อุปกิต ศิลปสาร, 2533) เป็นต้น ซึ่งงานวิจัยนี้จำเป็นต้องกำหนดนิยามของคำต่างๆที่เกี่ยวข้องอย่างชัดเจน เพื่อเป็นพื้นฐานในการดำเนินการขั้นต่อไป ดังนั้นผู้วิจัยจึงขอ กำหนดนิยามของคำหรือข้อความที่เกี่ยวข้องดังต่อไปนี้

3.1.1 คำ หมายถึง กลุ่มอักขระที่ได้จากการประมวลผลโดยโปรแกรมตัดคำโดยพจนานุกรมแบบการเทียบคำที่ยาวที่สุด

3.1.2 คำหยาบ หมายถึง คำที่ความหมายต่อไปนี้ในทางพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2542 คือ ชุกโคตรขึ้นมาค่า คัดจริต อาการทำกันขึ้นๆลงๆ อวัยวะสืบพันธุ์ของผู้หญิงหรือผู้ชาย เลว คำซ้ำ ชั่ว ลูกอ้นทะ หลึงมากชั่วหลายชั่ว การร่วมประเวณีหรือร่วมเพศ หลึงมีความใคร่ ชั่วร้าย เสนียดจัญไร ไม่ดีหรือไม่งาม คนเลวหรือคนโง่ หยาบเข้า หยาบคาย ทรามหรือเลวทราม หยาบโลน น่าขยะแขยง หลึงชั่ว คบชู้หรือมีชู้ ขนที่ลับ ปลายลึงค์ที่หนังหุ้มร่นเข้าไป อาการที่นั่งหรือนอนถ่างขาอย่างเปิดเผย มีใจฝักใฝ่ในทางชั่วร้าย ผู้ล่วงประเวณี คนแก่หรือคนมีอายุมากที่มีเล่ห์เหลี่ยมหรือกลอุบายหลอกเด็กผู้หญิงในทางกามารมณ์ อวัยวะส่วนล่างสุดของ

คนหรือสัตว์นับตั้งแต่ได้ซื้อเท้าลงไปสำหรับยืนหรือเดิน ทำนอกกลุ่มนอกทาง ชายที่ประพฤติดนอย่างนางบำเรอ ความเลวร้ายที่มีอยู่ในตัว ร่วมกันข่มขืนกระทำชำเรา หญิงที่หาเลี้ยงชีพด้วยการค้าประเวณี ปากร้าย ชอบพูดจาวาไรยคนอื่น ส่วนของร่างกายอยู่ระหว่างท้องน้อยกับอวัยวะสืบพันธุ์ รวมไปถึงคำที่แสดงความรู้สึกไม่สุภาพ คำที่มักใช้ในทางกามารมณ์หรือความใคร่ในกามคุณ คำที่มักใช้ในทางไม่ดี คำที่เรียกรูปจำลองอวัยวะเพศหญิงหรือชาย คำที่ใช้เป็นคำด่าหรือคำแข่ง คำที่ใช้เรียกหญิงที่มากด้วยราคะจริตและโทสะจริต คำที่มักใช้ในความประชดหรือแค้น คำที่ถือว่าผู้หญิงที่ประพฤติดัวไม่สมควร คำที่แสดงความรู้สึกเมื่อประสบสิ่งที่ขัดหูขัดตาหรือเป็นที่น่าสมเพชเป็นที่น่ารังเกียจ และคำที่ใช้ในการดูถูกดูหมิ่นหรือเหยียดหยาม

3.1.3 คำปกติ หมายถึง คำต่างๆที่มีความหมายไม่ตรงกับความหมายของคำหาบวมทั้งอักษรและกลุ่มอักษรที่ไม่ปรากฏในพจนานุกรมฉบับราชบัณฑิตยสถาน พ.ศ. 2542

3.1.4 ข้อความ หมายถึง อักษรตั้งแต่หนึ่งตัวขึ้นไปที่น่ามาเขียนรวมกันโดยมีช่องว่างหรือการขึ้นบรรทัดใหม่เป็นขอบเขตของการสิ้นสุด

3.1.5 ข้อความพิเศษ หมายถึง ข้อความที่เกิดจากการประสมของคำศัพท์ในพจนานุกรม แต่เมื่อตัดคำแบบเทียบคำที่ยาวที่สุดแล้วทำให้ได้ผลลัพธ์ไม่ถูกต้อง

3.2 ทฤษฎีที่ใช้ในงานวิจัย

3.2.1 การตัดคำ (Word Segmentation) เนื่องจากระบบจำเป็นอย่างยิ่งที่จะต้องหาขอบเขตของแต่ละคำเพื่อที่จะสามารถทำการประมวลผลกับข้อความในเอกสารได้อย่างสะดวก และจากลักษณะการเขียนของภาษาไทยนั้นไม่มีการใช้ตัวอักษรหรือสัญลักษณ์ที่น่ามาใช้คั่นระหว่างคำ หรือว่ามีช่องว่างระหว่างคำเหมือนภาษาอังกฤษ และงานต่างๆ ในด้านการประมวลผลภาษาไทยนั้นจำเป็นอย่างยิ่งที่ต้องทราบขอบเขตของคำหรือต้องตัดคำก่อนเป็นอันดับแรก (ไพศาล เจริญพร สวัสดิ์, 2541) ดังนั้นการตัดคำในภาษาไทยจึงถือเป็นสิ่งสำคัญ การตัดคำคือ การแบ่งตัวอักษรจากข้อความ (String) เพื่อหาขอบเขตของแต่ละหน่วยคำ (Morpheme) เนื่องจากส่วนใหญ่ภาษาไทยมีการเขียนในลักษณะที่ติดกัน โดยไม่มีการใช้เครื่องหมายวรรคตอนคั่นระหว่างคำเหมือนภาษาอังกฤษ ซึ่งใช้ช่องว่าง (Space) คั่นระหว่างคำ แต่ภาษาไทยจะมีการเว้นวรรคเป็นระยะ เพื่อให้

ผู้อ่านทำความเข้าใจกับความหมายของข้อความ วิธีการตัดคำภาษาไทยสามารถแบ่งได้เป็นสี่หลักการใหญ่ คือหลักการตัดคำโดยใช้กฎ (Rule Base Approach) หลักการตัดคำโดยใช้อัลกอริทึม (Algorithm Approach) หลักการตัดคำโดยใช้พจนานุกรม (Dictionary Approach) และหลักการตัดคำโดยใช้คลังข้อมูล (Corpus Base Approach) แต่ละวิธีการต่างก็ให้ผลในด้านความถูกต้อง ความรวดเร็วของการทำงานและปริมาณการใช้ทรัพยากรต่างๆที่แตกต่างกัน การตัดคำในวิทยานิพนธ์ฉบับนี้ ได้ใช้หลักการตัดคำโดยใช้พจนานุกรมแบบการเทียบคำที่ยาวที่สุด (Longest Matching) แต่จากการศึกษาการตัดคำด้วยวิธีนี้ยังมีบางกรณีที่ตัดคำไม่ถูกต้อง จึงได้ทำการปรับปรุงวิธีการตัดคำเพิ่มเติมโดยใช้วิธีการเทียบข้อความ (String Matching) ก่อน เพื่อให้การตัดคำมีความถูกต้องมากขึ้น ดังนั้นหลักในการตัดคำในวิทยานิพนธ์ฉบับนี้คือ จะใช้วิธีการเทียบข้อความก่อนถ้าเทียบข้อความไม่พบก็จะใช้วิธีการเทียบคำที่ยาวที่สุดแทน แต่ถ้าวิธีการเทียบคำที่ยาวที่สุดเทียบคำไม่พบก็จะทำการลดอักษรทางด้านขวาของข้อความลง เมื่อลดอักษรลงแล้วก็ให้กลับไปใช้วิธีเทียบข้อความอีกครั้ง ถ้าเทียบข้อความไม่พบก็จะใช้วิธีการเทียบคำยาวที่สุดแทน ซึ่งจะซ้ำทั้งสองวิธีการนี้ไปเรื่อยๆ จนได้ผลลัพธ์ในการตัดคำ การตัดคำแบบเทียบข้อความ เป็นการเก็บข้อความที่เกิดจากการประสมของคำศัพท์ในพจนานุกรมแต่เมื่อตัดคำแบบเทียบคำที่ยาวที่สุดแล้วทำให้ได้ผลลัพธ์ไม่ถูกต้อง โดยการจับเก็บข้อความนั้นก็จะทำการจับเก็บผลลัพธ์ในการตัดคำของข้อความนั้นด้วย ดังตารางต่อไปนี้

ตารางที่ 3.1 แสดงตัวอย่างข้อความที่เกิดจากการประสมของคำศัพท์ในพจนานุกรมซึ่งถ้าตัดคำแบบเทียบคำที่ยาวที่สุดแล้วจะทำให้ได้ผลลัพธ์ไม่ถูกต้อง และผลลัพธ์ที่จัดเก็บไว้เมื่อเทียบข้อความพบ

ข้อความ	เกิดจากการประสมระหว่าง	ผลลัพธ์จากการตัดคำแบบเทียบคำที่ยาวที่สุด	ผลลัพธ์ที่จัดเก็บไว้เมื่อเทียบข้อความพบ
ผลกระทบ	ผลก + ระ + ทบ หรือ ผล + กระทบ	ผลก / ระ / ทบ	ผล / กระทบ
ต้องตาย	ต้อง + ตาย	ต้องตา / ย	ต้อง / ตาย
หากระดาษ	หา + กระดาษ	หาก / ระดา / ษ	หา / กระดาษ
การกระทำ	การก + ระ + ทำ หรือ การ + กระทำ	การก / ระ / ทำ	การ / กระทำ

จากตารางที่ 3.1 เป็นตัวอย่างข้อความที่เกิดจากการประสมของคำศัพท์ในพจนานุกรมซึ่งถ้าตัดคำแบบเทียบคำที่ยาวที่สุดแล้วจะทำให้ได้ผลลัพธ์ไม่ถูกต้อง จึงต้องใช้วิธีการตัดคำแบบเทียบข้อความก่อน เช่นข้อความ “หากระดาษ” จะทำการนำข้อความไปเทียบกับข้อความที่ได้ทำการจัดเก็บ ถ้าเทียบพบก็จะนำผลลัพธ์ที่ได้จัดเก็บไว้มาเป็นผลลัพธ์ในการตัดคำ ดังนั้นผลลัพธ์ในการตัดคำของข้อความนี้คือ “หา / ระดาษ” การตัดคำแบบเทียบคำที่ยาวที่สุด เป็นวิธีการตัดคำที่ต้องใช้พจนานุกรมช่วยจำคำภาษาไทย ซึ่งก็คือต้องมีคำทุกคำที่มีอยู่ในภาษาไทยทั้งหมดลงพจนานุกรมโดยวิธีนี้จะทำการตรวจสอบหรือสแกนข้อความที่ป้อนเข้ามาจากซ้ายไปขวา การตรวจสอบคำจะเริ่มจากหน่วยของข้อความหรือสายอักขระที่สั้นที่สุดที่สามารถตรวจสอบได้ โดยจะตรวจสอบกับจุดสิ้นสุดของประโยค เช่น การขึ้นย่อหน้า การเว้นวรรคตอน เป็นต้น แล้วนำไปเทียบกับพจนานุกรม หากไม่พบก็ทำการลดข้อความลงไปเรื่อยๆ จนกว่าข้อความที่ตรวจสอบจะสามารถเทียบเป็นคำในพจนานุกรมได้ ก็จะทำการเครื่องหมายเพื่อเป็นจุดย้อนกลับ จากนั้นก็เริ่มทำงานจากจุดย้อนกลับนั้น เพื่อตรวจสอบข้อความที่เหลือว่าจะสามารถตัดข้อความใดต่อไปให้เป็นคำได้ หากตัวเลือกในตอนแรกนี้สามารถทำให้ขั้นตอนวิธีค้นหาคำที่เหลือได้ ตัวเลือกนี้ก็จะเป็คำแรกของข้อความได้จริงไม่เช่นนั้นขั้นตอนวิธีก็จะกลับไปยังจุดย้อนกลับที่ทำเครื่องหมายไว้ เพื่อแก้ไขคำแรกใหม่ จากนั้นจะเริ่มทำงานต่อไปโดยเริ่มจากจุดย้อนกลับ หากยังไม่สามารถเทียบสายอักขระกับคำในพจนานุกรมได้ ก็จะทำการลดตัวอักษรลงทีละตัวจนกว่าจะเทียบคำในพจนานุกรมได้ ซึ่งการลดอักษรตัวอักษรลงทีละตัวจะมีกฎทางอักขระวิธี (นัฐวุฒิ ไชยเจริญ, 2544) จะเป็นการพิจารณาความเป็นไปได้ในการที่จะแยกสายอักขระเพื่อพิจารณาคำ ซึ่งจะประกอบด้วย

อักขระกลุ่มที่ 1 กลุ่ม Non – spacing Character คือ กลุ่มของรูปสระ วรรณยุกต์ และเครื่องหมายพิเศษประกอบการเขียน ซึ่งเมื่อประสมเข้ากับพยัญชนะใดๆ แล้ว ไม่ทำให้มีการเคลื่อนขวาของตำแหน่งที่เขียนต่อไป ดังนั้นอักขระในกลุ่มนี้ไม่สามารถที่จะอยู่เดี่ยวๆ ได้ ประกอบด้วย ๐ ๑ ๒ ๓ ๔ ๕ ๖ ๗ ๘ ๙ และ ๐

อักขระกลุ่มที่ 2 กลุ่มอักขระที่จำเป็นที่จะต้องมึรูปพยัญชนะตามเสมอ ประกอบด้วยสระ แ โ ไ และ ใ

อักขระกลุ่มที่ 3 กลุ่มอักขระที่จำเป็นที่จะต้องมึรูปพยัญชนะอยู่หน้าเสมอ ประกอบด้วย ะ ำ ำ ฤ

อักขระกลุ่มที่ 4 กลุ่มของอักขระที่เป็นตัวการันต์ที่ไม่มีทัณฑฆาตบังคับข้างบน เช่น ย์ ใน “ทิพย์” และ ์ ใน “กาญจน์” เป็นต้น เนื่องจากว่าตัวการันต์เป็นพยัญชนะตัวสุดท้ายไม่อ่านออกเสียง จึงไม่มีการพิจารณาให้เป็นตัวอักษรแรกของคำ

อักขระกลุ่มที่ 5 กลุ่มของอักษรที่เหลือทั้งหมด

จากการตัดคำโดยวิธีการเทียบคำที่ยาวที่สุด (สัมพันธ ะระรินรมย์, 2534) ภายใต้เงื่อนไขของ อักษรวิธี เช่น ข้อความ “เขาไปตลาด” ข้อความนี้จะนำไปเทียบกับพจนานุกรมก่อน ถ้าไม่สามารถ เทียบคำกับพจนานุกรมได้ก็จะลดข้อความลงเหลือ “เขาไปตลาด” แล้วนำไปเทียบกับข้อความก่อน ถ้าเทียบข้อความไม่ได้จะนำไปเทียบกับพจนานุกรมอีก ถ้าเทียบกับพจนานุกรมไม่ได้ก็จะลดอักษร ทางด้านขวาของอีกเป็น “เขาไปต” โดยทำซ้ำแบบนี้เรื่อยๆ จนสามารถเทียบข้อความหรือเทียบคำกับ พจนานุกรมได้ ซึ่งก็คือคำว่า “เขา” จึงตัดเป็นคำแรกของข้อความและทำเครื่องหมายไว้เป็นจุด ย้อนกลับ จากนั้นจะทำการพิจารณาข้อความที่เหลือจากจุดย้อนกลับ ในที่นี้ก็คือ ข้อความ “ไป ตลาด” ซึ่งผลลัพธ์สุดท้ายที่ได้จะเป็น “เขา ไป ตลาด”

3.2.2 ไฟไนต์สเตตแมชชีนด้วยข้อมูลออก (Finite State Machines with Outputs)

เนื่องจากวิทยานิพนธ์ฉบับนี้ต้องการกรองคำหายาบ เพื่อไม่ให้คำหายาบปรากฏในข้อความ ภาษาไทย ดังนั้นในงานวิจัยฉบับนี้ขอเลือกตัวแบบทางคณิตศาสตร์ไฟไนต์สเตตแมชชีนด้วย ข้อมูลออก เพื่อเป็นแนวทางในการพัฒนาโปรแกรมกรองคำหายาบและอธิบายการแทนที่คำหายาบ ด้วยเครื่องหมายดอกจัน

ไฟไนต์สเตตแมชชีนด้วยข้อมูลออกคือ ตัวแบบทางคณิตศาสตร์ที่ใช้ในการอธิบายการ ทำงานที่เริ่มทำงานจากสถานะเริ่มต้นและรับข้อมูลเข้าเป็นตัวอักขระ แล้วเปลี่ยนสถานะไปยัง สถานะถัดไป การทำงานในแต่ละขั้นตอนนี้สามารถเปลี่ยนสถานะไปยังสถานะถัดไปได้ โดย ขึ้นอยู่กับสถานะปัจจุบันและตัวอักขระที่เป็นข้อมูลเข้าเท่านั้น การอธิบายการทำงานของ ไฟไนต์สเตตแมชชีนอีกวิธีหนึ่งคือ การใช้กราฟโดยจะแสดงสถานะด้วยวงกลม (Vertex) และการ เปลี่ยนสถานะด้วยเส้นเชื่อมระหว่างสถานะ (Edge) โดยบนเส้นเชื่อมจะอธิบายข้อมูลเข้าตามด้วย เครื่องหมายคอมม่าและข้อมูลออก หรืออาจจะไม่มีข้อมูลออกก็ได้ซึ่งขึ้นอยู่กับการทำงานของระบบ

ในทางคณิตศาสตร์ได้กำหนดให้ $M = (S, I, O, f, g, s_0)$ โดยที่

M เป็น ไฟไนต์สเตตแมชชีน

S เป็น เซตของสถานะ

I เป็น เซตของอักขระที่เป็นข้อมูลเข้า

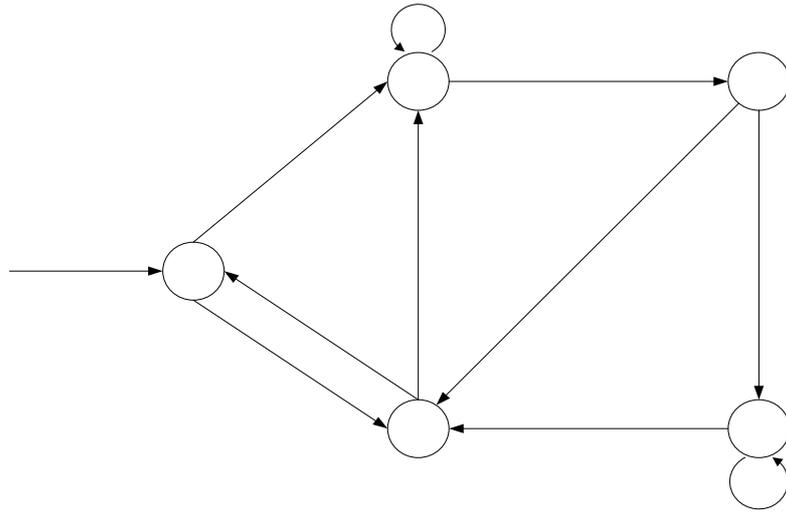
O เป็น เซตของอักขระที่เป็นข้อมูลออก

f เป็น ฟังก์ชันอธิบายการเปลี่ยนแปลงสถานะเมื่อมีข้อมูลเข้า

g เป็น ฟังก์ชันอธิบายข้อมูลออกเมื่อมีข้อมูลเข้า

และ s_0 เป็น สถานะเริ่มต้น

เราสามารถใช้อะไอะแกรมสถานะ (State Diagram) และตารางสถานะ (State Table) ในการอธิบายฟังก์ชัน f และ g ได้ (Rosen, 2003) ดังภาพที่ 3.1 และตารางที่ 3.2 ตามลำดับ



ภาพที่ 3.1 แสดงไอะแกรมสถานะอธิบายไฟในท์สเตทแมชชีนด้วยข้อมูลออก
แหล่งที่มา: Rosen, 2003: 754.

จากภาพที่ 3.1 สามารถอธิบายไอะแกรมสถานะได้คือ เริ่มต้นที่สถานะ s_0 ถ้าข้อมูลเข้าเป็น 0, 1 จะได้ข้อมูลออกเป็น 1 และจะเปลี่ยนสถานะเป็นสถานะ s_1 แต่ถ้าข้อมูลเข้าเป็น 1 จะได้ข้อมูลออกเป็น 0 และจะเปลี่ยนสถานะเป็นสถานะ s_3 ที่สถานะ s_1 ถ้าข้อมูลเข้าเป็น 0 จะได้ข้อมูลออกเป็น 1 และจะไม่เปลี่ยนสถานะ แต่ถ้าข้อมูลเข้าเป็น 1 จะได้ข้อมูลออกเป็น 1 และจะเปลี่ยนสถานะเป็นสถานะ s_2 ที่สถานะ s_2 ถ้าข้อมูลเข้าเป็น 0 จะได้ข้อมูลออกเป็น 0 และจะเปลี่ยนสถานะเป็นสถานะ s_3 แต่ถ้าข้อมูลเข้าเป็น 1 จะได้ข้อมูลออกเป็น 0 และจะเปลี่ยนสถานะเป็นสถานะ s_4 ที่สถานะ s_3 ถ้าข้อมูลเข้าเป็น 0 จะได้ข้อมูลออกเป็น 0 และจะเปลี่ยนสถานะเป็นสถานะ s_1 แต่ถ้าข้อมูลเข้าเป็น 1, 0 จะได้ข้อมูลออกเป็น 0 และจะเปลี่ยนสถานะเป็นสถานะ s_0 และที่สถานะ s_4 ถ้าข้อมูลเข้าเป็น 0 จะได้ข้อมูลออกเป็น 0 และจะเปลี่ยนสถานะเป็นสถานะ s_3 แต่ถ้าข้อมูลเข้าเป็น 1, 0 จะได้ข้อมูลออกเป็น 0 และจะไม่เปลี่ยนสถานะ นอกจากนี้ยังสามารถเขียนตารางสถานะซึ่งแสดงการทำงานเช่นเดียวกับไอะแกรมสถานะข้างต้นคือ

ตารางที่ 3.2 แสดงสถานะอธิบายไฟไนท์สเตตแมชชีนด้วยข้อมูลออกจากไดอะแกรมสถานะในภาพที่ 3.1

State	f		g	
	Input		Input	
	0	1	0	1
Output				
s_0	s_1	s_3	1	0
s_1	s_1	s_2	1	1
s_2	s_3	s_4	0	0
s_3	s_1	s_0	0	0
s_4	s_3	s_4	0	0

แหล่งที่มา: Rosen, 2003: 754.

การหาข้อมูลออกจากข้อมูลเข้า 101011 จะพิจารณาจากไดอะแกรมสถานะหรือตารางสถานะก็ได้ โดยพิจารณาข้อมูลเข้าทีละ 1 ตัวอักษร (จากซ้ายไปขวา) สมมติว่าพิจารณาจากตารางสถานะ เมื่อเริ่มต้นการทำงานของระบบจะอยู่ที่สถานะ s_0 ข้อมูลเข้าตัวแรกคือ 1 ฟังก์ชันข้อมูลออก g กำหนดให้ข้อมูลออก คือ 0 และฟังก์ชันการเปลี่ยนแปลงสถานะกำหนดให้เปลี่ยนสถานะไปยังสถานะ s_3 จาก s_0 ข้อมูลเข้าอักขระถัดไปคือ 0 ซึ่ง $g(s_3, 0) = 0$ และ $f(s_3, 0) = s_1$ ณ สถานะ s_1 ข้อมูลเข้าอักขระถัดไปคือ 1 ซึ่ง $g(s_1, 1) = 1$ และ $f(s_1, 1) = s_2$ ให้ดำเนินการเช่นนี้จนหมดข้อมูลเข้า ในกรณีนี้ระบบจะหยุดที่สถานะ s_3 โดยมีลำดับสถานะทั้งหมดจากสถานะเริ่มต้นคือ $s_0 s_3 s_1 s_2 s_3 s_0 s_3$ และข้อมูลออกคือ 001000

3.2.3 โครงสร้างพจนานุกรมและคำหยาบโดยต้นไม้เอวีแอล (AVL Tree)

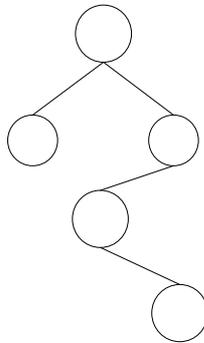
เนื่องจากวิทยานิพนธ์ฉบับนี้ จะต้องเลือกโครงสร้างในการจัดเก็บคำศัพท์ในพจนานุกรมข้อความพิเศษและคำหยาบที่ตรงกับนิยามที่ได้กำหนดไว้ เพื่อนำมาใช้ในการตัดคำและค้นหาคำหยาบ ซึ่งต้องเป็นโครงสร้างที่ใช้เนื้อที่ในการจัดเก็บข้อมูลน้อย สามารถค้นหาคำศัพท์ใน

พจนานุกรมและคำหยาบได้อย่างรวดเร็ว สามารถเพิ่มหรือแก้ไขคำศัพท์ในพจนานุกรมข้อความพิเศษและคำหยาบได้อย่างสะดวกรวดเร็ว รวมทั้งมีความง่ายในการพัฒนาโปรแกรม ดังนั้นในงานวิจัยฉบับนี้ขอเลือกโครงสร้างแบบต้นไม้เอวีแอล เพื่อเป็นโครงสร้างในการจัดเก็บคำศัพท์ในพจนานุกรม ข้อความพิเศษและคำหยาบ

ต้นไม้เอวีแอล เป็นต้นไม้ที่ประกอบด้วยคุณสมบัติสองข้อคือ ลูกทางซ้ายจะต้องมีค่าน้อยกว่าแม่และลูกทางขวาจะต้องมีค่ามากกว่าแม่ และความสูงของแต่ละต้นไม้ย่อย (Sub Tree) จะต้องต่างกันไม่เกินหนึ่ง ซึ่งความสูงของต้นไม้มีความสัมพันธ์ต่อการค้นหาข้อมูลในต้นไม้อย่างมาก ดังนั้นการพยายามค้นหาวิธีการที่ปรับให้ต้นไม้รักษาความสูงให้ดีที่สุด ได้ถูกค้นคว้าโดยนักวิทยาศาสตร์ชาวรัสเซียสองคน คือ Adel'son Velskii และ Landis ซึ่งชื่อต้นไม้ที่ตั้งเพื่อเป็นเกียรติกับผู้ค้นพบ และได้พิสูจน์ว่าความสูงของต้นไม้เอวีแอลจะไม่เกิน $1.44 \log_2 n$ ดังนั้นประสิทธิภาพของเวลาในการค้นหาข้อมูลในต้นไม้คือ $O(\log_2 n)$ โดยที่ n คือจำนวนข้อมูลทั้งหมด (Weiss, 1995) และในการตรวจสอบคุณสมบัติของต้นไม้ AVL จะใช้ค่า Balance Factor โดยสามารถตรวจสอบได้ดังนี้

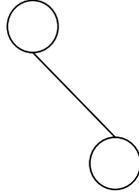
$$\text{Balance Factor} = \text{ความสูงของต้นไม้ย่อยทางซ้าย} - \text{ความสูงของต้นไม้ย่อยทางขวา}$$

โดยค่า Balance Factor ของแต่ละโหนดในต้นไม้ต้องมีค่าเท่ากับลบหนึ่ง ศูนย์ หรือหนึ่งเท่านั้น จึงจะเป็นต้นไม้เอวีแอล และถ้าในระหว่างการใช้งานซึ่งอาจจะเป็นการเพิ่มหรือลบโหนดแล้วทำให้ค่าของ Balance Factor ไม่เป็นไปตามค่านิยามของ Balance Factor ก็จะต้องทำการปรับต้นไม้ใหม่โดยวิธีการหมุน (Rotation) ซึ่งแบ่งได้สามแบบคือ หมุนครั้งเดียวไปทางซ้าย หมุนครั้งเดียวไปทางขวา และหมุนสองครั้ง (Double Rotation)



ภาพที่ 3.2 แสดงค่า Balance Factor ของแต่ละโหนดในต้นไม้

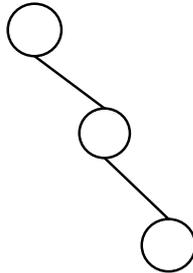
หลักในการหมุนครั้งเดียวจะง่ายกว่าหมุนสองครั้ง เช่น เดิมมีต้นไม้ดังภาพที่ 3.3 ซึ่งเป็นต้นไม้เอวีแอล



ภาพที่ 3.3 แสดงต้นไม้เอวีแอล

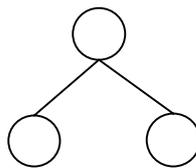
เมื่อเพิ่มโหนด C ในต้นไม้ดังภาพที่ 3.4 ก็จะทำให้ต้นไม้ขาดคุณสมบัติของต้นไม้เอวีแอล เนื่องจากค่า Balance Factor ของโหนด A มีค่าเท่ากับลบสอง

A



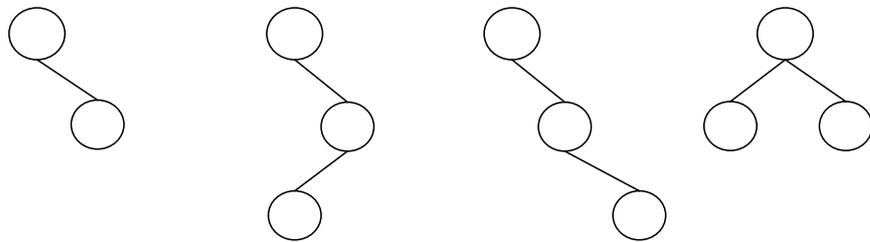
ภาพที่ 3.4 แสดงการเพิ่มโหนดในต้นไม้เอวีแอลในภาพที่ 3.4 ทำให้ต้นไม้ไม่เป็นต้นไม้เอวีแอล

จากภาพที่ 3.4 จะต้องทำการปรับต้นไม้ให้เป็นต้นไม้เอวีแอล โดยการหมุนครั้งเดียวทางซ้าย ดังภาพที่ 3.5



ภาพที่ 3.5 แสดงการปรับต้นไม้ให้เป็นต้นไม้เอวีแอล โดยการหมุนครั้งเดียวทางซ้าย

การหมุนซ้ายหรือหมุนขวาเมื่อส่วนที่ไม่เข้าคุณสมบัติของต้นไม้เอวีแอลนั้น ถ้าต้นไม้อยู่ในลักษณะเฉียงไปทางใดทางหนึ่ง การหมุนจะสามารถปรับให้เป็นต้นไม้เอวีแอลได้ ส่วนในกรณีภาพที่ 3.6 a) ที่เป็นต้นไม้เอวีแอล เมื่อเพิ่มโหนด B ซึ่งมีค่าน้อยกว่าโหนด C จึงทำให้ไม่เป็นต้นไม้เอวีแอล เนื่องจากค่า Balance Factor ของโหนด A มีค่าเท่ากับลบสอง ดังภาพที่ 3.6 b) และจะใช้วิธีหมุนซ้ายหรือขวาเพียงครั้งเดียวไม่ได้ ซึ่งจะต้องใช้วิธีการหมุนสองครั้ง คือครั้งแรกจะต้องหมุนเฉพาะโหนด C และ B ไปทางขวาก่อน ดังภาพที่ 3.6 c) จากนั้นให้ใช้วิธีหมุนครั้งเดียวทางซ้าย จะได้ดังภาพที่ 3.6 d)



ภาพที่ 3.6 แสดงการปรับต้นไม้ให้เป็นต้นไม้เอวีแอลโดยวิธีการหมุนสองครั้ง โดยเริ่มจากภาพที่ 3.6 a) 3.6 b) 3.6 c) และ 3.6 d) ตามลำดับ

3.2.4 การทดสอบประสิทธิภาพ ในวิทยานิพนธ์ฉบับนี้ มีความจำเป็นต้องทำการทดสอบประสิทธิภาพของโปรแกรม เพื่อทำการสรุปผลว่ามีประสิทธิภาพมากน้อยเพียงใด โดยในการทดสอบนั้น จะเน้นที่โปรแกรมสามารถแทนที่คำหยาบจริงด้วยเครื่องหมายดอกจันได้มากน้อยแค่ไหนเป็นสำคัญ โดยจะขอใช้คอนฟิวชันเมตริก (Confusion Matrix) ในการวิเคราะห์ประสิทธิภาพ ความถูกต้องของโปรแกรมดังนี้ คอนฟิวชันเมตริก เป็นการสร้างตารางสองทางระหว่างค่าที่เกิดขึ้นจริงกับค่าที่พยากรณ์จากตัวแบบ ณ จุดคัตออฟ (Cutoff) ต่างๆ (วิชิต หล่อจิระชุนห์กุล และจิราวัลย์ จิตรถเวช, 2547: 6) ดังในตารางที่ 3.3

A -1

A -2

C 0

C +1

B 0

a)

b)

ตารางที่ 3.3 แสดงคอนฟิชั่นเมตริก ในการทดสอบความถูกต้องของตัวแบบ

ค่าเป้าหมายที่เกิดขึ้นจริง	ค่าเป้าหมายจากสมการพยากรณ์		
	0	1	รวม
0	True Neg. A_1	False Pos. A_2	$A_1 + A_2$
1	False Neg. B_1	True Pos. B_2	$B_1 + B_2$
รวม	$A_1 + B_1$	$A_2 + B_2$	$A_1 + A_2 + B_1 + B_2$

แหล่งที่มา: วิจิต หล่อจิระชุนห์กุล และจิราวัลย์ จิตรถเวช, 2547: 6.

โดยจากตารางคอนฟิชั่นเมตริก นั้นจะใช้คำนวณค่าต่างๆ ดังต่อไปนี้

ความอ่อนไหว (Sensitivity) เป็นร้อยละที่สมการพยากรณ์สามารถจำแนกเหตุการณ์ที่สนใจได้ถูกต้องจากเหตุการณ์ที่สนใจทั้งหมด ($100 * B_2 / (B_1 + B_2)$)

ความเฉพาะเจาะจง (Specificity) เป็นร้อยละที่สมการพยากรณ์สามารถจำแนกเหตุการณ์ที่ไม่สนใจได้ถูกต้องจากเหตุการณ์ที่ไม่สนใจทั้งหมด ($100 * A_1 / (A_1 + A_2)$)

ความแม่นยำ (Accuracy) เป็นร้อยละที่สมการพยากรณ์สามารถจำแนกเหตุการณ์ที่สนใจและไม่สนใจได้ถูกต้องจากเหตุการณ์ทั้งหมด ($100 * (A_1 + B_2) / (A_1 + A_2 + B_1 + B_2)$)

ฟอลส์โพซิทีฟ (False Positive) เป็นร้อยละที่สมการพยากรณ์จำแนกเหตุการณ์ที่ไม่สนใจเป็นเหตุการณ์ที่สนใจจากเหตุการณ์ที่สมการพยากรณ์จำแนกว่าสนใจทั้งหมด ($100 * A_2 / (A_2 + B_2)$)

ฟอลส์เนกทีฟ (False Negative) เป็นร้อยละที่สมการพยากรณ์จำแนกเหตุการณ์ที่สนใจเป็นเหตุการณ์ที่ไม่สนใจจากเหตุการณ์ที่สมการพยากรณ์จำแนกว่าไม่สนใจทั้งหมด ($100 * B_1 / (A_1 + B_1)$)

(วิจิต หล่อจิระชุนห์กุล และจิราวัลย์ จิตรถเวช, 2547: 6) สรุปในบทนี้ ได้กำหนดนิยามของคำคำหยาบ คำปกติและข้อความ ส่วนทฤษฎีที่ใช้ในการวิจัยคือ การเทียบข้อความและเทียบคำที่ยาวที่สุดในการตัดคำ ไฟไนท์สเตทแมชชีนด้วยข้อมูลออก โครงสร้างข้อมูลแบบต้นไม้เอวีแอลในการจัดเก็บคำศัพท์ในพจนานุกรมข้อความพิเศษและคำหยาบ และคอนฟิชั่นเมตริกในการทดสอบประสิทธิภาพ ในบทต่อไปจะกล่าวถึงขั้นตอนการดำเนินการวิจัยในแต่ละขั้นตอนว่ามีความสัมพันธ์กับทฤษฎีที่นำมาใช้อย่างไร

บทที่ 4

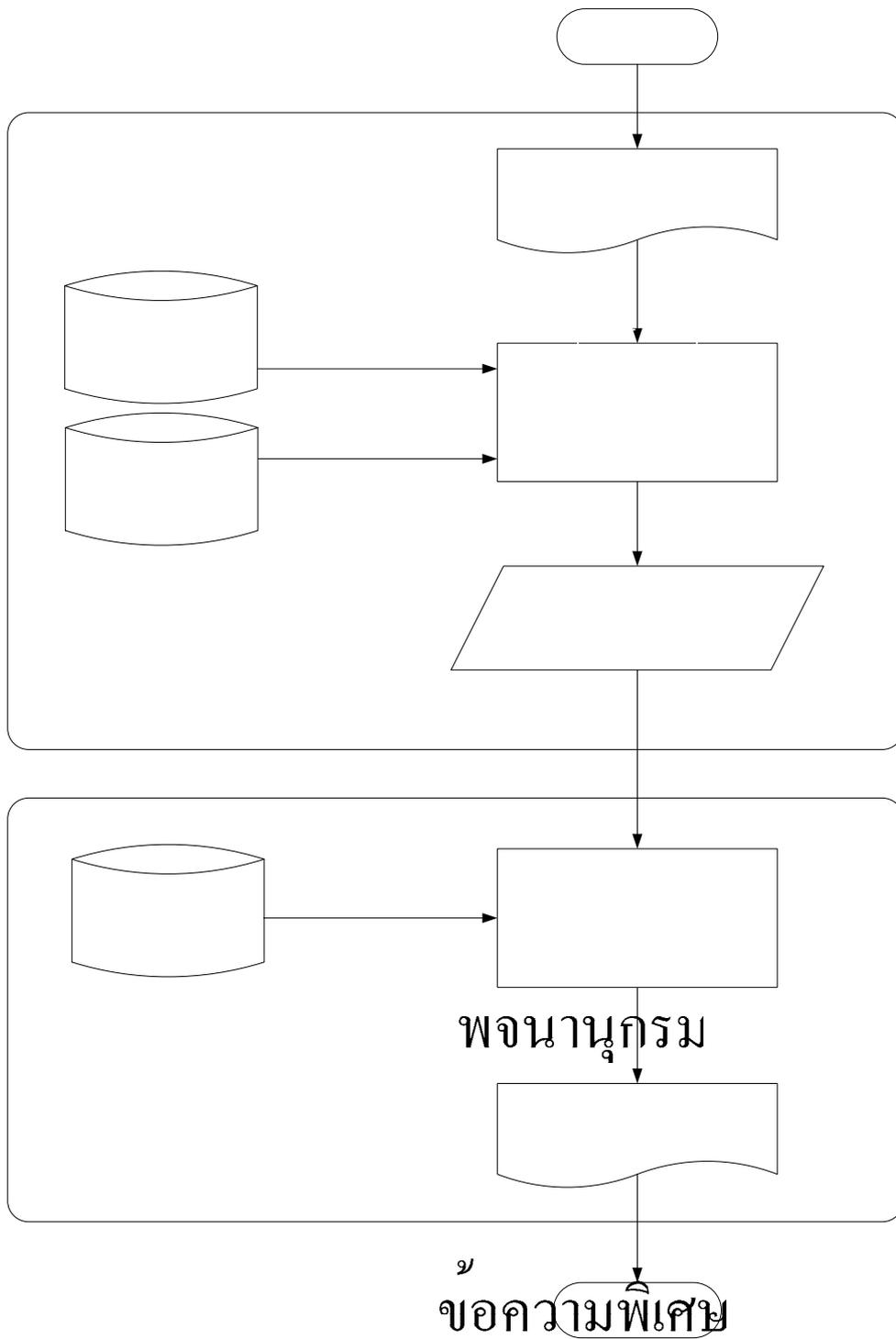
ขั้นตอนการวิจัย

การใช้ภาษาไทยในการสื่อสารปัจจุบันนี้ ยังขาดวัฒนธรรม จริยธรรม และจิตสำนึกในการใช้ โดยจะเห็นได้จากการใช้คำหยาบในสื่อต่างๆ เช่น หนังสือพิมพ์ และอินเทอร์เน็ต เป็นต้น หากไม่รีบแก้ไขและป้องกันจะทำให้ส่งผลเสียต่อภาษาไทยซึ่งเป็นเอกลักษณ์ประจำชาติไทยได้ ซึ่งหากเราสามารถสร้างเครื่องมือที่มีประสิทธิภาพช่วยในการตรวจสอบคำหยาบได้นั้น จะทำให้ช่วยป้องกันปัญหาดังกล่าวนี้ได้

ดังนั้นในงานวิจัยฉบับนี้ จึงมีความสนใจพัฒนาโปรแกรมค้นหาและแทนที่คำหยาบในข้อความภาษาไทยด้วยเครื่องหมายดอกจันเพื่อนำไปใช้ประมวลผลในเครื่องคอมพิวเตอร์ เนื่องจากปัจจุบันคอมพิวเตอร์มีบทบาทสำคัญในการติดต่อสื่อสาร สามารถรับส่งข่าวสารผ่านทางเครือข่ายอินเทอร์เน็ตและมีความรวดเร็วในการประมวลผล โดยได้แบ่งขั้นตอนการดำเนินงานออกเป็นสามขั้นตอนคือ การประมวลผล การเตรียมเอกสาร และการทดสอบประสิทธิภาพ ซึ่งมีรายละเอียดดังนี้

4.1 ขั้นตอนการประมวลผล

ขั้นตอนการประมวลผลนั้นมีอยู่สองขั้นตอนใหญ่ๆคือ การตัดคำ และการค้นหาและแทนที่คำหยาบในข้อความภาษาไทยด้วยเครื่องหมายดอกจัน โดยสามารถเขียนเป็นแผนภาพแสดงการทำงานของระบบได้ดังนี้



ภาพที่ 4.1 แสดงการทำงานของระบบการกรอกคำหยาบในข้อความภาษาไทย

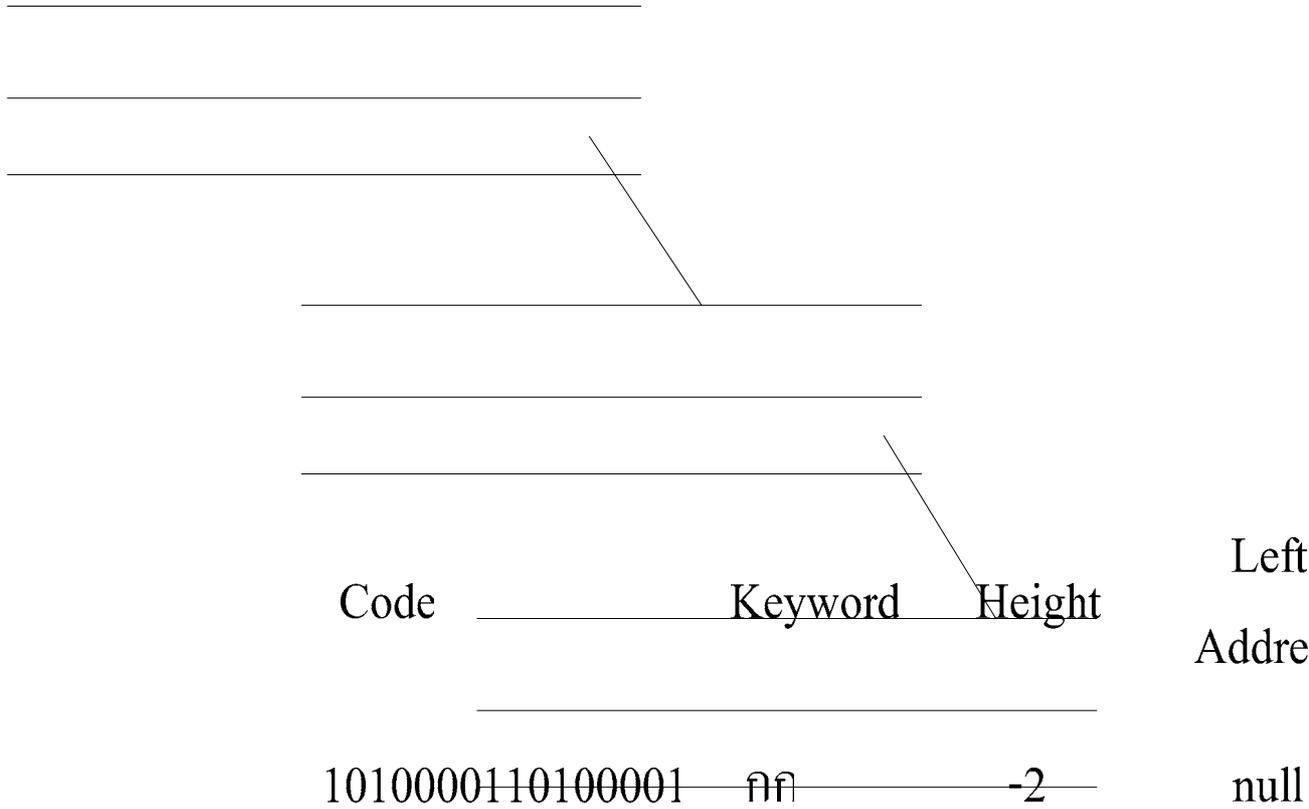
จากภาพที่ 4.1 เป็นการอธิบายวิธีการทำงานของระบบการกรองคำหยาบในข้อความภาษาไทย ซึ่งมีขั้นตอนทำงานหลักๆแบ่งเป็นสองส่วนด้วยกัน ซึ่งในแต่ละส่วนนี้จะมีกระบวนการทำงานและรายละเอียดที่แตกต่างกันไป สามารถอธิบายได้ดังนี้

4.1.1 การตัดคำ (Word Segmentation) ขั้นตอนนี้จะแบ่งการทำงานออกเป็นสองขั้นตอนดังนี้

4.1.1.1 การสร้างโครงสร้างต้นไม้เอวีแอล (AVL Tree) โครงสร้างต้นไม้เอวีแอลถูกใช้เพื่อจัดเก็บคำศัพท์ในพจนานุกรมให้มีความสะดวกและรวดเร็วในค้นหาคำศัพท์ และข้อความพิเศษในการตัดคำ รวมทั้งค้นหาคำหยาบในการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน ต้นไม้เอวีแอล เป็นต้นไม้ที่แต่ละโหนดมีความสูงของแต่ละต้นไม้ย่อย (Sub Tree) ต่างกันไม่เกินหนึ่ง ซึ่งความสูงของต้นไม้มีความสัมพันธ์ต่อการค้นหาข้อมูลในต้นไม้อย่างมาก ดังนั้นการพยายามค้นหาวิธีการที่ปรับให้ต้นไม้รักษาความสูงให้ดีที่สุด โดยประกอบด้วยคุณสมบัติสองข้อคือ ข้อหนึ่ง ลูกทางซ้ายจะต้องมีค่าน้อยกว่าแม่และลูกทางขวาจะต้องมีค่ามากกว่าแม่ และข้อสอง ค่าความสูงของต้นไม้ย่อยทางซ้ายลบด้วยค่าความสูงของต้นไม้ย่อยทางขวาต้องมีค่าเท่ากับลบหนึ่ง ศูนย์ หรือหนึ่ง เท่านั้น (Weiss, 1995) ในขั้นตอนนี้จะทำการสร้างโครงสร้างต้นไม้เอวีแอลสองแบบดังนี้

1) การสร้างโครงสร้างคำศัพท์และคำหยาบโดยต้นไม้เอวีแอล หลักการในการสร้างคือ ทำการกำหนดรหัสเลขฐานสองให้กับตัวอักษรในภาษาไทยแต่ละตัว เช่น “ก” = 10100001, “ข” = 10100010, “ฃ” = 10100011, “ค” = 10100100, ..., “จ” = 11101101 และอักขระอื่นเป็น 11111111 ซึ่งจะแสดงในรหัสต้นฉบับ (Source Code) ในภาคผนวก ก โดยการสร้างต้นไม้จะนำเอาโครงสร้างข้อมูลแบบลิงคีสต์ (Link List) มาใช้ในการพัฒนาโครงสร้างของต้นไม้ ซึ่งโหนดแต่ละโหนดจะเก็บข้อมูลทั้งหมดห้าชนิดคือ ข้อมูลรหัสเลขฐานสองของคำศัพท์ เช่นคำว่า “กก” = 1010000110100001 ข้อมูลคำศัพท์ ค่า Balance Factor ซึ่งได้จากค่าความสูงของต้นไม้ย่อยทางซ้ายลบด้วยค่าความสูงของต้นไม้ย่อยทางขวา และข้อมูลที่อยู่ของต้นไม้ที่เป็นลูกทางซ้ายและขวา โดยระหว่างที่มีการเพิ่มข้อมูลคำศัพท์ในต้นไม้จะทำการตรวจสอบคุณสมบัติของต้นไม้เสมอ หากมีการเพิ่มข้อมูลแล้วทำให้ค่า Balance Factor มีค่าไม่เท่ากับ ลบหนึ่ง ศูนย์ หรือหนึ่ง ซึ่งทำให้ขาดคุณสมบัติของต้นไม้เอวีแอล จะต้องทำการปรับต้นไม้ใหม่โดยวิธีการหมุน (Rotation) ซึ่งแบ่งได้สามแบบคือ หมุนครั้งเดียวไปทางซ้าย หมุนครั้งเดียวไปทางขวาและหมุนสองครั้ง (Double Rotation) โดยขั้นตอนของการสร้างต้นไม้เอวีแอลสามารถแสดงได้ดังนี้

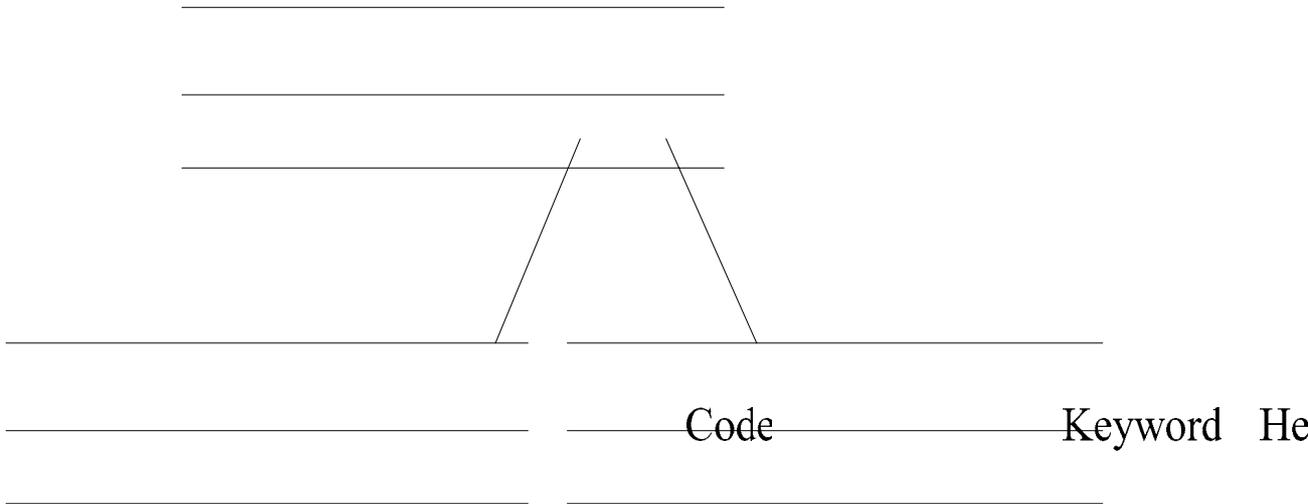
สมมติมีคำในพจนานุกรม คือ คำว่า “กก” “กระบอก” และ “กระรอก” จะสร้างต้นไม้ได้
เป็นดังภาพที่ 4.2



ภาพที่ 4.2 แสดงการสร้างโครงสร้างของต้นไม้เอวีแอลจากคำว่า “กก” “กระบอก” และ “กระรอก”

จากภาพที่ 4.2 เมื่อเพิ่มคำว่า “กระรอก” เข้าไปแล้วจะทำให้ค่าความสูงของต้นไม้ย่อยทางซ้ายลดด้วยค่าความสูงของต้นไม้ย่อยทางขวามีค่าไม่เท่ากับลบหนึ่ง ศูนย์ และหนังสือจึงจำเป็นต้องปรับโดยการหมุนครั้งเดียวทางซ้าย จะได้ต้นไม้เอวีแอลดังภาพต่อไปนี้

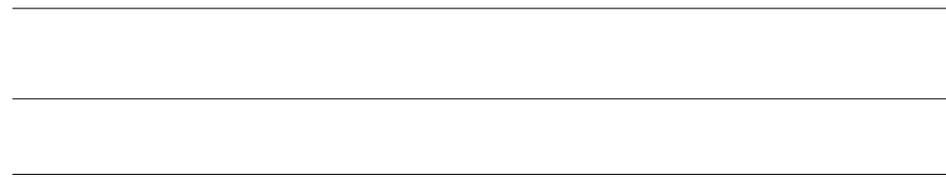
1010000111000011...



ภาพที่ 4.3 แสดงการปรับต้นไม้ที่ไม่เป็นคุณสมบัติของต้นไม้เอวีแอลจากภาพที่ 4.2 ให้เป็นต้นไม้เอวีแอลโดยการหมุนครั้งเดียวทางซ้าย

2) การสร้างโครงสร้างข้อความพิเศษโดยต้นไม้เอวีแอล เนื่องจากในการวิจัยช่วงแรกๆ ได้เลือกวิธีการตัดคำแบบเทียบคำที่ยาวที่สุด ซึ่งในขั้นตอนการพัฒนาโปรแกรมตัดคำนั้น ได้มีการทดลองประมวลผลโปรแกรม ซึ่งจากการทดสอบโปรแกรมผู้วิจัยได้สังเกตเห็นผลลัพธ์จากการตัดคำของบางข้อความที่ไม่ถูกต้อง ทั้งๆที่ข้อความนั้นเป็นข้อความที่เกิดจากการประสมของคำที่ปรากฏในพจนานุกรม ทั้งนี้เนื่องมาจากการตัดคำแบบเทียบคำที่ยาวที่สุดนั้น จะทำการเลือกคำที่มีจำนวนอักขระมากที่สุดเป็นอันดับแรกในการประมวลผลแต่ละครั้ง เช่นข้อความ “ต้องตาย” จะได้ผลลัพธ์เป็น “ต้อง ตาย” ทั้งๆที่ควรจะเป็น “ต้อง ตาย” แต่เนื่องจากคำว่า “ต้องตา” มีอักขระมากกว่าคำว่า “ต้อง” จึงทำให้คำว่า “ต้องตา” ถูกเลือกมาเป็นผลลัพธ์ ซึ่งไม่ถูกต้อง ดังนั้นจากปัญหาดังกล่าวทำให้ผู้วิจัยได้คิดหาแนวทางในการแก้ไขปัญหานี้ โดยการสร้างคลังข้อความพิเศษ ซึ่งจะอยู่ในรูปของเอกสารข้อความพิเศษ ทั้งนี้ก็เพื่อให้่ายในการเพิ่มและแก้ไขข้อความ โดยจะทำการจัดเก็บข้อมูลสองส่วนคือ ข้อความที่เกิดจากการประสมของคำที่ปรากฏในพจนานุกรม แต่ถ้านำข้อความนั้นไปตัดคำแบบเทียบคำที่ยาวที่สุดแล้วจะทำให้ได้ผลลัพธ์ที่ไม่ถูกต้อง และผลลัพธ์ของการตัดคำที่มีความเป็นไปได้มากที่สุดของข้อความนั้นจำนวนหนึ่งผลลัพธ์ เช่นข้อความ “หากระดาษ” ผลลัพธ์คือ “หา กระดาษ” ข้อความ “ผลกระทบ” ผลลัพธ์คือ “ผล กระทบ” ซึ่งจะแสดงอยู่ในภาคผนวก จ อีกอย่างหนึ่งในเอกสารข้อความพิเศษนั้นยังมีความยืดหยุ่นและง่ายในการเพิ่ม ลบและแก้ไขข้อความ โดยไม่ส่งผลกระทบต่อโปรแกรมที่พัฒนาขึ้น ถึงแม้ในอนาคตข้างหน้าผู้ใช้อาจจะ

พบข้อความที่ทำให้การตัดคำไม่ถูกต้อง ก็สามารถที่จะทำการเพิ่มเติมได้โดยง่าย จากนั้นก็จะนำข้อมูลในเอกสารพิเศษมาทำการสร้างเป็นโครงสร้างต้นไม้เอวีแอล เพื่อเพิ่มประสิทธิภาพในการค้นหาในการเทียบข้อความให้มีความรวดเร็วมากยิ่งขึ้น หลักการในการสร้างคือ ทำการกำหนดรหัสเลขฐานสองให้กับตัวอักษรในภาษาไทยแต่ละตัว เช่น “ก” = 10100001 , “ข” = 10100010 , “ฅ” = 10100011 , “ค” = 10100100 ,..., “จ” = 11101101 และอักขระอื่นเป็น 11111111 ซึ่งจะแสดงในรหัสต้นฉบับ (Source Code) ในภาคผนวก ก โดยการสร้างต้นไม้จะนำเอาโครงสร้างข้อมูลแบบลิงคัลลิสต์ (Link List) มาใช้ในการพัฒนาโครงสร้างของต้นไม้ ซึ่งโหนดแต่ละโหนดจะเก็บข้อมูลทั้งหมดทุกชนิด คือข้อมูลรหัสเลขฐานสองของข้อความพิเศษ ข้อความพิเศษ ผลลัพธ์ที่มีความเป็นไปได้ในการตัดคำของข้อความพิเศษนั้น ค่า Balance Factor และข้อมูลที่อยู่ของต้นไม้ที่เป็นลูกทางซ้ายและขวา ยกตัวอย่างเช่นข้อความ “หากระดาช” ซึ่งมีผลลัพธ์เป็น “หา กระดาช” สามารถสร้างเป็นโหนดของต้นไม้เอวีแอลได้ดังรูปต่อไปนี้



ภาพที่ 4.4 แสดงโหนดของต้นไม้เอวีแอลของข้อความเอกสารพิเศษ

4.1.1.2 การตัดคำแบบเทียบข้อความ (String Matching) วิธีการที่ใช้ในการตัดคำนั้นเป็นวิธีการแบบการเทียบข้อความ (String Matching) ซึ่งเป็นวิธีการที่ได้ปรับปรุงเพิ่มเติมเพื่อให้การตัดคำมีความถูกต้องมากยิ่งขึ้น โดยจะเก็บข้อความที่เกิดจากการประสมกันระหว่างคำที่ปรากฏในพจนานุกรมตั้งแต่สองคำขึ้นไป ซึ่งเป็นข้อความที่ตัดคำแบบวิธีเทียบคำที่ยาวที่สุดแล้วได้ผลลัพธ์ไม่ถูกต้อง โดยได้พิจารณาจากข้อความในเอกสารที่เก็บรวบรวมได้แล้วนำไปทำการทดสอบตัดคำด้วยวิธีการแบบเทียบคำที่ยาวที่สุด ถ้าได้ผลลัพธ์ไม่ถูกต้องก็จะทำการเก็บข้อความนั้นไว้ในเอกสารข้อความพิเศษ ยกตัวอย่างได้จากข้อความ “เขาหากระดาช” โดยข้อความนี้ถ้าตัดคำแบบเทียบคำที่ยาวที่สุดแล้วจะได้ผลลัพธ์เป็น “เขา / หาก / ระดา / ช” ซึ่งไม่ถูกต้อง ซึ่งแนวทางในการแก้ไขปัญหานี้คือใช้วิธีการย้อนรอย (ยื่น กุ๊ววรรณ และวิวรรธ อัมอรณม, 2529) แต่จากการศึกษาการตัดคำโดยวิธีการย้อนรอนั้นจะใช้เวลาในการประมวลผลค่อนข้างมาก ดังนั้นผู้วิจัยจึงขอเสนอวิธีการตัด

คำแบบเทียบข้อความ ซึ่งเป็นวิธีที่ใช้เวลาในการประมวลผลน้อย โดยผู้วิจัยได้ทำการปรับปรุงเพื่อแก้ไขปัญหาการตัดคำแบบวิธีเทียบคำที่ยาวที่สุดแล้วได้ผลลัพธ์ไม่ถูกต้อง

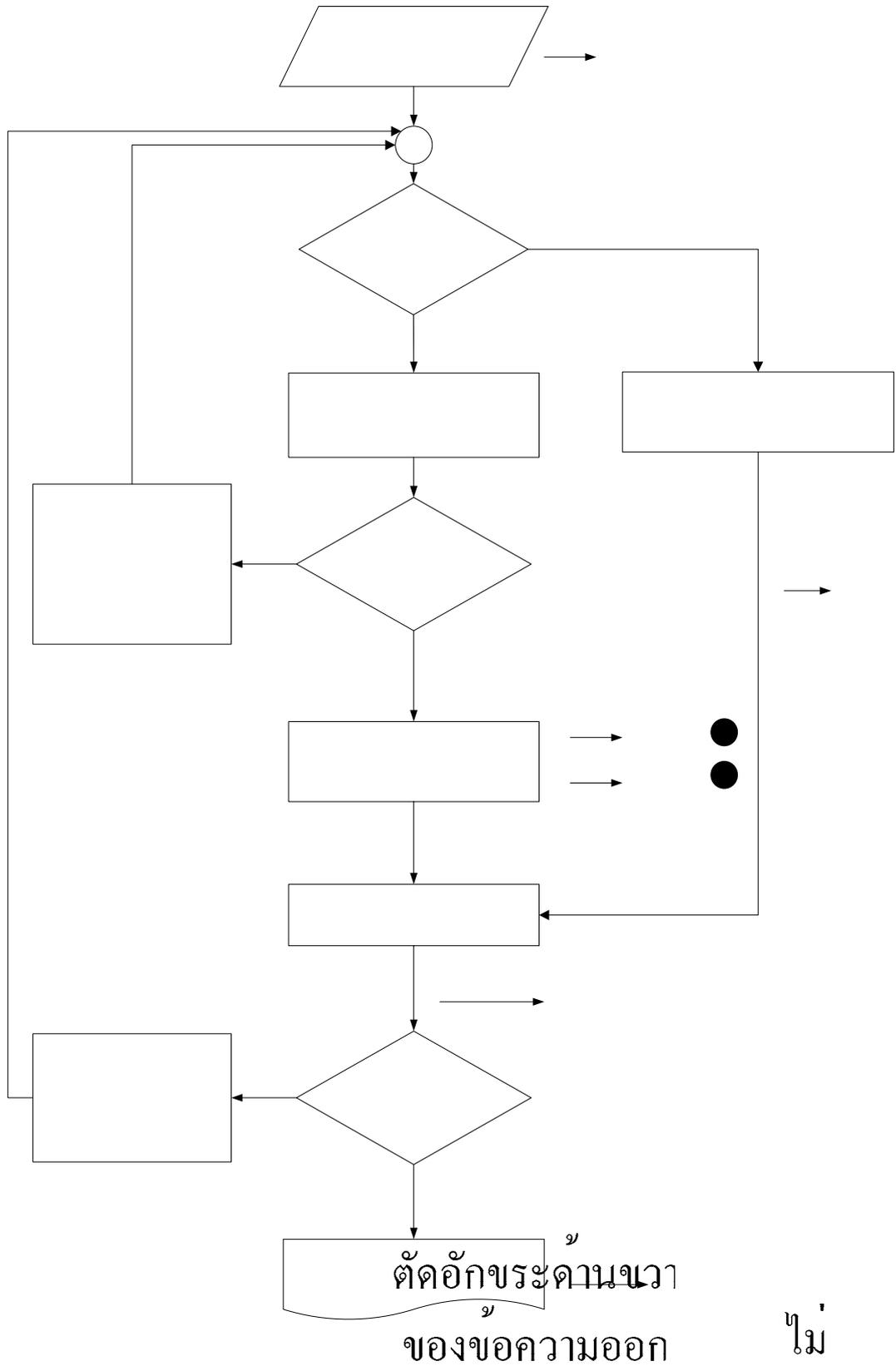
หลักการตัดคำโดยวิธีเทียบข้อความคือ ทำการเก็บข้อความพิเศษในเอกสารก่อนโดยเอกสารพิเศษจะมีข้อมูลสองอย่างคือ ข้อความพิเศษและผลลัพธ์การตัดคำที่ควรจะเป็นไปได้มากที่สุดของข้อความพิเศษนั้น ยกตัวอย่างเช่นข้อความ “หากระดาษ” ในเอกสารพิเศษก็จะทำการจัดเก็บข้อความ “หากระดาษ” แล้วตามด้วยผลลัพธ์คือ “หา / กระดาษ” ถ้าดับต่อไปก็จะนำเอกสารพิเศษมาใช้ในการตัดคำแบบเทียบข้อความ ถ้าเทียบกันได้ก็จะนำผลลัพธ์ที่ได้จัดเก็บไว้มาเป็นผลลัพธ์ในการตัดคำ แต่ถ้าไม่สามารถเทียบได้ก็จะทำการตัดคำโดยวิธีเทียบคำที่ยาวที่สุดแทนสามารถแสดงตัวอย่างการตัดแบบเทียบข้อความได้จากข้อความ “เขาหากระดาษ” ซึ่งจะทำการเทียบข้อความดังกล่าวก่อน ในที่นี้จะเทียบไม่ได้ เนื่องจากไม่ได้ทำการจัดเก็บข้อความนี้ไว้ในเอกสารพิเศษ จึงต้องทำการตัดคำแบบเทียบคำที่ยาวที่สุดก่อน ซึ่งจะได้ผลลัพธ์ในการตัดคำแบบเทียบคำที่ยาวที่สุดเป็นคำว่า “เขา /” แล้วก็ให้นำอักขระที่เหลือไปทำการตัดคำโดยเทียบข้อความอีกครั้ง ซึ่งในที่นี้อักขระที่เหลือคือ “หากระดาษ” และเมื่อทำการตัดคำแบบเทียบข้อความก็จะได้ผลลัพธ์เป็น “หา / กระดาษ” แล้วนำผลลัพธ์ที่ได้มาต่อกัน ดังนั้นผลลัพธ์ที่ได้ในการตัดคำจากข้อความ “เขาหากระดาษ” คือ “เขา / หา / กระดาษ”

4.1.1.3 การตัดคำแบบเทียบคำที่ยาวที่สุด (Longest Matching) เนื่องจากการตัดคำด้วยวิธีการเทียบข้อความนั้น มีความยากลำบากในการจัดเก็บข้อความทั้งหมดที่สามารถเกิดขึ้นได้ รวมทั้งยังทำให้สิ้นเปลืองเนื้อที่ในการจัดเก็บ ดังนั้นจากปัญหาดังกล่าวจึงทำการจัดเก็บเฉพาะข้อความที่ตัดคำแบบวิธีเทียบคำที่ยาวที่สุดแล้วได้ผลลัพธ์ไม่ถูกต้องเท่านั้น นอกเหนือจากนี้แล้วจะใช้วิธีการแบบเทียบคำที่ยาวที่สุดจากพจนานุกรมในการตัดคำเป็นหลัก โดยจะทำการตรวจสอบคำ ซึ่งจะเริ่มจากหน่วยของข้อความที่สั้นที่สุดที่สามารถตรวจสอบได้ โดยจะตรวจสอบกับจุดสิ้นสุดของประโยค เช่น การขึ้นย่อหน้า การเว้นวรรคตอน เป็นต้น จากนั้นจะทำการลดความยาวของข้อความลงทีละตัวไปตามเกณฑ์ทางอักขรวิธีที่ใช้พิจารณาความเป็นไปได้ในการแยกสายอักขระออกเพื่อการพิจารณาคำ สามารถแสดงตัวอย่างได้จากข้อความ “เขาไปตลาด” โดยจะเริ่มเทียบข้อความนี้กับพจนานุกรมก่อน ถ้าข้อความนี้ไม่สามารถเทียบคำกับพจนานุกรมได้ก็จะลดลงเหลือ “เขาไปตลาด” ถ้าไม่สามารถเทียบได้อีก ก็จะลดอีก “เขาไปตลาด” ลดไปเรื่อยๆ จนได้คำว่า “เขา” ซึ่งสามารถเทียบคำในพจนานุกรมได้จึงตัดเป็นคำแรกของข้อความ และทำเครื่องหมายไว้เป็นจุดย้อนกลับ โดยสามารถแสดงได้ดังตารางต่อไปนี้

ข้อความ	ผลลัพธ์ในการตัดคำ	ข้อความที่เหลือ
เขาไปตลาด	-	-
เขาไปตลาด	-	ค
เขาไปตลาด	-	ตลาด
เขาไปตลาด	-	ตลาด
เขาไปตลาด	เขา	ไปตลาด (ทำจุดย้อนกลับ)
เขาไปตลาด	-	-
เขาไปตลาด	-	ค
เขาไปตลาด	-	ตลาด
เขาไปตลาด	ไป	ตลาด (ทำจุดย้อนกลับ)
เขาไปตลาด	ตลาด	-

ตารางที่ 4.1 แสดงผลการตัดคำแบบเทียบคำที่ยาวที่สุดของข้อความ “เขาไปตลาด”

จากสองวิธีที่ได้กล่าวมานี้ สามารถแสดงเป็นแผนผังการประมวลผลการตัดคำโดยวิธีเทียบข้อความและเทียบคำที่ยาวที่สุดได้ดังภาพที่ 4.5



ขอความ



เทียบ
ได้

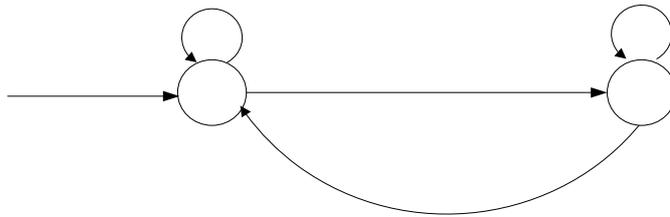
นำขอความ
วิธีเทียบ

เทียบ

ภาพที่ 4.5 แสดงการตัดคำโดยวิธีเทียบข้อความและเทียบคำที่ขาด
และเก็บอีกประที่ตัดออก
ไว้ประมวลผลต่อไป

4.1.2 การกรองคำหยาบในข้อความภาษาไทย

ในขั้นตอนนี้จะดำเนินการต่อจากขั้นการตัดคำ โดยจะอ่านคำที่ตัดได้จากขั้นตอนการตัดคำ เป็นข้อมูลนำเข้าในการประมวลผลทำการตรวจสอบว่าคำที่อ่านเข้ามาเป็นคำหยาบหรือไม่ซึ่งสามารถเทียบได้จากนิยามที่ได้กล่าวไปแล้วในบทที่ 3 ถ้าเทียบแล้วเป็นคำหยาบก็จะทำการแทนที่คำหยาบนั้นด้วยเครื่องหมายดอกจัน แต่ถ้าไม่เป็นก็จะได้ผลลัพธ์ออกมาเป็นคำนั้นคำเดิม โดยจะทำการอ่านคำเข้ามาทีละคำ ตั้งแต่คำแรกจากซ้ายไปขวาจนกว่าจะหมดทุกคำจึงหยุดการทำงาน เพื่อให้ไม่ให้คำหยาบปรากฏในข้อความภาษาไทย ดังนั้นขอเลือกตัวแบบทางคณิตศาสตร์ไฟไนท์สเตตแมชชีนด้วยข้อมูลออกเป็นแนวทางในพัฒนาโปรแกรม และอธิบายการค้นหาและแทนที่คำหยาบด้วยเครื่องหมายดอกจัน โดยสามารถอธิบายการทำงานเป็นไดอะแกรมสถานะ ดังต่อไปนี้



ภาพที่ 4.6 แสดงไดอะแกรมสถานะในการอธิบายการค้นหาและแทนที่คำหยาบในข้อความภาษาไทยด้วยเครื่องหมายดอกจัน

จากภาพที่ 4.6 แสดงไฟไนท์สเตตแมชชีนการค้นหาและแทนที่คำหยาบด้วยเครื่องหมายดอกจันประกอบไปด้วย สถานะทั้งหมดสองสถานะ คือ s_0 และ s_1 ข้อมูลนำเข้าคือ คำปกติและคำหยาบ ข้อมูลออกคือ คำปกติและเครื่องหมายดอกจัน (*) ซึ่งสามารถอธิบายไดอะแกรมสถานะได้คือ เริ่มต้นที่สถานะ s_0 ถ้าข้อมูลเข้าเป็น คำปกติ จะได้ข้อมูลออกเป็น คำปกติ และจะไม่เปลี่ยนสถานะ แต่ถ้าข้อมูลเข้าเป็น คำหยาบ จะได้ข้อมูลออกเป็น เครื่องหมายดอกจัน (*) และจะเปลี่ยนสถานะเป็นสถานะ s_1 และที่สถานะ s_1 ถ้าข้อมูลเข้าเป็น คำปกติ จะได้ข้อมูลออกเป็น เครื่องหมายดอกจัน (*) และจะไม่เปลี่ยนสถานะ แต่ถ้าข้อมูลเข้าเป็น คำปกติ จะได้ข้อมูลออกเป็น คำปกติ และจะเปลี่ยนสถานะเป็นสถานะ s_0 ยกตัวอย่างเช่นข้อความ “ไอ้หน้าตัวเมีย” จากโปรแกรมตัดคำได้ผลลัพธ์เป็น ไอ้ / หน้า / ตัวเมีย จากการประมวลผลโดยอ่านข้อมูลเข้าทีละคำจนหมด เริ่มจากสถานะ s_0 ข้อมูลเข้าคือคำว่า “ไอ้” ซึ่งเมื่อเทียบจากนิยามคำว่า “ไอ้” มีความหมายแสดงความดู

หมิ่นเหยียดหยาม (พจนานุกรมฉบับราชบัณฑิตสถาน พ.ศ. 2542) จึงถือว่า “ไอ้” เป็นคำหยาบ จากไคอะแกรมสถานะถ้าที่สถานะ s_0 ข้อมูลเข้าเป็น คำหยาบ จะได้ข้อมูลออกเป็น เครื่องหมายดอกจัน (*) และจะเปลี่ยนสถานะเป็นสถานะ s_1 จากสถานะ s_1 ก็อ่านคำต่อไปคือคำว่า “หน้า” มีความหมายว่า ส่วนของศีรษะตั้งหน้าผากลงมาจดคาง (พจนานุกรมฉบับราชบัณฑิตสถาน พ.ศ. 2542) เมื่อเทียบจากนิยามถือได้ว่าเป็นคำปกติ จึงได้ข้อมูลออกเป็นคำว่า “หน้า” เหมือนเดิม และจะเปลี่ยนสถานะเป็นสถานะ s_0 และจากสถานะ s_0 ก็อ่านคำต่อไปคือคำว่า “ตัวเมีย” มีความหมายว่า เพศเมีย (เฉพาะสัตว์หรือพืชบางชนิด) หรือเรียกสิ่งของบางอย่างที่มีรูหรือขอรับให้สิ่งอื่นเกาะหรือสอดเข้าได้ว่าตัวเมีย (พจนานุกรมฉบับราชบัณฑิตสถาน พ.ศ. 2542) และเมื่อเทียบจากนิยามถือได้ว่าเป็นคำปกติ จึงได้ข้อมูลออกเป็นคำว่า “ตัวเมีย” เหมือนเดิม และไม่เปลี่ยนสถานะ และจะสิ้นสุดการทำงานเนื่องไม่มีข้อมูลเข้าแล้ว ดังนั้นจากการทำงานครั้งนี้โดยมีข้อมูลเข้าเป็นข้อความ “ไอ้หน้าตัวเมีย” จะได้ผลลัพธ์เป็น “*หน้าตัวเมีย”

4.2 ขั้นตอนการเตรียมเอกสาร

เป็นขั้นตอนการเตรียมเอกสารที่นำมาทดสอบประสิทธิภาพ โดยเอกสารที่นำมาทดสอบนั้นจะเป็นเอกสารจากหนังสือพิมพ์และข้อความจากเว็บบอร์ดในสื่ออินเทอร์เน็ต โดยไม่มีรูปภาพ บทกลอน หรือสมการใดๆ ที่ผ่านการตรวจสอบแก้ไขคำต่างๆ ที่เขียนไม่ถูกต้อง ให้ถูกต้องเรียบร้อยแล้ว เพื่อนำมาทดสอบกับโปรแกรม โดยได้แบ่งชุดข้อมูลออกเป็นสองกลุ่ม คือ ชุดข้อมูลที่เป็นเอกสารประเภทข่าวหนังสือพิมพ์ ประกอบด้วยหนังสือพิมพ์ไทยรัฐ มติชน ข่าวสด และคมชัดลึก จำนวนหนึ่งร้อยคอลัมน์ และชุดข้อมูลจากข้อความแสดงความคิดเห็นบนเว็บบอร์ดต่างๆ ในสื่ออินเทอร์เน็ต จำนวนเจ็ดร้อยเจ็ดข้อคิดเห็น

4.3 ขั้นตอนการทดสอบประสิทธิภาพ

เป็นขั้นตอนการทดสอบความถูกต้องของโปรแกรมการตัดคำ และโปรแกรมค้นหาและแทนที่คำหยาบในข้อความภาษาไทยด้วยเครื่องหมายดอกจัน โดยการทดสอบจะแบ่งชุดข้อมูลในการทดสอบออกเป็นสองชุดตามที่ได้กล่าวถึงในขั้นตอนการเตรียมเอกสาร ซึ่งมีรายละเอียดขั้นตอนในการทดสอบดังต่อไปนี้

4.3.1 การทดสอบโปรแกรมตัดคำ

4.3.1.1 นำข้อมูลจากเอกสารชุดที่หนึ่งมาประมวลผลในโปรแกรมตัดคำ แล้วทำการนับจำนวนคำที่ตัดถูกต้องและที่ตัดไม่ถูกต้อง

4.3.1.2 คำนวณหาร้อยละความถูกต้อง จากการทดสอบการตัดคำของโปรแกรมที่ได้จากข้อมูลเอกสารในชุดที่หนึ่ง

4.3.1.3 นำข้อมูลจากเอกสารชุดที่สองมาประมวลผลในโปรแกรมตัดคำ แล้วทำการนับจำนวนคำที่ตัดถูกต้องและที่ตัดไม่ถูกต้อง

4.3.1.4 คำนวณหาร้อยละความถูกต้อง จากการทดสอบการตัดคำของโปรแกรมที่ได้จากข้อมูลเอกสารในชุดที่สอง

4.3.2 การทดสอบโปรแกรมค้นหาและแทนที่คำหยาบในข้อความภาษาไทยด้วยเครื่องหมายดอกจัน

4.3.2.1 ทำการนับคำหยาบโดยเทียบจากนิยามที่ได้กำหนดไว้ในบทที่ 3 โดยผู้ทำการทดสอบเองในข้อมูลเอกสารชุดที่หนึ่ง

4.3.2.2 ทำการนับคำปกติโดยเทียบจากนิยามที่ได้กำหนดไว้ในบทที่ 3 โดยผู้ทำการทดสอบเองในข้อมูลเอกสารชุดที่หนึ่ง

4.3.2.3 นำข้อมูลในเอกสารชุดที่หนึ่ง เข้าประมวลผลในโปรแกรมค้นหาและแทนที่คำหยาบในข้อความภาษาไทยด้วยเครื่องหมายดอกจัน และทำการนับจำนวนคำดังต่อไปนี้คือ จำนวนคำปกติจริงที่โปรแกรมบอกว่าเป็นคำปกติ จำนวนคำปกติจริงที่โปรแกรมบอกว่าเป็นคำหยาบ จำนวนคำหยาบจริงที่โปรแกรมบอกว่าเป็นคำหยาบ และจำนวนคำหยาบจริงที่โปรแกรมบอกว่าเป็นปกติ แล้วนำค่าที่นับได้ไปใส่ในตารางคอนฟิซชันเมตริก ซึ่งได้กล่าวไปแล้วในบทที่ 3

4.3.2.4 ทำการนับคำหยาบโดยเทียบจากนิยามที่ได้กำหนดไว้ในบทที่ 3 โดยผู้ทำการทดสอบเองในข้อมูลเอกสารชุดที่สอง

4.3.2.5 ทำการนับคำปกติโดยเทียบจากนิยามที่ได้กำหนดไว้ในบทที่ 3 โดยผู้ทำการทดสอบเองในข้อมูลเอกสารชุดที่สอง

4.3.2.6 นำข้อมูลจากเอกสารชุดที่สอง เข้าประมวลผลในโปรแกรมค้นหาและแทนที่คำหยาบในข้อความภาษาไทยด้วยเครื่องหมายดอกจัน และทำการนับคำดังต่อไปนี้คือ จำนวนคำปกติจริงที่โปรแกรมบอกว่าเป็นคำปกติ จำนวนคำปกติจริงที่โปรแกรมบอกว่าเป็นคำหยาบ จำนวนคำหยาบจริงที่โปรแกรมบอกว่าเป็นคำหยาบ และจำนวนคำหยาบจริงที่โปรแกรมบอกว่าเป็นปกติ แล้วนำค่าที่นับได้ไปใส่ในตารางคอนฟิซชันเมตริก ซึ่งได้กล่าวไปแล้วในบทที่ 3

4.3.2.7 นำเอาผลจากตารางคอนฟิซชันเมตริกจากข้อ 4.3.2.1 และ 4.3.2.6 มารวมกันในตารางคอนฟิซชันเมตริก ตารางใหม่

4.3.2.8 นำค่าต่างๆที่ได้ในตารางคอนฟิซชันเมตริกมาทำการคำนวณค่าที่สนใจในการทดสอบ โดยในการทดสอบนั้นจะเน้นการวัดค่าร้อยละที่โปรแกรมสามารถบอกได้ว่าเป็นคำหยาบจริงและแทนที่ด้วยเครื่องหมายดอกจันได้ถูกต้องจากคำหยาบจริงทั้งหมด ร้อยละที่โปรแกรมบอกว่าเป็นคำหยาบและแทนที่ด้วยเครื่องหมายดอกจัน ทั้งๆที่ค่านั้นเป็นคำปกติจากคำที่โปรแกรมบอกว่าเป็นคำหยาบทั้งหมด และร้อยละที่โปรแกรมบอกว่าเป็นปกติทั้งๆที่ค่านั้นเป็นคำหยาบที่ต้องแทนที่ด้วยเครื่องหมายดอกจันจากคำที่โปรแกรมบอกว่าเป็นคำปกติทั้งหมด

4.3.2.9 ทำการสรุปผลที่ได้จากการทดสอบ ว่ามีประสิทธิภาพการแทนที่คำหยาบด้วยเครื่องหมายดอกจันมากน้อยเท่าใด

สรุปในบทนี้ได้อธิบายถึงขั้นตอนการวิจัย ตั้งแต่ขั้นตอนในการประมวลผล ขั้นตอนการเตรียมเอกสารที่ใช้ในการทดสอบ และขั้นตอนการทดสอบประสิทธิภาพของโปรแกรม ในบทต่อไปนี้จะกล่าวถึงผลการทดสอบประสิทธิภาพ การสรุปผลและข้อเสนอแนะต่างๆที่ได้จากการทำวิจัยในครั้งนี้

บทที่ 5

ผลการทดสอบ สรุปผลการวิจัยและข้อเสนอแนะ

ในบทนี้จะกล่าวถึงชุดข้อมูลที่ใช้ในการทดสอบ ผลการทดสอบข้อมูลตัวอย่างที่ใช้ในทดสอบการตัดคำ การค้นหาและแทนที่คำหายาบในข้อความภาษาไทยด้วยเครื่องหมายดอกจันของโปรแกรมกรองคำหายาบ การวัดประสิทธิภาพ สรุปผลการวิจัย และข้อเสนอแนะ โดยมีรายละเอียดดังต่อไปนี้

5.1 ข้อมูลที่ใช้ในการทดสอบ

ข้อมูลที่ใช้ในการทดสอบเป็นข้อมูลที่ไม่มีรูปภาพ บทกลอน หรือสมการใดๆ ที่ผ่านการตรวจสอบแก้ไขคำต่างๆที่เขียนไม่ถูกต้อง ให้ถูกต้องเรียบร้อยแล้ว โดยได้แบ่งชุดข้อมูลออกเป็นสองกลุ่มคือ ชุดข้อมูลที่เป็นเอกสารประเภทข่าวหนังสือพิมพ์ ประกอบด้วยหนังสือพิมพ์ไทยรัฐ มติชน ข่าวสด และคมชัดลึก ประจำเดือนพฤษภาคม พ.ศ. 2548 ถึง เดือนธันวาคม พ.ศ. 2548 จำนวนหนึ่งร้อยคอลัมน์ และชุดข้อมูลจากข้อความแสดงความคิดเห็นบนเว็บบอร์ดต่างๆในสื่ออินเทอร์เน็ต ตั้งแต่ เดือนสิงหาคม พ.ศ. 2546 ถึง เดือนธันวาคม พ.ศ. 2548 จำนวนเจ็ดร้อยเจ็ดข้อคิดเห็น

5.2 ผลการทดสอบประสิทธิภาพ

ในการทดสอบประสิทธิภาพของโปรแกรมนั้นจะทำการทดสอบทั้งหมดสองส่วน โดยได้ผลการทดสอบดังนี้

5.2.1 ผลการทดสอบประสิทธิภาพของโปรแกรมตัดคำในข้อความภาษาไทย

การทดสอบประสิทธิภาพของโปรแกรมจะทำการวัดประสิทธิภาพหรือละความถูกต้องของการตัดคำจากจำนวนคำทั้งหมดที่นำมาประมวลผล ซึ่งผลของการวัดประสิทธิภาพของโปรแกรมการตัดคำ จากข้อมูลเอกสารชุดที่หนึ่งมีจำนวนคำทั้งหมดหนึ่งหมื่นห้าพันสี่ร้อยเจ็ดสิบหกคำ มีจำนวนคำที่ตัดถูกต้องหนึ่งหมื่นห้าพันสองร้อยหกสิบกคำ และคำที่ตัดไม่ถูกต้องสองร้อยสิบสองคำ โดยคิดเป็นค่าความถูกต้องเท่ากับ 98.63% ตัวอย่างคำที่ตัดไม่ถูกต้องเช่น ข้อความ “อีกบ” ผลการตัดคำคือ “อีก บ” ซึ่งผลลัพธ์ที่ถูกต้องควรจะเป็น “อี กบ” และจากข้อมูลเอกสารชุดที่สองมีจำนวนคำทั้งหมดหกหมื่นสามพันห้าร้อยหกสิบห้าคำ มีจำนวนคำที่ตัดถูกต้องหกหมื่นหนึ่งพันเจ็ดร้อยยี่สิบหกคำและคำที่ตัดไม่ถูกต้องหนึ่งพันแปดร้อยสามสิบเก้าคำ โดยคิดเป็นค่าความถูกต้องเท่ากับ 97.11% ตัวอย่างคำที่ตัดไม่ถูกต้องเช่น ข้อความ “คนโทรศัพท์” ผลการตัดคำคือ “คน โทร ศัพท์” และข้อความ “ยี่ห้ออิชูชู” ผลการตัดคำคือ “ยี่ห้อ อี ชูชู” ซึ่งผลลัพธ์ที่ถูกต้องควรจะเป็น “คน โทรศัพท์” และ “ยี่ห้อ อิชูชู” ตามลำดับ

5.2.2 ผลการทดสอบประสิทธิภาพของโปรแกรมค้นหาและแทนที่คำหายไปในข้อความภาษาไทยด้วยเครื่องหมายดอกจัน

ในการการทดสอบประสิทธิภาพของโปรแกรมนั้นจะใช้ตารางคอนฟิซชันเมตริกดังต่อไปนี้

ตารางที่ 5.1 แสดงคอนฟิซชันเมตริก ในการทดสอบประสิทธิภาพของโปรแกรมค้นหาและแทนที่คำหายไปในข้อความภาษาไทยด้วยเครื่องหมายดอกจัน

สถานะของคำที่ถูกต้องจริง	สถานะของคำจากการประมวลผลของโปรแกรม		
	ปกติ	หาย	รวม
ปกติ	True Neg. A_1	False Pos. A_2	$A_1 + A_2$
หาย	False Neg. B_1	True Pos. B_2	$B_1 + B_2$
รวม	$A_1 + B_1$	$A_2 + B_2$	$A_1 + A_2 + B_1 + B_2$

โดย	A_1	เป็นจำนวนคำปกติจริงและโปรแกรมบอกว่าเป็นคำปกติ
	A_2	เป็นจำนวนคำปกติจริงแต่โปรแกรมบอกว่าเป็นคำหยาบ
	B_1	เป็นจำนวนคำหยาบจริงแต่โปรแกรมบอกว่าเป็นคำปกติ
	B_2	เป็นจำนวนคำหยาบจริงและโปรแกรมบอกว่าเป็นคำหยาบ

โดยจาก ตารางคอนฟิซันเมตริก นั้นจะใช้คำนวณค่าต่างๆ ดังต่อไปนี้

ความอ่อนไหว (Sensitivity) เป็นร้อยละที่โปรแกรมบอกได้ว่าเป็นคำหยาบจริงและแทนที่ด้วยเครื่องหมายดอกจันได้ถูกต้องจากคำหยาบจริงทั้งหมด ($100 * B_2 / (B_1 + B_2)$)

ความเฉพาะเจาะจง (Specificity) เป็นร้อยละที่โปรแกรมบอกได้ว่าเป็นคำปกติจริงได้ถูกต้องจากคำปกติจริงทั้งหมด ($100 * A_1 / (A_1 + A_2)$)

ความแม่นยำ (Accuracy) เป็นร้อยละจากคำทั้งหมดที่โปรแกรมบอกได้ว่าเป็นคำหยาบจริงและแทนที่ด้วยเครื่องหมายดอกจันได้ถูกต้อง รวมกับร้อยละที่โปรแกรมบอกได้ว่าเป็นคำปกติจริงได้ถูกต้องจากคำทั้งหมด ($100 * (A_1 + B_2) / (A_1 + A_2 + B_1 + B_2)$)

ฟอลส์โพซิทีฟ (False Positive) เป็นร้อยละที่โปรแกรมบอกว่าเป็นคำหยาบและแทนที่ด้วยเครื่องหมายดอกจันต่างๆที่คำนั้นเป็นคำปกติจากคำที่โปรแกรมบอกว่ายาบทั้งหมด ($100 * A_2 / (A_2 + B_2)$)

ฟอลส์เนกทีฟ (False Negative) เป็นร้อยละที่โปรแกรมบอกว่าเป็นคำปกติ ทั้งๆที่คำนั้นเป็นคำหยาบที่จะต้องแทนที่ด้วยเครื่องหมายดอกจันจากคำที่โปรแกรมบอกว่าปกติทั้งหมด ($100 * B_1 / (A_1 + B_1)$)

ตารางที่ 5.2 แสดงคอนฟิซันเมตริก ในการทดสอบประสิทธิภาพของโปรแกรมค้นหาและแทนที่คำหยาบในข้อความภาษาไทยด้วยเครื่องหมายดอกจันจากข้อคิดเห็นในเว็บบอร์ด

สถานะของคำที่ถูกต้องจริง	สถานะของคำจากการประมวลผลของโปรแกรม		
	ปกติ	หยาบ	รวม
ปกติ	14,816	8	14,824
หยาบ	6	646	652
รวม	14,822	654	15,476

เมื่อนำข้อมูลจากเว็บบอร์ดมาประมวลผล เพื่อทำการทดสอบโปรแกรมค้นหาและแทนที่คำ
 หยาบในข้อความภาษาไทยด้วยเครื่องหมายดอกจันแล้ว สามารถเขียนให้อยู่ในรูปตารางคอน
 ฟิซันเมตริกดังตารางที่ 5.2 โดยสามารถอธิบายข้อมูลที่อยู่ในตารางได้คือ จำนวนคำปกติจริงและ
 โปรแกรมบอกว่าเป็นคำปกติเท่ากับหนึ่งหมื่นสี่พันแปดร้อยสิบหกคำ จำนวนคำปกติจริงแต่
 โปรแกรมบอกว่าเป็นคำหยาบเท่ากับแปดคำ จำนวนคำปกติจริงทั้งหมดเท่ากับหนึ่งหมื่นสี่พันแปด
 ร้อยยี่สิบสี่คำ จำนวนคำหยาบจริงแต่โปรแกรมบอกว่าเป็นคำปกติเท่ากับหกคำ จำนวนคำหยาบ
 จริงและโปรแกรมบอกว่าเป็นคำหยาบเท่ากับหกหรือสี่สิบหกคำ จำนวนคำหยาบจริงทั้งหมดเท่ากับ
 หกหรือห้าสิบสองคำ จำนวนคำทั้งหมดที่โปรแกรมบอกว่าเป็นคำปกติเท่ากับหนึ่งสี่พันแปดร้อย
 ยี่สิบสองคำ จำนวนคำทั้งหมดที่โปรแกรมบอกว่าเป็นคำหยาบเท่ากับหกหรือห้าสิบสี่คำ และ
 จำนวนคำปกติและคำหยาบทั้งหมดเท่ากับหนึ่งหมื่นห้าพันสี่ร้อยเจ็ดสิบหกคำ

ตารางที่ 5.3 แสดงคอนฟิซันเมตริกในการทดสอบประสิทธิภาพของ โปรแกรมค้นหาและแทนที่
 คำหยาบในข้อความภาษาไทยด้วยเครื่องหมายดอกจันจากข้อความข่าวหนังสือพิมพ์

สถานะของคำที่ถูกต้องจริง	สถานะของคำจากการประมวลผลของโปรแกรม		
	ปกติ	หยาบ	รวม
ปกติ	63,465	55	63,520
หยาบ	0	45	45
รวม	63,465	100	63,565

เมื่อนำข้อมูลจากข่าวหนังสือพิมพ์มาประมวลผล เพื่อทำการทดสอบโปรแกรมค้นหาและ
 แทนที่คำหยาบด้วยเครื่องหมายดอกจันแล้ว สามารถเขียนให้อยู่ในรูปตารางคอนฟิซันเมตริกดัง
 ตารางที่ 5.3 โดยสามารถอธิบายข้อมูลที่อยู่ในตารางได้คือ จำนวนคำปกติจริงและโปรแกรมบอกว่
 เป็นคำปกติเท่ากับหกหมื่นสามพันสี่ร้อยหกสิบห้าคำ จำนวนคำปกติจริงแต่โปรแกรมบอกว่าเป็น
 คำหยาบเท่ากับห้าสิบห้าคำ จำนวนคำปกติจริงทั้งหมดเท่ากับหกหมื่นสามพันห้าร้อยยี่สิบคำ คำ
 หยาบจริงแต่โปรแกรมบอกว่าเป็นคำปกติไม่มี จำนวนคำหยาบจริงและโปรแกรมบอกว่าเป็นคำ
 หยาบเท่ากับสี่สิบห้าคำ จำนวนคำหยาบจริงทั้งหมดเท่ากับสี่สิบห้าคำ จำนวนคำทั้งหมดที่
 โปรแกรมบอกว่าเป็นคำปกติเท่ากับหกหมื่นสามพันสี่ร้อยหกสิบห้าคำ จำนวนคำทั้งหมดที่
 โปรแกรมบอกว่าเป็นคำหยาบเท่ากับหนึ่งร้อยห้า และจำนวนคำปกติและคำหยาบทั้งหมดเท่ากับหก
 หมื่นสามพันห้าร้อยหกสิบห้าคำ

ในการทดสอบประสิทธิภาพ ของโปรแกรมค้นหาและแทนที่คำหายาในข้อความภาษาไทยด้วยเครื่องหมายดอกจัน จากข้อมูลทั้งสองชุดนั้น ต้องกระทำด้วยความระมัดระวังและไม่ควรคิดเป็นค่าเฉลี่ยเชิงเลขคณิต (Arithmetic Means) ดังนั้นจึงต้องสร้างตารางคอนฟิซชันเมตริกใหม่โดยนำตารางคอนฟิซชันเมตริก จากการทดสอบประสิทธิภาพจากข้อมูลทั้งสองชุดมารวมกัน ซึ่งสามารถแสดงได้ดังตารางต่อไปนี้

ตารางที่ 5.4 แสดงคอนฟิซชันเมตริก ในการทดสอบประสิทธิภาพของโปรแกรมค้นหาและแทนที่คำหายาในข้อความภาษาไทยด้วยเครื่องหมายดอกจันจากการรวมคอนฟิซชันเมตริก ในตารางที่ 5.2 และตารางที่ 5.3

สถานะของคำที่ถูกต้องจริง	สถานะของคำจากการประมวลผลของโปรแกรม		
	ปกติ	หายา	รวม
ปกติ	78,281	63	78,344
หายา	6	691	697
รวม	78,287	754	79,041

จากตารางที่ 5.4 สามารถอธิบายข้อมูลที่อยู่ในตารางได้คือ จำนวนคำปกติจริงและโปรแกรมบอกว่าเป็นคำปกติเท่ากับเจ็ดหมื่นแปดพันสองร้อยแปดสิบเอ็ดคำ จำนวนคำปกติจริงแต่โปรแกรมบอกว่าเป็นคำหายาเท่ากับหกสิบสามคำ จำนวนคำปกติจริงทั้งหมดเท่ากับเจ็ดหมื่นแปดพันสามร้อยสี่สิบสี่คำ จำนวนคำหายาจริงแต่โปรแกรมบอกว่าเป็นคำปกติเท่ากับหกคำ จำนวนคำหายาจริงและโปรแกรมบอกว่าเป็นคำหายาเท่ากับหกร้อยเก้าสิบเอ็ดคำ จำนวนคำหายาจริงทั้งหมดเท่ากับหกร้อยเก้าสิบเจ็ดคำ จำนวนคำทั้งหมดที่โปรแกรมบอกว่าเป็นคำปกติเท่ากับเจ็ดหมื่นแปดพันสองร้อยแปดสิบเจ็ดคำ จำนวนคำทั้งหมดที่โปรแกรมบอกว่าเป็นคำหายาเท่ากับเจ็ดร้อยห้าสิบสี่คำ และจำนวนคำปกติและคำหายาทั้งหมดเท่ากับเจ็ดหมื่นเก้าพันสี่สิบเอ็ดคำ

เนื่องจากงานวิจัยนี้ มีความสนใจในแง่ความสามารถในการค้นหาและแทนที่คำหายาในข้อความภาษาไทยด้วยเครื่องหมายดอกจันของโปรแกรม ด้วยเหตุนี้จึงขอพิจารณาประสิทธิภาพในการค้นหาและแทนที่คำหายาด้วยเครื่องหมายดอกจันได้ถูกต้อง ซึ่งก็คือ ค่าความอ่อนไหว (Sensitivity) ในทำนองเดียวกันก็จะพิจารณาความผิดพลาดของโปรแกรมที่ค้นหาและแทนที่คำหายาไม่ถูกต้อง ซึ่งก็คือ ค่าฟอลส์โพซิทีฟ (False Positive) และพิจารณาความผิดพลาดของโปรแกรมที่ไม่สามารถค้นหาและแทนที่คำหายาได้ ซึ่งก็คือค่าฟอลส์เนกทีฟ (False Negative)

โดยสามารถคำนวณร้อยละของทั้งสามค่า ได้ดังนี้

การคำนวณการค้นหาและแทนที่คำหยาบด้วยเครื่องหมายดอกจันได้ถูกต้อง สามารถคำนวณได้จากสูตร ต่อไปนี้

$$E_c = (100 * B_2 / (B_1 + B_2))$$

โดยที่ E_c เป็นร้อยละของการค้นหาและแทนที่คำหยาบด้วยเครื่องหมายดอกจันได้ถูกต้อง

B_1 เป็นจำนวนคำหยาบจริงแต่โปรแกรมบอกว่าเป็นคำปกติ

B_2 เป็นจำนวนคำหยาบจริงและโปรแกรมบอกว่าเป็นคำหยาบ
แทนค่าจากตารางที่ 5.4 ในสูตรจะได้เป็น

$$E_c = ((100 * 691) / 697)$$

$$= 99.14\%$$

การคำนวณความผิดพลาดของโปรแกรมที่ค้นหาและแทนที่คำหยาบไม่ถูกต้อง สามารถคำนวณได้จากสูตร ต่อไปนี้

$$R_g = (100 * A_2 / (A_2 + B_2))$$

โดย R_g เป็นร้อยละของความผิดพลาดของโปรแกรมที่ค้นหาและแทนที่คำหยาบไม่ถูกต้อง

A_2 เป็นจำนวนคำปกติจริงแต่โปรแกรมบอกว่าเป็นคำหยาบ

B_2 เป็นจำนวนคำหยาบจริงและโปรแกรมบอกว่าเป็นคำหยาบ

แทนค่าจากตารางที่ 5.4 ในสูตรจะได้เป็น

$$R_g = ((100 * 63) / 754)$$

$$= 8.36\%$$

การคำนวณความผิดพลาดของโปรแกรมที่ไม่สามารถค้นหาและแทนที่คำหยาบได้ สามารถคำนวณได้จากสูตร ต่อไปนี้

$$F_g = (100 * B_1 / (A_1 + B_1))$$

โดย	F_n	เป็นร้อยละของความผิดพลาดของโปรแกรมที่ไม่สามารถค้นหาและแทนที่คำหยาบได้
	A_1	เป็นจำนวนคำปกติจริงและโปรแกรมบอกว่าเป็นคำปกติ
	B_1	เป็นจำนวนคำหยาบจริงแต่โปรแกรมบอกว่าเป็นคำปกติ

แทนค่าจากตารางที่ 5.4 ในสูตรจะได้เป็น

$$\begin{aligned} F_n &= (100 * 6) / 78,287 \\ &= 0.008\% \end{aligned}$$

จากการทดสอบประสิทธิภาพความถูกต้องของโปรแกรมได้ค่า E_c จากการวัดประสิทธิภาพของโปรแกรมค้นหาและแทนที่คำหยาบในข้อความภาษาไทยด้วยเครื่องหมายดอกจัน มีค่าเท่ากับ 99.14% ซึ่งต่ำกว่า 100% เพียงเล็กน้อยเท่านั้น ค่า R_g ในความผิดพลาดของโปรแกรมที่ค้นหาและแทนที่คำหยาบไม่ถูกต้อง มีค่าเท่ากับ 8.36% และค่า F_n ในความผิดพลาดของโปรแกรมที่ไม่สามารถค้นหาและแทนที่คำหยาบได้ มีค่าเท่ากับ 0.008%

5.3 สรุปผลการวิจัย

โปรแกรมค้นหาและแทนที่คำหยาบในข้อความภาษาไทยด้วยเครื่องหมายดอกจัน ตามที่ได้พัฒนานั้นได้กำหนดนิยามของคำหยาบและคำปกติเพื่อเป็นพื้นฐานในการวิจัยครั้งนี้ และเป็นการพัฒนาโปรแกรมอยู่สองส่วนใหญ่ๆ ส่วนแรกคือ โปรแกรมการตัดคำในข้อความภาษาไทย โดยใช้วิธีการตัดคำด้วยพจนานุกรมแบบเทียบข้อความและเทียบคำที่ยาวที่สุด ซึ่งเป็นวิธีการตัดคำที่ให้ความถูกต้องสูง ประหยัดพื้นที่ในการจัดเก็บและใช้เวลาในการประมวลผลต่ำ เนื่องจากได้นำโครงสร้างต้นไม้เอวีแอลมาประยุกต์ใช้ ซึ่งผลการทดสอบประสิทธิภาพความถูกต้องของโปรแกรมการตัดคำด้วยวิธีดังกล่าว คิดเป็นร้อยละ 97.87 แสดงให้เห็นว่าโปรแกรมสามารถตัดคำได้ถูกต้องสูง

มาก และส่วนที่สองคือ โปรแกรมค้นหาและแทนที่คำหายไปในข้อความภาษาไทยด้วยเครื่องหมายดอกจันโดยใช้ตัวแบบไฟไนต์สเตทแมชชีนด้วยข้อมูลออกมาเป็นแนวทางในการพัฒนา และอธิบายการแทนที่คำหายไปด้วยเครื่องหมายดอกจัน ในการทำงานจะอาศัยข้อมูลนำเข้าที่ได้จากการตัดคำไปประมวลผลค้นหาและแทนที่คำหายไปด้วยเครื่องหมายดอกจัน ซึ่งผลการทดสอบประสิทธิภาพความถูกต้อง ของโปรแกรมค้นหาและแทนที่คำหายไปในข้อความภาษาไทยด้วยเครื่องหมายดอกจัน พบว่าได้ร้อยละของการค้นหาและแทนที่คำหายไปด้วยเครื่องหมายดอกจันได้ถูกต้องเท่ากับ 99.14 แสดงให้เห็นว่าโปรแกรมสามารถค้นหาและแทนที่คำหายไปด้วยเครื่องหมายดอกจันได้ถูกต้องจากคำหายจริงทั้งหมดเกือบ 100% ร้อยละความผิดพลาดของโปรแกรมที่ค้นหาและแทนที่คำหายไม่ถูกต้องเท่ากับ 8.36 แสดงให้เห็นว่ามีคำปกติจริงบางคำที่โปรแกรมแทนที่ด้วยเครื่องหมายดอกจันจากคำที่โปรแกรมบอกว่าเป็นคำหายทั้งหมด และร้อยละความผิดพลาดของโปรแกรมที่ไม่สามารถค้นหาและแทนที่คำหายได้เท่ากับ 0.008 ซึ่งแสดงให้เห็นว่ามีคำหายบางคำที่โปรแกรมไม่สามารถแทนที่ด้วยเครื่องหมายดอกจัน จากคำที่โปรแกรมบอกว่าเป็นคำปกติทั้งหมด ดังนั้นประสิทธิภาพของโปรแกรมแทนที่คำหายไปในข้อความภาษาไทยด้วยเครื่องหมายดอกจันจึงถือว่าเป็นที่น่าพอใจมาก

5.4 ข้อเสนอแนะ

จากผลการทดสอบ ถึงแม้ว่าการทดสอบโปรแกรมจะเป็นที่น่าพอใจก็ตามแต่ก็ยังมีความผิดพลาดอยู่ โดยดูได้จากค่าร้อยละความผิดพลาดของโปรแกรมที่ค้นหาและแทนที่คำหายไม่ถูกต้องและร้อยละความผิดพลาดของโปรแกรมที่ไม่สามารถค้นหาและแทนที่คำหายได้ถูกต้อง ทั้งนี้เนื่องมาจากการตัดคำที่ไม่ถูกต้อง จึงทำให้ได้คำที่เป็นข้อมูลเข้าในการประมวลผลค้นหาและแทนที่คำหายนั้นผิดพลาดไปด้วย เช่น ในกรณีที่ข้อความมีคำที่ไม่ปรากฏในพจนานุกรมรวมอยู่ด้วย ทั้งนี้อาจเกิดจากการเขียนผิดหรือเป็นความตั้งใจ จึงทำให้ตัดคำออกมาผิดพลาด เช่น ข้อความ “ผู้หญิงคนนั้นขับรถอีซูซุ” จะได้ผลลัพธ์คือ “ผู้หญิง คน นั้น ขับ รถ อี ซูซุ” และผลลัพธ์สุดท้ายคือ “ผู้หญิงคนนั้นขับรถ*ซูซุ” ซึ่งโดยความเป็นจริงแล้วข้อความนี้ควรจะได้ผลลัพธ์ “ผู้หญิงคนนั้นขับรถอีซูซุ” เหมือนเดิม แต่เนื่องจากตัดคำแล้วได้คำหายเกิดขึ้น คือ คำว่า “อี” แทนที่ที่จะเป็นคำว่า “อีซูซุ” ทั้งนี้เป็นเพราะคำว่า “อีซูซุ” ไม่มีอยู่พจนานุกรม อีกกรณีหนึ่งคือวิธีการตัดคำแบบเทียบคำที่ยาวที่สุดนั้นยังมีกรณีที่ตัดคำไม่ถูกต้อง เช่น ข้อความ “ผลกระทบจากน้ำท่วม” จะได้ผลลัพธ์เป็น “ผลก ระ ทบ จาก น้ำ ท่วม” ถึงแม้จะตัดได้คำที่มีความหมายตามพจนานุกรมทุกคำก็ตาม และข้อความ “อิสาวชื่อดังหวังรวยทางลัด” จะได้ผลลัพธ์เป็น “อิสาว ชื่อ

คังหวัง รวย ทาง ลัด” ซึ่งนอกจากตัดคำไม่ถูกต้องแล้วยังทำให้คำหายขาดไม่ถูกแทนที่อีกด้วย โดยที่ความเป็นจริงแล้วข้อความนี้ควรจะตัดคำได้เป็น “อี สาว ชื่อ คัง หวัง รวย ทาง ลัด” แล้วคำว่า “อี” ที่เป็นคำหายขาดก็จะถูกแทนที่ด้วยเครื่องหมายดอกจันได้ถูกต้อง

ด้วยเหตุนี้ จึงขอเสนอวิธีการแก้ไขข้อผิดพลาดดังกล่าว เพื่อปรับปรุงให้ประสิทธิภาพในการค้นหาและแทนที่คำหายขาดในข้อความภาษาไทยด้วยเครื่องหมายดอกจัน ให้สูงขึ้น โดยการเพิ่มคำศัพท์ที่ไม่ปรากฏในพจนานุกรม โดยเฉพาะคำที่มีคำหายขาดประสมอยู่ด้วยเข้าไปในพจนานุกรม เช่นคำว่า “อีชูชู” “อีลินนิง” ซึ่งคำหายขาดคือคำว่า “อี” ทั้งสองคำ อีกอย่างหนึ่งให้เพิ่มคำศัพท์ที่ทำให้การตัดคำแบบเทียบคำที่ยาวที่สุดไม่ถูกต้อง ซึ่งคำเหล่านี้อาจจะเกิดมาจากการเขียนประสมกันระหว่างคำที่ปรากฏในพจนานุกรมทุกคำ โดยเฉพาะอย่างยิ่งคำศัพท์ที่มีคำหายขาดประสมอยู่ด้วย เช่น คำว่า “อีสาว” เกิดจากการประสมกันระหว่างคำว่า “อี” กับ “สาว” ซึ่งในที่นี้การตัดคำแบบเทียบคำที่ยาวที่สุดจะได้เป็น “อีสาว” ในกรณีเช่นนี้ให้แก้ไขโดยทำการเก็บข้อความว่า “อีสาว” และกำหนดผลลัพธ์ในการตัดคำในที่นี้จะเป็น “อี สาว” ก็จะทำให้การตัดคำนั้นมีความถูกต้องมากยิ่งขึ้น

งานวิจัยนี้ พัฒนาขึ้นเพื่อใช้ในการสกัดคำหายขาดไม่ปรากฏในข้อความภาษาไทยโดยการค้นหาและแทนที่คำหายขาดในข้อความภาษาไทยด้วยเครื่องหมายดอกจัน ซึ่งในการนำไปใช้จริงนั้น ผู้ใช้อาจจะทำการประยุกต์ใช้ในเว็บไซต์ของตนเองหรือขององค์กร โดยผู้ใช้สามารถกำหนดคำหายขาดหรือคำอื่นๆที่ไม่ต้องการให้ปรากฏในข้อความภาษาไทยขึ้นมาเอง และเป็นแนวทางการปรับปรุงในการวิจัยต่อไป

คลังข้อความพิเศษ เป็นคลังข้อความที่ได้จัดเก็บข้อความที่มีความจำเป็นต่อเอกสารในการทดสอบของงานวิจัยฉบับนี้ เพื่อให้การตัดคำมีความถูกต้องมากที่สุด โดยข้อความที่ได้ทำการจัดเก็บในคลังข้อความพิเศษนั้น จะประกอบด้วยข้อความที่ตัดคำแบบเทียบคำที่ยาวที่สุดแล้วได้ผลลัพธ์ที่ไม่ถูกต้อง และผลลัพธ์ของการตัดคำที่น่าจะเป็นไปได้มากที่สุดของข้อความพิเศษนั้น เช่นข้อความ “การกล่าว” จะมีผลลัพธ์เป็น “การ กล่าว” และข้อความ “ต้องเตรียม” จะมีผลลัพธ์เป็น “ต้อง เตรียม” โดยข้อความสองข้อความนั้น ถ้าตัดคำแบบเทียบคำที่ยาวที่สุดแล้วจะได้ผลลัพธ์เป็น “การก กล่าว” และ “ต้องเต รี ยม” ตามลำดับ ดังนั้นในการประมวลผลโปรแกรมจากข้อมูลเอกสารอื่น อาจจะมีข้อความอื่นที่ตัดคำแบบเทียบคำที่ยาวที่สุดแล้วได้ผลลัพธ์ที่ไม่ถูกต้อง ก็สามารถเพิ่มข้อความพิเศษและผลลัพธ์ที่น่าจะเป็นไปได้มากที่สุดของข้อความนั้นในคลังข้อความพิเศษเพื่อทำให้การตัดคำในข้อมูลเอกสารนั้นถูกต้องมากยิ่งขึ้น

บรรณานุกรม

- ชมภูนุช คุปติวุฒิ. 2542. การตัดคำข้อความภาษาไทยด้วยโปรแกรมตรรกะเชิงอุปนัย.
วิทยานิพนธ์ปริญญาโทมหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย.
- นัฐวุฒิ ไชยเจริญ. 2544. การตัดคำและการกำกับหมวดคำภาษาไทยแบบเบ็ดเสร็จด้วยคอมพิวเตอร์.
วิทยานิพนธ์ปริญญาโทมหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย.
- ไพศาล เจริญพรสวัสดิ์. 2541. การตัดคำภาษาไทยโดยใช้คุณลักษณะ. วิทยานิพนธ์ปริญญา
มหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย.
- ยี่น ภู่วรรณ และวิวรรธ อิมอาร์มณ. 2529. การแบ่งแยกพยางค์ไทยด้วยดิคชันนารี. รายงาน
ประชุมวิชาการวิศวกรรมไฟฟ้า สถาบันอุดมศึกษาแห่งประเทศไทยครั้งที่ 9, 3-4
ธันวาคม 2529. ขอนแก่น: หอประชุมศูนย์วิทยาศาสตร์สุขภาพ
มหาวิทยาลัยขอนแก่น. หน้า 1-13.
- วิจิตร ภาณุพงศ์. 2524. โครงสร้างภาษาไทย: ระบบไวยากรณ์. กรุงเทพมหานคร: มหาวิทยาลัย
รามคำแหง.
- วิจิต หล่อจิระชุนท์กุล และเจริญ คุวินทร์พันธุ์. 2524. อัลกอริทึมการแยกพยางค์ภาษาไทย.
รายงานการวิจัย เสนอต่อคณะสถิติประยุกต์ สถาบันบัณฑิตพัฒนบริหารศาสตร์.
- วิจิต หล่อจิระชุนท์กุล และจิราวัลย์ จิตรถเวช. 2547. เอกสารประกอบการประชุมวิชาการสถิติ
ประยุกต์ระดับชาติ: การประยุกต์การทำเหมืองข้อมูลกับการประกันรถยนต์, 14-15
ตุลาคม 2547. กรุงเทพมหานคร: โรงแรม Merchant Court. หน้า 6-7.
- วิรัช ศรีเลิศล้ำวานิช. 2535. การตัดคำไทยในระบบแปลภาษา. ใน การแปลภาษาด้วย
คอมพิวเตอร์. กรุงเทพมหานคร: ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์
แห่งชาติ. หน้า 3-6.
- สมนึก สิ้นชูปวน. 2546. การวิเคราะห์การกระจายคำในประโยคภาษาไทยโดยการโปรแกรมเชิง
เจเนติก. วิทยานิพนธ์ปริญญาโทมหาบัณฑิต สถาบันบัณฑิตพัฒนบริหารศาสตร์.
- สัมพันธ์ ธรินทร์มย์. 2534. การแบ่งคำไทยด้วยพจนานุกรม. กรุงเทพมหานคร: ภาควิชาวิศวกรรม
คอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย.

- อมรทิพย์ กวินปณิธาน. 2546. การศึกษาริบทบงบอกชื่อเฉพาะในภาษาไทยตามแนว
 ภาษาศาสตร์คอมพิวเตอร์. วิทยานิพนธ์ปริญญาโทมหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย.
 อุปกิตติปลสาร, พระยา. 2533. หลักภาษาไทย. กรุงเทพมหานคร: ไทยวัฒนาพานิช.
- Asanee Kawtrakul, Chalathip Thumkano and Surapol Seriburi. 1995. A Statistical Approach to
 Thai Word Filtering. In **Proceedings of the Symposium on Natural Language
 Processing in Thailand'95**. Bangkok: Chulalongkorn University. Pp. 398-406.
- Golding, Andrew R. Golding and Roth, Dan. 1996. Applying Winnow to Context-Sensitive
 Spelling Correction. In **Proceedings of the Twelfth International Conference on
 Machine Learning**. Tahoe City, Calif: Morgan Kaufmann. Pp. 182-190.
- Rosen, Kenneth H. 2003. **Discrete Mathematics and Its Applications**. 5th ed. Boston:
 McGraw Hill.
- Surin Charnyapornpong. 1983. **A Thai Syllable Separation Algorithm**. Master's thesis, Asian
 Institute of Technology.
- Surapant Meknavin, Paisarn Chareonponsawat and Boonserm Kijisirikul. 1997. Feature-based
 Thai Word Segmentation. In **Proceedings of the Natural Language Processing
 Pacific Rim Symposium'97**. Phuket: Chulalongkorn University. Pp. 41-48.
- Weiss, Mark Allen. 1995. **Data Structures and Algorithm Analysis**. 2nd ed. Redwood City,
 Calif: Benjamin/Cumming.
- Cohen, William W. 1995. Fast Effective Rule Induction. In **Proceedings of the Twelfth
 International Conference on Machine Learning**. Tahoe City, Calif: Morgan
 Kaufmann. Pp. 115-123.
- Yupin Thairatananond. 1981. **Towards the design of a Thai text syllable analyzer**. Master's
 thesis, Asian Institute of Technology.

ภาคผนวก

ภาคผนวก ก

รหัสเลขฐานสองของตัวอักษรในภาษาไทยที่ใช้ในรหัสต้นฉบับ

ก	=	10100001
ข	=	10100010
ฃ	=	10100011
ค	=	10100100
ฅ	=	10100101
ฉ	=	10100110
ง	=	10100111
จ	=	10101000
ฉ	=	10101001
ช	=	10101010
ฌ	=	10101011
ฉ	=	10101100
ญ	=	10101101
ฎ	=	10101110
ฏ	=	10101111
ฐ	=	10110000
ฑ	=	10110001
ฒ	=	10110010
ณ	=	10110011
ด	=	10110100
ด	=	10110101
ถ	=	10110110
ท	=	10110111
ธ	=	10111000

น	= 10111001
บ	= 10111010
ป	= 10111011
ผ	= 10111100
ฝ	= 10111101
พ	= 10111110
ฟ	= 10111111
ภ	= 11000000
ม	= 11000001
ย	= 11000010
ร	= 11000011
ล	= 11000101
ว	= 11000111
ศ	= 11001000
ษ	= 11001001
ส	= 11001010
ห	= 11001011
ฬ	= 11001100
อ	= 11001101
ฮ	= 11001110
ฤ	= 11000100
ฦ	= 11000110
สระเอ	= 11100000
สระแอ	= 11100001
สระไ้มลาย	= 11100010
สระโaim่วน	= 11100011
สระโอ	= 11100100
ไ้ปยาล	= 11100110
ไ้ม่การ์นต์	= 11010001
สระอา	= 11100101

ไม้ไต่คู้	= 11100111
สระอำ	= 11010011
สระอะ	= 11010000
สระอิ	= 11010100
สระอี	= 11010101
สระอุ	= 11011000
สระอู	= 11011001
สระือ	= 11010111
สระืออ	= 11010110
หยาดน้ำค้าง	= 11101101
วรรณยุกต์เอก	= 11101000
วรรณยุกต์โท	= 11101001
วรรณยุกต์ตรี	= 11101010
วรรณยุกต์จัตวา	= 11101011
ไม้ก้าน	= 11101100
ฯ	= 11101101
อักขระอื่นๆ	= 11111111

ภาคผนวก ข

คำหยาบที่มีความหมายตามนิยามในงานวิจัย

ก่นโคตร

กรรกศ

กระเจี๊ว

อ้ายเจี๊ว

ไฮ้เจี๊ว

กระดอ

กระเต้า

กระแต้แร่

กระแตะ

กระแตะกระแต้

กระบาล

กริกกริว

กั๊ดมัน

กะห่า

กาที

กามกิจ

กามินี

กาลี

ก้ามะลอ

กำหนด

ก๊วย

กู

แก่นแก้ว

ขม้

ของลับ

จิก

จีจาบ

จีถัง

จีซ้า

จีคุก

จีหน้า

ขุนเพ็ด

ไผ่ห้า

คบชู้

ควาย

ค้ำประเวณี

คุษฐาน

คุษประเทศ

พระคุษฐาน

เจียน

ง่าน

จังไร

จัญไร

จันท์

จันฑาล

ใจบาป

ฉิบหาย

ชั่ว

ชั่วช้า

ช้าเรา

ชู้

ช่องเสพ

ช้าสาม

เต่าหัวงู

คัสกรี

คิริจัน

เค้า

เคร์จัน

เคียร์จัน

แคก

แคค

แคะเค้

ค้ำซ่า

ถวัล

ถ้อย

ถะถุนถะถัน

ถุย

ถูล

ถูล

ถูลละ

ทราชาตี

ทราพล

ทราม

ทุ

ทุราจาร

ทุเรศ

โทรมหญิง

นครโสภินี

หญิงนครโสภินี

นครโสภณี

หญิงนครโสภณี

นัยกัษ

นางบำเรอ

นางโคม

นิรคุณ
เนรคุณ
บัก
ประทุฐ
ประทุษ
ประลัย
ปลัดจิก
ปากม้า
ปากหมา
ปี่
เปิดบริสุทธิ
ผ่าเหล่า
ผิดเมีย
ผิดลูกผิดเมีย
ผีทะเล
ฝัก
แพศยา
ไพร่
ฟอนเฟะ
มีชู
มึง
เม็ดละมุด
เมहनะ
แม่เล่า
ยะ
ยาหัด
เย็ด
โยนี
ร่วมประเวณี
ร่วมเพศ

ร่วมรัก
คอกทอง
ระยำ
ระยำดำบอน
ราคะ
ร้าน
ริ
เราะร้าย
ร้าย
แรด
ลอยคอก
ลักเพศ
ลามก
ลึงค์
ลูกสวาท
เล่นจัญไร
เล่นระยำ
เล่นอับริย์
เล่นอุบาทว์
เล่นซู้
เล่นสวาท
เล่นสัปคน
เลว
เลือดชั่ว
โลน
วฤษณะ
วสณะ
วสดี
วะ
ว้ายร้าย

เวศยา
ไว้ย
เวี้ย
ศิวลิ่งค์
สกปรก
สฤถ
สันติน
สบจ
สมคบ
สมจร
สมยอม
สมรัก
สมฐู้
สมฐู่
สมพาส
สมโยค
สวมเขา
สวะ
สอคฐู้
สอคฐู้สอคเห็น
สะเก็ดตินเมรุ
สะคั้ง
สคั้งสะคั้ง
สังวาส
หณิงสัจจรโรค
สัด
สันดาน
สาแกใจ
สาธารณ์
สามหา

สามานย์
สารระยำ
สาวไส้
ลำสอน
สุ่มหัว
เต็งเค็ง
เสนียด
เสียดา
เสียดายดา
เสียดิน
เสียดู
เสื่อมทรม
แสบ
เสือก
โสกโคก
โสกโครก
โสกถึ
โสกถึ
โสมม
หญิงงามเมือง
หญิงสามพร้าว
หญิงหากิน
หมอย
หมาจู้เรื้อน
หมากไข่ห้า
หยาบ
หยาบๆ
หยาบคาย
หยาบซ้ำ
หยาบโลน

หยาบหยาม

หยาม

หยามน้ำหน้า

หญฺโหด

หลุกหลิก

หัวถอก

หัวหน่าว

หำ

หีน

หี

หีน

หีน

หั้น

หือรือโหด

โหด

เฮงชวย

อกุศล

อวชาต

อสาธุ

อสาธร

น้ำกาม

อันทะ

อันตชาติ

อัปรีย์

อ้าซ่า

อ้าย

ไเฮ้

อี

อีดอก

อุบาทว์

อุปีทวะ
อุปีทวันตราย
อุลามก

ภาคผนวก ก

ตารางแสดงข้อความในเอกสารพิเศษที่ใช้ในการตัดคำโดยวิธีเทียบข้อความ

ข้อความ	ผลลัพธ์จากการตัดคำโดยวิธีเทียบคำที่ยาวที่สุด	ผลลัพธ์ที่ควรจะได้
การกระจาย	การก / ระ / าย	การ / กระจาย
การกระโดด	การก / ระ / โดด	การ / กระโดด
การกระทำ	การก / ระ / ทำ	การ / กระทำ
การกระทัีบ	การก / ระ / ทัีบ	การ / กระทัีบ
การกระพือ	การก / ระ / พือ	การ / กระพือ
การกล่าว	การก / ล่าว	การ / กล่าว
คำของ	คำขอ / ง	คำ / ของ
จงแสดง	จง / แสดง / ง	จง / แสดง
จอแสดง	จอ / แสดง / ง	จอ / แสดง
ต้องตาม	ต้องตา / ม	ต้อง / ตาม
ต้องตาย	ต้องตา / ย	ต้อง / ตาย
ต้องเตรียม	ต้องเต / รี / ยม	ต้อง / เตรียม
ตาบวม	ตาบ / วม	ตา / บวม
แต่งบ	แต่ง / บ	แต่ง / บ
แต่ละคร	แต่ละ / คร	แต่ละ / คร
ทำตาม	ทำตาม / ม	ทำ / ตาม
ทางโทรศัพท์	ทางโท / ร / ศัพท์	ทาง / โทรศัพท์
ประสานงาน	ประสานงา / น	ประสาน / งาน
ปากของ	ปากขอ / ง	ปาก / ของ
เป็นบ้าน	เป็นบ้าน / น	เป็น / บ้าน
เป็นอันตราย	เป็นอัน / ต / ราย	เป็น / อันตราย
ผลกระทบ	ผลก / ระ / ทบ	ผล / กระทบ

ตารางแสดงข้อความในเอกสารพิเศษที่ใช้ในการตัดคำโดยวิธีเทียบข้อความ (ต่อ)

ข้อความ	ผลลัพธ์จากการตัดคำโดยวิธีเทียบคำที่ยาวที่สุด	ผลลัพธ์ที่ควรจะได้
ผลออก	ผล / ออก	ผล / ออก
ผิดกฎหมาย	ผิด / กฎ / หมายถึง	ผิด / กฎหมาย
พักตาม	พักตา / ม	พัก / ตาม
พักรอ	พักร / อ	พัก / รอ
พาหญิง	พาห / หญิง	พา / หญิง
มันตาย	มันตา / ย	มัน / ตาย
มากกระทรวง	มาก / ระ / ทรวง	มา / กระทรวง
มากระแทก	มาก / ระ / แทก	มา / กระแทก
มาตรวจ	มาตร / วจ	มา / ตรวจ
มาตาก	มาตา / ก	มา / ตาก
มาตาม	มาตา / ม	มา / ตาม
มาทดสอบ	มาท / ด / สอบ	มา / ทดสอบ
มาบรรยาย	มาบ / ร / ร / ยาย	มา / บรรยาย
มาบริเวณ	มาบ / ริ / เวณ	มา / บริเวณ
มาบริหาร	มาบ / ริ / หาร	มา / บริหาร
มาบอก	มาบ / ออก	มา / บอก
มายก	มาย / ก	มา / ยก
มารวม	มาร / วม	มา / รวม
มารอ	มาร / อ	มา / รอ
มาสรุป	มาส / รุ / ป	มา / สรุป
มาสอบ	มาส / อบ	มา / สอบ
มีคนตรี	มีด / น / ตรี	มี / คนตรี
มีนโยบาย	มีน / โย / บาย	มี / นโยบาย
ยานอนหลับ	ยาน / อน / หลับ	ยา / นอน / หลับ
ร้องของ	ร้องขอ / ง	ร้อง / ของ
ถูกตาม	ถูกตา / ม	ถูก / ตาม

ตารางแสดงข้อความในเอกสารพิเศษที่ใช้ในการตัดคำโดยวิธีเทียบข้อความ (ต่อ)

ข้อความ	ผลลัพธ์จากการตัดคำโดยวิธีเทียบคำที่ยาวที่สุด	ผลลัพธ์ที่ควรจะได้
ลูกนาย	ลูกนา / ย	ลูก / นาย
ว่านอน	ว่าน / อน	ว่า / นอน
หน้าหนาว	หน้าหนา / ว	หน้า / หนาว
หมอนวด	หมอน / นวด	หมอ / นวด
หลากว่า	หลาก / ว่า	หลา / กว่า
หอกกระจายข่าว	หอก / ระ / จา / ย / ข่าว	หอก / กระจาย / ข่าว
หากระดาษ	หา / ระดา / ษ	หา / กระดาษ
เหรด	เห / รด	เหรด
หฺรออ	เห / รออ	หฺรออ
อีสาว	อีสา / ว	อี / สาว

ภาคผนวก ง

ตัวอย่างข้อความจากเว็บบอร์ดที่ใช้ทดสอบ
ประสิทธิภาพของโปรแกรมในงานวิจัย

อินเทอร์เน็ตใช้ในทางสื่อสารมวลชนอย่างไรและเกี่ยวข้องกันอย่างไร

ประโยชน์ หลักการใช้งาน และการให้บริการจากอินเทอร์เน็ต

อยากมีเซ็กมาทางนี้

อยากรักเขาแต่ไม่รู้ว่าจะรัก

อีเจ็อยบ้า อ้วนด้วย

อีหี้ยอีหน้าตัวเมีย

บักพิลิตบ้ามี่เมียชื่อน้องปรา

ดีมากๆ

ดีคะ แล้วชอบเข้าไปเล่นด้วยคะ

เฮ็ดแม่ช่วยตอบมาดีๆหน่อย

ทำไมเขียนกันไม่ค่อยสุภาพเลยนะ

จะเลิกกันทำไมรักกันดีไม่ใช่หรอ

ตกใจเหมือนกันนะคะพีนุ่น ตอนแรกเนนว่าเหมาะสมกันดี เป็นคู่ที่น่ารัก เนนยังแอบอิจฉาเลย แต่คนเราก็มียุติผลของตัวเองนะ ยังไงก็ยังชอบและปลื้มพีนุ่นมากๆนะคะ

เขตต์น่าจะพีนุ่นไปเลย ให้อ้อมไม่มีพ่อเลยอะ

ทำไมผู้หญิงต้องเป็นอย่างนี้ทุกคนก็ไม่รู้คบกันมาตั้งนานแทนที่จะคุยกัน

สมน้ำหน้าสะใจดี ฮ่าๆๆๆๆๆๆ

รักแท้ดูแลไม่ได้

ไม่น่าเลิกคบกันเลย คาราคู่อื่นไม่เหมาะสมกันเท่าคู่นี้เลยสู่อุ่น+เขตต์ไม่ได้เพราะน่ารักกันทั้งคู่

ทุกคนอย่าโกรธนะ มือที่สามคือเราเองนะ ล้อเล่น

ก๊นสวยน่ารักจัง

ดำ จีบหาย แล้วดูใส่เสื้อสีขาว สีเขียว ไม่ค่อยจะตัดกับสีผิวเลยนะ รู้ว่าสวยแม่ แล้วดูสายเดี่ยวสีขาว ไม่ใส่ยกทรงหรือสงสัยอยากให้พี่เขาเห็นนะ

ไอ้บ้า อยู่ดีดีไม่ชอบ

แล้วไปเสือกอะไรกับเรื่องของเค้าวะ

กอด จูบ มีเช็ก เป็นเรื่องธรรมดาของเด็กๆอยู่แล้ว จะไปสนใจอะไรกันนักหนาเรื่องส่วนตัวของเขา จะดีจะเลวช่างเขาเถอะ

ไม่เคยชอบบอลเลย ไม่รู้มาช่วยอมคบด้วยได้ไง ไม่เห็นเป็นสุภาพบุรุษเลย ว่าแต่ว่าตอนนี้มาช้คบกับใครอยู่คะ?

มายสาวจอมฉาว..นี่เคยเห็นแถวย่านนาวากับเสี่ยและเพื่อนก็บอกว่าเคยเห็นที่เชียงใหม่ในโรงแรมกับเสี่ยก่อนที่จะเข้าวงการ

นึกถึงเรื่องไอ้บอลกับมาซ่า แล้วอยากถามมันว่า "มึงเป็นลูกผู้ชายเปล่าวะ"

วิกหลายสี่ช่วยคุณปูไปรยาหน่อยเถอะ เดี่ยวนี้ดูไม่ได้เลยคิดจะขายนม (ยานๆ) หรือจ๊ะ ดูไม่ได้เลยน่าเกลียดมาก

กินบนเรือนขึ้นหลังคาจริงๆ เลยคารานี้ หาผลประโยชน์กับทุกเรื่อง แม้เรื่องตั่วๆ ของพวกเธอๆ ทั้งหลาย

ยังเอามาขายประจานกันอีก เฮ้อ....เบื้อ

ไม่รับผิดชอบงานเลขนะซ้อ ป่านี่ละ เดี่ยวก้อโดนไล่ออกหรอก

ซ้อจำ 5 โมงแล้วนะยังไม่มาอีกหรอรออยู่นะจ๊ะ

นางเอกหน้าอเป็นเมียน้อยผู้จัดด้วยเธอ ผัวเยอะจัง

ทุกอย่างที่เกิดขึ้นเป็นเรื่องระหว่างคน 2 คน ใครทำอะไรย่อมรู้แก่ใจดี แต่การเอาคนที่เคยรักมาพูดให้คนอื่นฟังมันเป็นการกระทำที่ผู้ชายดีๆ เขาไม่ทำกัน หนังสือที่จะออกมานี้คิดให้ดีกว่าก่อน คงจะมีคนด่ามากกว่ามีคนชม

นี้ๆ ซ้อ มาเถอะ จะเข้าห้องน้ำก็ไม่เข้า รออยู่นะ ปวดมากๆ แต่อยากอ่าน

อยากให้เยาวชน ฟึกและศึกษา การใช้ภาษาไทย จาก ซ้อ นะครับ อาจจะมีเขียน คล้ายของซ้อ แล้วให้เพื่อนๆ ใน Web ทายกัน ผมคิดว่าจะช่วย เสริมสร้าง สิ่งดีๆ นะครับ

ช่วยแจ้งเจ้าแม่ช่องหลายสี่ด้วยว่า คาราของท่านทำไมชอบโชว์นมจังเลย โดยเฉพาะคาราหน้าใหม่อย่าง ปู ไปรยา เพราะเด็กๆ ดูแล้วรู้สึกเป็นตัวอย่างที่ไม่ดีเลย เดี่ยวโชว์โน่นโชว์นี้สารพัดไม่รู้จะโชว์

ทำใหม่ คุณอย่างนางรองที่ชื่อสา คำเรียบร้อย น่ารัก คุณแล้วสวยกว่าต้องเยอะ ทำมัยไม่ทำแบบคำข้าง
 ควรจะสรรหาเนื้อหาละครที่เป็นการส่งเสริมวัฒนธรรม หรือเป็นตัวอย่งที่ดีเหมือนกับช่อง 3 บ้าง
 เพราะคำเน้นละครที่ส่งเสริมวัฒนธรรม หรือมีอะไรที่น่าดูกว่าเยอะ เดี่ยวนี้ละครช่อง 7 ไม่ค่อยมี
 ละครที่ดีๆ เลย ไม่รู้ว่าหมดความสามารถแค่นี้หรือ ผากด้วยนะ คนไทย

เจมีพฤติกรรมของ อ้อม สุณิสรา สุขบุญสังข์มารายงานบ้างใหม่ ไม่ค่อยเห็นเป็นข่าวเลย หัวจริงหรือ

กำลังร้อนรนไปทั้งวงการเลยตอนนี้ เรื่องที่หนุ่มมาดเซอร์ "บอล" อธิป นานา จะออกพ็อกเก็ตบุ๊ก
 แออดิตคนรัก มาซา วัฒนพานิช นักร้องและดารารุ่นพี่ที่ตลอดกาล

พ็อกเก็ตบุ๊กของหนุ่มบอลยังไม่รู้หน้าตาเป็นยังไง รู้แต่ว่าดังกระหึ่มเมืองไปแล้ว เมื่อหนุ่มบอล
 ออกมาและเหยียดขี้บิในรายการตีสิบของเสี่ยทีวี วิทวัส สุนทรวิเนตร เมื่อค่ำวันที่ 19 เมษายน ที่ผ่าน
 มา

อยากให้พูดถึงอีกบเกลียดมัน กระแสะ แม่บรูคดีด้วยแล้วหรือ จี้คุย บรูคยังพูดเลยในรายการของ
 สัญญาว่าคบให้ชื้อของให้แล้วออกอากาศว่าประมาทหลายบาทนะ มันจ้องจะจับชื้อๆ ทำไมมันถึงมี
 แพนคลับหน้ามีคตามัวกันอยู่ได้

ตอนนี้อยากให้ชื้อพูดถึงสามไฮโซให้ฟังหน่อย อยากรู้ว่าชื้อคิดยังไง

โดยเฉพาะฉากล้วง ฉากเดินกอดคนมอธสาวไซค์ไลน์ของนักร้อง

ธสาวชื้อคังหวังรวยทางลัด

อีกบมันก็ชอบเขียนแต่สิ่งไร้สาระ และยุงชาวบ้านให้ทะเลาะกัน

อีกบมันคือใคร เลวจริงๆ

ถ้าพวกมึงมีปัญหา มาเจอกูได้ กูคลราชบุรี

ภาคผนวก จ

ผลการตัดคำและการแทนที่คำหายับด้วยเครื่องหมายดอกจันของตัวอย่าง ข้อความจากระบบบอร์ดที่ใช้ทดสอบประสิทธิภาพของโปรแกรมในงานวิจัย

ผลการตัดคำ อิน / เต / อร์เน็ต / ใช้ / ใน / ทาง / สื่อสารมวลชน / อย่างไม่ / และ / เกี่ยวข้อง / กัน /
อย่างไร /

ผลการแทนที่คำหายับด้วยเครื่องหมายดอกจัน อินเทอร์เน็ตใช้ในทางสื่อสารมวลชนอย่างไรและ
เกี่ยวข้องกันอย่างไร

ผลการตัดคำ ประโยชน์ / หลักการ / ใช้ / งาน / และ / การ / ให้ / บริการ / จาก / อิน / เต / อร์
เน็ต /

ผลการแทนที่คำหายับด้วยเครื่องหมายดอกจัน ประโยชน์ หลักการใช้งาน และการให้บริการจาก
อินเทอร์เน็ต

ผลการตัดคำ อยาก / มี / เช็ก / มา / ทาง / นี้ /

ผลการแทนที่คำหายับด้วยเครื่องหมายดอกจัน อยากมีเช็กมาทางนี้

ผลการตัดคำ อยาก / รัก / เขา / แต่ / ไม่ / รู้ / ว่า / เขา / รัก /

ผลการแทนที่คำหายับด้วยเครื่องหมายดอกจัน อยากรักเขาแต่ไม่รู้ว่าเขารัก

ผลการตัดคำ อี / เจ้ / จอ / ย / บ้า / อ้วน / ด้วย /

ผลการแทนที่คำหายับด้วยเครื่องหมายดอกจัน *เจ้จอยบ้า อ้วนด้วย

ผลการตัดคำ อี / เหี้ย / อี / หน้า / ตัวเมีย /

ผลการแทนที่คำหายับด้วยเครื่องหมายดอกจัน *เหี้ย*หน้าตัวเมีย

ผลการตัดคำ บัก / พิ / ลี / ท / บ้า / มี / เมีย / ซื่อ / นื่อง / ป / รา /

ผลการแทนที่คำหายับด้วยเครื่องหมายดอกจัน *พิลิตบ้ามีเมียซื่อนื่องปรา

ผลการตัดคำ ดี / มาก / ๑ /

ผลการแทนที่คำหายับด้วยเครื่องหมายดอกจัน ดีมาก๑

ผลการตัดคำ ดี / คะ / แล้ว / ชอบ / เข้าไป / เล่น / ด้วย / คะ /

ผลการแทนที่คำหายับด้วยเครื่องหมายดอกจัน ดีคะ แล้วชอบเข้าไปเล่นด้วยคะ

ผลการตัดคำ เย็ด / แม่ / ช่วย / ตอบ / มา / ดี / ๑ / น้อย /

ผลการแทนที่คำหายับด้วยเครื่องหมายดอกจัน *แม่ช่วยตอบมาดี๑น้อย

ผลการตัดคำ ทำไม / เขียน / กัน / ไม่ค่อย / สุภาพ / เลย / นะ /

ผลการแทนที่คำหายับด้วยเครื่องหมายดอกจัน ทำไมเขียนกันไม่ค่อยสุภาพเลยนะ

ผลการตัดคำ จะ / เลิก / กัน / ทำไม / รัก / กัน / ดี / ไม่ / ไร / เหนอ /

ผลการแทนที่คำหายับด้วยเครื่องหมายดอกจัน จะเลิกกันทำไมรักกันดีไม่ไรเหนอ

ผลการตัดคำ ตกใจ / เหมือน / กัน / นะ / คะ / พี่ / นุ่น / ตอน / แรก / แน่น / ว่า / เหมาะสม / กัน /
ดี / เป็น / คู่ / ที่ / น่า / รัก / แน่น / ยัง / แอบ / อิจฉา / เลย / แต่ / คน / เรา / ก็ / มี / เหตุผล / ของ
ตัว / เรา / เอง / นะ / ยัง / ไร / ก็ / ยัง / ชอบ / และ / ปลื้ม / พี่ / นุ่น / มาก / ๑ / นะ / คะ /

ผลการแทนที่คำหายับด้วยเครื่องหมายดอกจัน ตกใจเหมือนกันนะคะพี่นุ่น ตอนแรกแน่นว่า
เหมาะสมกันดี เป็นคู่ที่น่ารัก แน่นยังแอบอิจฉาเลย แต่คนเราก็มีเหตุผลของตัวเองนะ ยังไรก็ยัง
ชอบและปลื้มพี่นุ่นมาก๑นะคะ

ผลการตัดคำ เข / ตต์ / นำ / จะ / ฟัน / นุ่น / ไป / เลย / ให้ / ท้อง / ไม่ / มี / พ่อ / เลย / วะ /

ผลการแทนที่คำหายับด้วยเครื่องหมายดอกจัน เขตต์นำจะฟันนุ่นไปเลยให้ท้องไม่มีพ่อเลย*

ผลการตัดคำ ทำไม / ผู้หญิง / ต้อง / เป็น / อย่าง / นี้ / ทุก / คน / ก็ / ไม่ / รู้ / คบ / กัน / มา / ตั้ง / นาน / แทนที่จะ / ลุย / กัน /

ผลการแทนที่คำหายไปด้วยเครื่องหมายดอกจัน ทำไมผู้หญิงต้องเป็นอย่างนี้ทุกคนก็ไม่รู้คบกันมาตั้งนานแทนที่จะลุยกัน

ผลการตัดคำ สมน้ำหน้า / สะใจ / ดี / ฮ่าๆๆๆๆๆ /

ผลการแทนที่คำหายไปด้วยเครื่องหมายดอกจัน สมน้ำหน้าสะใจดี ฮ่าๆๆๆๆๆ

ผลการตัดคำ รัก / แทะ / ดูแล / ไม่ / ได้ /

ผลการแทนที่คำหายไปด้วยเครื่องหมายดอกจัน รักแท้ดูแลไม่ได้

ผลการตัดคำ ไม่ / นำ / เลิก / คบ / กัน / เลย / ดารา / คู่ / อื่น / ไม่ / เหมาะสม / กัน / เท่า / คู่ / นี้ / เลย / คู่ / นุ่น / + / เจ / ตต์ / ไม่ / ได้ / เพราะ / นำ / รัก / กัน / ทั้ง / คู่ /

ผลการแทนที่คำหายไปด้วยเครื่องหมายดอกจัน ไม่นำเลิกคบกันเลย ดาราคู่อื่นไม่เหมาะสมกันเท่าคู่นี้เลยคู่คู่+เจตต์ไม่ได้เพราะนำรักกันทั้งคู่

ผลการตัดคำ ทุก / คน / อย่า / โกรธ / นะ / มือ / ที่ / สาม / คือ / เรา / เอง / นะ / ล้อ / เล่น /

ผลการแทนที่คำหายไปด้วยเครื่องหมายดอกจัน ทุกคนอย่าโกรธนะ มือที่สามคือเราเองนะ ล้อเล่น

ผลการตัดคำ กัน / สวย / นำ / รัก / จัง /

ผลการแทนที่คำหายไปด้วยเครื่องหมายดอกจัน กันสวยนำรักจัง

ผลการตัดคำ คำ / นิบหาย / แล้ว / ดู / ใต้ / เสื้อ / สี / ขาว / สี / เขียว / ไม่ / ค่อย / จะ / ตัด / กับ / สี / ผิว / เลย / นะ / รู้ / ว่า / สวย / แม่ / แล้ว / ดู / สาย / เดี่ยว / สี / ขาว / ไม่ / ใต้ / ยก / ทรง / เหนือ / สงสัย / อยาก / ให้ / ผัว / เขา / เห็น / นะ /

ผลการแทนที่คำหายไปด้วยเครื่องหมายดอกจัน คำ * แล้วดูใต้เสื้อสีขาว สีเขียว ไม่ค่อยจะตัดกับสีผิวเลยนะ รู้ว่าสวยแม่ แล้วดูสายเดี่ยวสีขาว ไม่ใ้ยกทรงเหนือสงสัยอยากให้ผัว เขาเห็นนะ

ผลการตัดคำ ไอ้ / บ้า / อยู่ / ดี / ดี / ไม่ / ชอบ /

ผลการแทนที่คำหายับด้วยเครื่องหมายดอกจัน *บ้า อยู่ดีดีไม่ชอบ

ผลการตัดคำ แล้วไป / เสือก / อะไร / กับ / เรื่อง / ของ / คำ / วะ /

ผลการแทนที่คำหายับด้วยเครื่องหมายดอกจัน แล้วไป*อะไรกับเรื่องของคำ*

ผลการตัดคำ กอด / จูบ / มี / เช็ก / เป็น / เรื่อง / ธรรมดา / ของ / เด็ก / ลูก / ครั้ง / อยู่ / แล้ว /
จะ / ไป / สนใจ / อะไร / กัน / นักหนา / เรื่อง / ส่วนตัว / ของ / เขา / จะ / ดี / จะ / เลว / ช่าง / เขา /
เถอะ /

ผลการแทนที่คำหายับด้วยเครื่องหมายดอกจัน กอด จูบ มีเช็ก เป็นเรื่องธรรมดาของเด็กลูกครั้งอยู่
แล้ว จะไปสนใจอะไรกันนักหนาเรื่องส่วนตัวของเขา จะดีจะ*ช่างเขาเถอะ

ผลการตัดคำ ไม่ / เคย / ชอบ / บอล / เลย / ไม่ / ฐึ / มา / ซ่า / ขอม / คบ / ด้วย / ได้ / ใจ / ไม่ /
เห็น / เป็น / สุภาพบุรุษ / เลย / ว่า / แต่ว่า / ตอน / นี้ / มา / ซ่า / คบ / กับ / ใคร / อยู่ / ค่ะ / ? /

ผลการแทนที่คำหายับด้วยเครื่องหมายดอกจัน ไม่เคยชอบบอลเลย ไม่ฐึมาซ่าขอมคบด้วยได้ใจ ไม่
เห็นเป็นสุภาพบุรุษเลย ว่าแต่ว่าตอนนี้มาซ่าคบกับใครอยู่ค่ะ?

ผลการตัดคำ ม่าย / สาว / จอม / ฉาว / .. / นี้ / ฎ / เคย / เห็น / แฉว / ย่าน / นาวา / กับ / เลี / ย / และ /
เพื่อน / ก็ / บอก / ว่า / เคย / เห็น / ที่ / เชียงใหม่ / ใน / โรงแรม / กับ / เลี / ย / ก่อน / ที่ / จะ / เข้า /
วงการ /

ผลการแทนที่คำหายับด้วยเครื่องหมายดอกจัน ม่ายสาวจอมฉาว..นี้*เคยเห็นแฉวย่านนาวากับเลี
และเพื่อนก็บอกว่าเคยเห็นที่เชียงใหม่ในโรงแรมกับเลียก่อนที่จะเข้าวงการ

ผลการตัดคำ นึกถึง / เรื่อง / ไร้ / บอล / กับ / มา / ซ่า / แล้ว / อยาก / ถาม / มัน / ว่า / " / มี /
เป็น / ลูก / ผู้ชาย / เปลา / วะ / " /

ผลการแทนที่คำหายับด้วยเครื่องหมายดอกจัน นึกถึงเรื่อง*บอลกับมาซ่า แล้วอยากถามมันว่า "*
เป็นลูกผู้ชายเปลา*"

ผลการตัดคำ วิ / ก / หลาย / ที / ช่วย / ดู / ไป / ร / ยา / หน้อย / เอะ / เดียวนี้ / ดู / ไม่ / ได้ /
เลย / คิด / จะ / ขาย / นม / (/ ยาน / ๗ /) / หรือ / จะ / ดู / ไม่ / ได้ / เลย / นำ / เกลียด / มาก /

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน วิกหลายสี่ช่วยดูไปรยาหน่อยเถอะ เตียวนี้ดูไม่ได้
 เลยคิดจะขายนม (ยานๆ) หรือจะ ดูไม่ได้เลยน่าเกลียดมาก

ผลการตัดคำ กิน / บน / เรือน / จี้ / บน / หลังคา / จริง / ๗ / เลย / ดารา / นี้ / หา / ผล /
 ประโยชน์ / กับ / ทุก / เรื่อง / แม่ / เรื่อง / คำ / ๗ / ของ / พวก / เธอ / ๗ / ทั้งหลาย /

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน กินบนเรือนขึ้นหลังคาจริงๆ เลยดารานี้ หา
 ผลประโยชน์กับทุกเรื่อง แม่เรื่องต่างๆ ของพวกเธอๆ ทั้งหลาย

ผลการตัดคำ ยัง / เอา / มา / ขาย / ประจัน / กัน / อีก / เชื้อ / / เบื่อ /

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน ยังเอามาขายประจันกันอีก เชื้อ....เบื่อ

ผลการตัดคำ ไม่ / รับผิดชอบ / งาน / เลย / นะ / ซ้อ / ป่านนี้ / ละ / เดี่ยว / ก้อ / โคน / ไล่ออก /
 หรอก /

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน ไม่รับผิดชอบงานเลยนะซ้อ ป่านนี้ละ เดี่ยวก้อ
 โคนไล่ออกหรอก

ผลการตัดคำ ซ้อ / จำ / 5 / โมง / แล้ว / นะ / ยัง / ไม่ / มา / อีก / เหรอ / รอ / อยู่ / นะ / จะ /

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน ซ้อจำ 5 โมงแล้วนะยังไม่มาอีกเหรอรออยู่นะจะ

ผลการตัดคำ นางเอก / หน้าอ / เป็น / เมียน้อย / ผู้ / จัด / ด้วย / เหรอ / ฝั้ว / เยอะ / จัง /

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน นางเอกหน้าอเป็นเมียน้อยผู้จัดด้วยเหรอ ฝั้วเยอะ
 จัง

ผลการตัดคำ ทุก / อย่าง / ที่ / เกิด / ขึ้น / เป็น / เรื่อง / ระหว่าง / คน / 2 / คน / ใคร / ทำ /
 อะไร / ย่อม / รู้ / แก่ / ใจดี / แต่ / การ / เอา / คน / ที่ / เคย / รัก / มา / พุด / ให้ / คน / อื่น / ฟัง / มัน
 / เป็นการ / กระทำ / ที่ / ผู้ชาย / ดี / ๗ / เขา / ไม่ / ทำ / กัน / หนังสือ / ที่ / จะ / ออก / มา / นี้ / คิด
 / ให้ / ดี / ก่อน / คง / จะ / มี / คน / ค่า / มาก / กว่า / มี / คน / ชม /

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน ทุกอย่างที่เกิดขึ้นเป็นเรื่องระหว่างคน 2 คน ใครทำอะไรย่อมรู้แก่ใจดี แต่การเอาคนที่เคยรักมาพูดให้คนอื่นฟังมันเป็นการกระทำที่ผู้ชายดีๆ เขาไม่ทำกัน หนังสือที่จะออกมานี้คิดให้ดีกว่าก่อน คงจะมีคนค่ามากกว่ามีคนชม

ผลการตัดคำ นี้ / ๑ / ซ้อ / มา / เกอะ / จะ / เข้า / ห้องน้ำ / ก็ / ไม่ / เข้า / รอ / อยู่ / นะ / ปวด / มาก / ๑ / แต่ / อยาก / อ่าน /

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน นี้ ๑ ซ้อ มาเกอะ จะเข้าห้องน้ำก็ไม่เข้า รออยู่นะปวดมากๆ แต่อยากอ่าน

ผลการตัดคำ อยาก / ให้ / ยาวชน / ฝึก / และ / ศึกษา / การ / ใช้ / ภาษา / ไทย / จาก / ซ้อ / นะ / ครับ / อาจ / จะ / ฝึก / เขียน / คล้าย / ของ / ซ้อ / แล้ว / ให้ / เพื่อน / ๑ / ใน / Web / ทาย / กัน / ผม / คิด / ว่า / จะ / ช่วย / เสริม / สร้าง / สิ่ง / ดี / ๑ / นะ / ครับ /

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน อยากให้ยาวชน ฝึกและศึกษา การใช้ภาษาไทย จาก ซ้อ นะครับ อาจจะมีฝึกเขียน คล้ายของซ้อ แล้วให้เพื่อนๆ ใน Web ทายกัน ผมคิดว่าจะช่วยเสริมสร้าง สิ่งดีๆ นะครับ

ผลการตัดคำ ช่วย / แจ้ง / เจ้าแม่ / ช่อง / หลาย / สี / ด้วยว่า / ดารา / ของ / ท่าน / ทำ / ไม่ / ชอบ / โชว์ / นม / จัง / เลย / โดย / เฉพาะ / ดารา / หน้า / ใหม่ / อย่าง / ปู / ไป / รยา / เพราะ / เด็ก / ๑ / ๑ / ดู / แล้ว / รู้สึก / เป็น / ตัวอย่าง / ที่ / ไม่ / ดี / เลย / เดี่ยว / โชว์ / โน่น / โชว์ / นี้ / สารพัด / ไม่ / รู้ / จะ / โชว์ / ทำ / ไหม / ดู / อย่าง / นางรอง / ที่ / ชื่อ / सा / เจ้า / เรียบร้อย / นำ / รัก / ดู / แล้ว / สวย / กว่า / ต้อง / เยอะ / ทำ / มั้ย / ไม่ / ทำ / แบบ / เจ้า / บ้าง / ควร / จะ / สรร / หา / เนื้อหา / ละคร / ที่ / เป็น / การ / ส่งเสริม / วัฒนธรรม / หรือ / เป็น / ตัวอย่าง / ที่ / ดี / เหมือนกับ / ช่อง / 3 / บ้าง / เพราะ / เจ้า / เน้น / ละคร / ที่ / ส่งเสริม / วัฒนธรรม / หรือ / มี / อะไร / ที่ / นำ / ดู / กว่า / เยอะ / เดี่ยว / นี้ / ละคร / ช่อง / 7 / ไม่ / ค่อย / มี / ละคร / ที่ / ดี / ๑ / เลย / ไม่ / รู้ / ว่า / หมด / ความ / สามารถ / แค / นี้ / เหรอ / ฝาก / ด้วย / นะ / คน / ไทย /

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน ช่วยแจ้งเจ้าแม่ช่องหลายสีด้วยว่า ดาราของท่าน ทำไม่ชอบโชว์นมจังเลย โดยเฉพาะดาราน้ำใหม่อย่าง ปู ไปรยา เพราะเด็กๆ ดูแล้วรู้สึกเป็นตัวอย่างที่ไม่ดีเลย เดี่ยวโชว์โน่นโชว์นี้สารพัดไม่รู้จะโชว์ทำไหม ดูอย่างนางรองที่ชื่อสา เจ้า เรียบร้อย นำรัก ดูแล้วสวยกว่าต้องเยอะ ทำมั้ยไม่ทำแบบเจ้าบ้าง ควรจะสรรหาเนื้อหาละครที่เป็น การส่งเสริมวัฒนธรรม หรือเป็นตัวอย่างที่ดีเหมือนกับช่อง 3 บ้าง เพราะเน้นละครที่ส่งเสริม

วัฒนธรรม หรือมีอะไรที่น่าดูกว่าเยอะ เคี้ยวนี้ละครช่อง 7 ไม่ค่อยมีละครที่ดีๆ เลย ไม่รู้ว่าหมดความสามารถแค่นี้หรือ ผากด้วยนะ คนไทย

ผลการตัดคำ เจ้/มี/พฤติกรรม/ของ/ อ้อม/ สุ/นิ/สา/ สุข/บุญ/สังข์/มา/รายงาน/ บ้าง/ไหม/ ไม่ค่อย/เห็น/เป็น/ข่าว/เลย/ หัว/จริง/หรือ /

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน เจ้มีพฤติกรรมของ อ้อม สุนิสา สุขบุญสังข์มา รายงานบ้างไหม ไม่ค่อยเห็นเป็นข่าวเลย หัวจริงหรือ

ผลการตัดคำ กำ/ลั้ง/ร้อน/นำ/ไป/ทิ้ง/วงการ/เลย/ตอน/นี้/ เรื่อง/ที่/หนุ่ม/มาด/เซ/ อร์/ " /บอล/ " / อธิป/ นานา/ จะ/ออก/พี่/อก/เก้/ตบู้ค/ แฉ/อดีต/คน/รัก/ มา /ชำ/ วัฒน/พา/นิ/ช/ นัก/ร้อง/และ/ดารา/เซ็ก/ซี่/ตลอด/กาล/

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน กำลั้งร้อนนำไปทิ้งวงการเลยตอนนี้ เรื่องที่หนุ่มมาดเซอ "บอล" อธิป นานา จะออกพี่อกเก้ตบู้ค แฉอดีตคนรัก มาชำ วัฒนพานิช นักร้องและดารา เซ็กซี่ตลอดกาล

ผลการตัดคำ พี่/อก/เก้/ตบู้ค/ของ/หนุ่ม/บอล/ยัง/ไม่/รู้/หน้าตา/เป็น/ยัง/ไง/ รู้/ แต่/ว่า/ดั่ง/กระหึ่ม/เมือง/ไป/แล้ว/ เมื่อ/หนุ่ม/บอล/ออก/มา/แฉ/และ/เอียด/ยิบ/ใน/ รายการ/ดี/ลิป/ของ/เสี่ย/ย/ที/วี/ วิทวัส/ สุนทร/วินต/ร/ เมื่อ/คำ/วัน/ที่/ 19/ เมษายน/ ที่/ผ่าน/มา/

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน พี่อกเก้ตบู้คของหนุ่มบอลยังไม่รู้หน้าตาเป็นยังไร รู้แต่ว่าดั่งกระหึ่มเมืองไปแล้ว เมื่อหนุ่มบอลออกมาแฉและเอียดยิบในรายการดีลิปของเสี่ยทีวี วิทวัส สุนทรวินต เมื่อคำวันที่ 19 เมษายน ที่ผ่านมา

ผลการตัดคำ อยาก/ให้/พูด/ถึง/อีก/บ/เกลียด/มัน/ กระเดะ/ แม่/บ/รู/คดี/ด้วย/ แล้ว/หรือ/ จี้/คุย/ บ/รู/ค/ยัง/พูด/เลย/ใน/รายการ/ของ/สัญญา/ว่า/กบ/ให้/ซื้อ/ ของ/ให้/แล้ว/ออกอากาศ/ว่า/ประมาณ/หลาย/บาท/นะ/ มัน/จ้อง/จะ/จับ/ซัด/ๆ/ ทำ/ไม/มัน/ถึง/มี/แพ/นค/ลับ/หน้า/มิด/ตา/มัว/กัน/อยู่/ได้/

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน อยากให้พูดถึงอีกบเกลียดมัน * แม่บรูคดีด้วยแล้วหรือ จี้คุย บรูคยังพูดเลยในรายการของสัญญาว่ากบให้ซื้อของให้แล้วออกอากาศว่าประมาณหลายบาทนะ มันจ้องจะจับซัดๆ ทำไมมันถึงมีแพนคลับหน้ามิดตามัวกันอยู่ได้

ผลการตัดคำ ตอน / นี้ / อยาก / ให้ / ซื่อ / พูด / ถึง / สาม / โส / โส / ให้ / ฟัง / หน่อย / อยาก / รู้ /
ว่า / ซื่อ / คิด / ยัง / ไง /

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน ตอนนี้อะอยากให้ซื่อพูดถึงสามโสโสให้ฟังหน่อย
อยากรู้ว่าซื่อคิดยังไง

ผลการตัดคำ โดย / เฉพาะ / ฉาก / ล้วง / ฉาก / เต็ม / กอด / นม / อี / สาว / ไซค์ / โลน / ของ / นัก
/ ร้อง /

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน โดยเฉพาะฉากล้วง ฉากเต็มกอดนม*สาวไซค์โลน
ของนักร้อง

ผลการตัดคำ อี / สาว / ชื่อ / ดัง / หวัง / รวย / ทาง / ลัด /

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน *สาวชื่อดังหวังรวยทางลัด

ผลการตัดคำ อีก / บ / มั่น / ก็ / ชอบ / เขียน / แต่ / สิ่ง / ไร / สาระ / และ / ยุ้ง / ชาว / บ้าน / ให้ /
ทะเลาะ / กัน /

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน อีกบมั่นก็ชอบเขียนแต่สิ่งไรสาระ และยุ้งชาวบ้าน
ให้ทะเลาะกัน

ผลการตัดคำ อีก / บ / มั่น / คือ / ใคร / เสว / จริง / ๆ /

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน อีกบมั่นคือใคร *จริงๆ

ผลการตัดคำ ถ้า / พวก / มี / มี / ปัญหา / มา / เจอ / ฤ / ได้ / ฤ / ด / ราชนูรี /

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน ถ้าพวก*มีปัญหา มาเจอ*ได้ ฤดราชนูรี

ภาคผนวก ฉ

ตัวอย่างข้อความจากข่าวหนังสือพิมพ์ที่ใช้ทดสอบ ประสิทธิภาพของโปรแกรมในงานวิจัย

‘เจ้าทุกข์’ โผล่เพียบ แท้กซ์หื่น! โดนชี้ตัว-หลายคดี

ความคืบหน้าคดีแท็กซี่ฆ่าภรรยา หนี น.ส.อารีรัตน์ พวงทอง อายุ 29 ปี สาวพริตตี้บาร์ ร้านชพราโน ซอย สุขุมวิท 26 เรียกกรณแท็กซี่เขียวเหลือง ยี่ห้อมิตซูบิชิ ทะเบียน มค 4973 กรุงเทพมหานคร จากปากชอชยลาตพริ้ว 95 ให้ไปส่งศูนย์การค้า สยามดิศคัฟเวอรี ย่านปทุมวัน แต่เมื่อรถวิ่งถึงสะพานหัวช้าง เขตราชเทวี กทม.

โชเฟอร์แท็กซี่ออกฉายโจร จอดรถทำที่ขอลงไปปีสสาวะ เสร็จแล้วชักมีดจี้จับมัดมือ โผล่หลัง ก่อนปลดทรัพย์สิน บังคับให้นั่งเบาะหน้ารถ แล้วขับมุ่งหน้าไปตามถนนเพชรบุรี แต่เมื่อถึงแยกอรุณพงษ์ รถติดสัญญาณไฟแดง ผู้เสียหายตัดสินใจเปิดประตูรถแล้วถ่วงตัวลงจากรถ หนีเอาตัวรอดมาได้ หวาดหวิด ส่วนคนร้ายขับรถหลบหนีไป ต่อมาเจ้าหน้าที่ตำรวจ สน.พญาไท ติดตามจับกุมโชเฟอร์โจรได้ รับสารภาพว่าก่อคดีมาแล้วโชกโชก โดยเลือกเหยื่อผู้โดยสารผู้หญิงหน้าตาดี ปลดทรัพย์สินแล้วพาไปข่มขืนมาแล้วหลายรายนั้น

เมื่อเวลา 11.30 น. วันที่ 31 ต.ค. พล.ต.ต.ปราโมช ปทุมวงศ์ ผบก.น.1 พ.ต.อ.การุณ กลิ่นประยูร ผกก.สน. พญาไท พ.ต.ท.สมิง รอดรัศยะ สว.สส. ได้ควบคุมตัวนายณรงค์ หรือบังซูป ปั้นหิ อายุ 40 ปี โชเฟอร์แท็กซี่ภัยสังคม ผู้ต้องหา ไปทำแผนประทุษกรรมประกอบคำรับสารภาพบริเวณเชิงสะพานหัวช้าง ถนนพญาไท แขวงถนนเพชรบุรี เขตราชเทวี กทม. จุดที่นายณรงค์จอดรถทำที่ไปปีสสาวะ เสร็จแล้วชักมีดจี้จับมัดผู้เสียหาย พร้อมกับปลดทรัพย์สิน ส่วนจุดที่ 2 บริเวณแยกอรุณพงษ์ จุดที่ผู้เสียหายเปิดประตูรถแล้วถ่วงตัวลงจากรถแท็กซี่ ขณะจอดติดสัญญาณไฟแดง ใช้เวลาทำแผนทั้งสองแห่ง ประมาณครึ่งชั่วโมงจึงเสร็จสิ้น

ต่อมาได้มี น.ส.ธิดา เศิกสงบ อายุ 27 ปี พนักงานขายจิวเวลรี่ย่านสี่พระยา อยู่บ้านเลขที่ 52/203 หมู่บ้านกฤษณา ถนนกรุงเทพกรีฑา แขวงและเขตสะพานสูง กทม. ผู้เสียหายเหยื่อแท็กซี่ฆ่าภรรยา เข้าพบพนักงานสอบสวนเจ้าของคดี โดยนำบันทึกประจำวันที่แจ้งความไว้ สน.ประเวศ เหตุเกิดเมื่อเวลา 08.00 น. วันที่ 10 ต.ค. ที่ผ่านมา เพื่อชี้ตัวคนร้ายเพิ่มเติม เหยื่อสาวผู้เสียหายให้รายละเอียดว่า

เรียกรถแท็กซี่เที่ยวเหลืองจากหน้าหมู่บ้านกฤษณาให้ไปส่งย่านสี่พระยา แต่เมื่อรถผ่านบริเวณถนน
 คู่ขนานมอเตอร์เวย์ ใกล้สุเหร่าหัวหมากใหม่ แขวงและเขตสวนหลวง กทม. โชเฟอร์ยัดข้าวได้จอด
 รถอ้างว่าขอไปตรวจเช็ควาล์วแก๊สด้านหลังรถ แต่กลับเปิดประตูซ็อกมีดปลายแหลมจี้บังคับปลด
 ทรัพย์สินได้โทรศัพท์มือถือยี่ห้อฟิลิป รูนจิ 790 และเงินสด 1,500 บาท พร้อมกับใช้เชือกมัดมือไพล่
 หลัง แต่หลังจากขับรถไปได้ไม่กี่ไกล ตนแก้มัดมือจนเป็นอติสภาพ เสร็จแล้วพยายามแย่งมีดจาก
 คนร้าย แต่โชเฟอร์โจร จอดรถแล้วแย่งมีดกลับคืนไปได้

สาวเหยื่อแท็กซี่โจรกรุงให้รายละเอียดต่อว่า พยายามทำใจดีสู้เสือ วิงวอนร้องขอชีวิต ว่าได้
 ทรัพย์สินไปแล้วอย่าทำอะไร ทำให้โชเฟอร์แท็กซี่ใจอ่อนขว้างมีดทิ้ง แต่ก็ยังมัดมือพาขึ้นรถขับไป
 ต่อ โดยขับอ้อมเข้าไปในซอยคาราบาว ถนนกรุงเทพกรีฑา ซึ่งเป็นเส้นทางเปลี่ยว เสร็จแล้วจอดรถ
 เพื่อจะทำนิติมิร้าย จึงอาศัยจังหวะที่คนร้ายไม่ทันตั้งตัว รวบรวมสติพละกำลังเปิดประตูรถวิ่งหนี
 อย่างไม่คิดชีวิต โดยไม่ทันได้จำหน่ายเลขทะเบียนรถ ทราบแต่ว่าเป็นแท็กซี่เที่ยวเหลือง กระทั่งเมื่อ
 เห็นหน้าของคนร้ายทางหนังสือพิมพ์ จึงจำได้ว่าเป็นคนเดียวกับที่ก่อเหตุ

ต่อมาเจ้าหน้าที่ได้นำตัวนายณรงค์ออกมาให้เหยื่อสาวรายนี้ชี้ตัว น.ส.ธิดา ยืนยันว่าจำหน้าได้อย่าง
 แม่นยำเพราะว่านั่งอยู่ในรถด้วยกันนานกว่าชั่วโมงเศษ เจ้าหน้าที่จึงอายัดตัวและแจ้งข้อหาชิงทรัพย์
 ในเวลากลางคืนโดยมีอาวุธ โดยใช้ยานพาหนะและหน่วงเหนี่ยว หรือกักขังผู้อื่น หรือกระทำความ
 ประการใดให้ผู้อื่นปราศจากเสรีภาพในร่างกาย เพิ่มอีกคดี โดยล่าสุดทาง สภ.อ.เมืองสมุทรปราการ
 กำลังขออนุมัติหมายจับมาอายัดตัวเพิ่มอีกราย

ด้าน พล.ต.ต.ปราโมช ปทุมวงศ์ ผบ.ก.น. 1 เปิดเผยว่า ในวันพรุ่งนี้จะนำตัวผู้ต้องหาไปฝากขังและ
 คิดว่าคงต้องนำตัวกลับมาควบคุม ต่อเพื่อขยายผลและให้ผู้เสียหายชี้ตัวยืนยัน เพราะน่าจะมีผู้ที่ตก
 เป็นเหยื่ออีกหลายราย สำหรับรถแท็กซี่คันที่ใช่ก่อเหตุ ผู้ต้องหาเข้ามาจากอู่ เจ้แดง ในซอย
 พัฒนาการ 64 โดยเช่าเหมา 24 ชั่วโมง เป็นรถรุ่นเก่า เจ้าของอยู่ตลิ่งใต้ไปชำระค่าเช่าอาทิตย์ละครั้ง
 โดยคิดวันละ 800 บาท วันอาทิตย์ขับฟรี 1 วัน

วันเดียวกัน พ.ต.ท.กิตติพัทธ์ เทียงน้อย สว.สส.สน. ลาดพร้าว ได้ติดต่อ น.ส.นินา (นามสมมติ) อายุ
 20 ปี นักศึกษาชั้นปี 1 มหาวิทยาลัยอัสสัมชัญ หรือเอแบค ลูกสาวของสมาชิกสภาจังหวัดสุราษฎร์
 ธานี และเป็นหลานสาวนายสุเทพ เทือกสุบรรณ เลขาธิการพรรคประชาธิปัตย์ ผู้เสียหายอีกราย ให้
 ยืนยันตัวโชเฟอร์แท็กซี่มหาภัยรายนี้ หลังจากเมื่อคืนวันที่ 12 ต.ค. ที่ผ่านมา เวลาประมาณตีสอง
 น.ส.นินาตกเป็นเหยื่อ โดยเรียกรถแท็กซี่จากเอแบคคอนโด หลังมหาวิทยาลัยรามคำแหง ให้ไปส่ง
 ในซอยลาดพร้าว 35 เพื่อหาเพื่อน แต่ปรากฏว่าเมื่อผ่านหน้าร้านอาหารแม่ลา ปลาเผา กลางซอย
 ลาดพร้าว 130 แขวงคลองจั่น เขตบางกะปิ คนขับรถแท็กซี่ได้จอดรถ อ้างว่าขอลงไปทำธุระ จากนั้น
 เดินอ้อมหลังรถไปเปิดประตูหลัง ใช้มือจิกหัวแล้วกระชากโทรศัพท์มือถือถือกับกระเป๋าสตางค์มีเงิน

สด 3,300 บาท ก่อนที่เหยื่อสาวจะสะบัดตัว เปิดประตูรถอีกด้านวิ่งหลบหนีได้หวุดหวิด โดยผู้เสียหายบอกว่า เห็นรูปของคนร้ายแล้ว ยืนยันว่าเป็นคนเดียวกัน แต่ตอนก่อเหตุผมยาวกว่าในขณะนี้ ทางตำรวจจึงประสานให้เข้าพบพนักงานสอบสวน สน.ลาดพร้าว เพื่อสอบปากคำเพิ่มเติม ก่อนขออนุมัติหมายจับและอายัดตัวผู้ต้องหาต่อไป

ส่วนคดีแท็กซี่ราคน้ำมันแล้วจุดไฟเผาสาวยังไม่ทราบชื่ออย่างใดเห็นก่อนไปสิ้นใจที่โรงพยาบาล เหตุเกิดในซอยพระรามสอง 38 แยก 3 ถนนพระรามสอง แขวงและเขตจอมทอง ฝั่งธนฯ โดยตำรวจยังสืบสนชื่อ “จำปี มีชัย” ที่ผู้ตายบอกก่อนหมดลมหายใจ เนื่องจากมีอยู่จำนวนมากและยังมีชีวิตอยู่นั้น

ความสืบหน้าเมื่อเวลา 11.00 น. วันที่ 31 ต.ค. ที่ กก.สส.บก.น.8 พ.ต.อ.จิรศักดิ์ ขำคง ผกก.สส.น.8 ได้ ประชุมชุดสืบสวนเพื่อคลี่คลายคดี จากนั้นเปิดเผยว่า ได้ เรียกฝ่ายสืบสวนที่ออกไปหาข่าว กลับมาหารือเกี่ยวกับประเด็นข้อมูลที่ได้มาในเบื้องต้น โดยเฉพาะชื่อ “จำปี มีชัย” อายุ 35 ปี ชาว จ.ร้อยเอ็ด ทางเจ้าหน้าที่ตรวจสอบรายชื่อและนามสกุลอย่างละเอียด รวมทั้งคำที่ใกล้เคียงแล้ว ปรากฏว่ามีอยู่เป็นร้อยชื่อ ซึ่งจากการสุ่มตัวอย่างทุกคนยังมีชีวิตอยู่ทั้งสิ้น ทำให้การตามหาว่าผู้ตายเป็นใคร เริ่มริบหรี่ ส่วนที่พยานเห็นว่าสาวผู้ตายลงมาจากรถ ล้อเล็ก หรือรถกะบ้อ โดยใส่เสื้อสีจืดคาดยี่ห้อยี่ห้อโกลฟ มีรูปมือที่หน้าอก ก่อนเรียกรถแท็กซี่หายไป ในซอยดังกล่าว เจ้าหน้าที่ฝ่ายสืบสวนตรวจสอบแล้วพบว่าผู้หญิงคนนั้นยังมีชีวิตอยู่ โดยพักอยู่การเคหะชุมชนพระราม 2 ทำให้ ต้องเริ่มต้นค้นหาคนใหม่

ผกก.สส.น.8 กล่าวด้วยว่า จากการวิเคราะห์สภาพศพในเบื้องต้น เหยื่อสาวสูงประมาณ 150 ซม.เศษ หน้าตาดูแล้วไม่น่าจะถึง 35 ปี ส่วนชื่อที่ผู้ตายบอกก่อนตายว่า “จำปี มีชัย” อายุ 35 ปี ชาว จ.ร้อยเอ็ด หลังจากเจ้าหน้าที่ตรวจสอบชื่อของผู้หญิงไปแล้วไม่น่าจะใช่ ได้หันมาตรวจ สอบรายชื่อของผู้ชายบ้าง เพราะอาจจะป็นชื่อของไซเฟอร์ แท็กซี่ที่ก่อเหตุก็เป็นได้ นอกจากนี้ยังให้ชุดสืบสวนตรวจสอบหญิงขายบริการทางเพศในละแวกเกิดเหตุ ย่านถนนพระราม 2 วงเวียนใหญ่ พระประแดง ว่ามีสาวขายบริการคนใดหายตัวไปบ้าง

สำหรับข้อมูลที่พยานระบุว่า เห็นรถแท็กซี่เขียวเหลืองออกมาจากซอยเกิดเหตุ ตรวจสอบแล้วพบว่า มีอยู่ 2 ยี่ห้อ คือ โตโยต้า และนิสสัน แต่ชุดสืบสวนประมวลจากคำบอกเล่าของพยานแล้วเชื่อว่า น่าจะเป็นรถนิสสัน ซาลูน ที่มาน้ำสีฟ้าติดด้านหลังมากกว่า ขณะนี้อยู่ระหว่างตรวจสอบตามอุ้งรถแท็กซี่ต่างๆ ย่านบางปะกอก บางปะแก้ว และราษฎร์บูรณะให้ชัดเจน นอกจากนี้กำลังรอผลการตรวจ พิสูจน์ของแพทย์ รพ.ศิริราช โดยเจ้าหน้าที่ฝ่ายสืบสวนได้ทำหนังสือขอให้ตรวจสอบอย่างละเอียด โดยเฉพาะในกระเพาะอาหารและเลือดให้มากที่สุดเท่าที่จะทำได้

ศาลจำคุกแม่เฒ่า 2 ปี ส่งสาวค้ากามมาเลย

ที่ศาลอาญากรุงเทพใต้ วันที่ 1 พ.ย. ศาลอ่านคำพิพากษาในคดีนางสาวศิริณี เนื่องสมศรี อายุ 34 ปี บ้านเดิมอยู่ที่ ต.ขามเฒ่าพัฒนา อ.กันทรวิชัย จ.มหาสารคาม เป็นจำเลยฐานเป็นธุระจัดหาหญิงเพื่อสนองความใคร่ของผู้อื่น คดีนี้จำเลยให้การรับสารภาพมาตลอด และยื่นหลักทรัพย์ 4 แสนบาท เพื่อขอประกันตัวระหว่างพิจารณาแต่ศาลไม่อนุญาต คดีนี้โจทก์ฟ้องว่า ตำรวจกองสวัสดิภาพเด็กและเยาวชนสืบทราบว่า มีคนไทยทำตัวเป็นเอเยนต์ส่งหญิงสาวหน้าตาดี ไปทำงานขายบริการทางเพศ ที่ร้านอาหารชื่อลีซิงคาราโอเกะ อยู่ในรัฐยะโฮร์ ประเทศมาเลเซีย โดยให้หญิงไทยขายตัววันละ 3 รอบ ได้ค่าตอบแทนรอบละ 1,000 บาท แต่มีข้อแม้ว่าต้องทำงานครบ 2 ปี ตำรวจสดจึงประสานไปยังกองตำรวจรถไฟ วางแผนจับกุม กระทั่งวันที่ 9 ธ.ค. 2547 เวลา 20.00 น. ตำรวจสามารถดักจับกุมนางสาวศิริณีได้บนขบวนรถไฟสายกรุงเทพฯ-หาดใหญ่ ตรวจค้นตัวพบตัวรถไฟ 2 ใบ หนังสือเดินทาง 2 เล่ม โทรศัพท์มือถือ และบัญชีเงินสดของ ธนาคารพาณิชย์ นางสาวศิริณีรับสารภาพว่ากำลังจะพาหญิงสาวไปส่งที่ประเทศมาเลเซียจริง ศาลพิเคราะห์แล้วเห็นว่าโจทก์มีพยานมาสืบฟังได้ว่านางสาวศิริณีเป็นหนึ่งในขบวนการส่งหญิงสาวคนไทยไปขายบริการที่ประเทศมาเลเซียจริง จึงมีความผิดฐานเป็นธุระจัดหาหญิงเพื่อสนองความใคร่ผู้อื่น พิพากษาลงโทษจำคุก 4 ปี คำรับสารภาพเป็นประโยชน์คงรับโทษจำคุก 2 ปี หลังฟังคำพิพากษานางสาวศิริณีได้ถูกคุมตัวไปขังที่ทัณฑสถานหญิง แม้พยายามยื่นคำร้องขอประกันตัว แต่ศาลไม่อนุญาต โดยให้เหตุผลว่าจำเลยมีพฤติกรรมร้ายแรงเป็นภัยต่อสังคม

ล่อซื้อจับคาโรงแรม แม่พาลูกสาววัย 15 ค้ากาม

เมื่อเวลา 19.00 น. วันที่ 1 พ.ย. พ.ต.อ.เรวัช กลิ่นเกษร รอง ผบก.กลุ่มงานสืบสวนสอบสวนภาค 7 พ.ต.อ.สมชาย รักเสนาะ ผกก.กลุ่มงานสืบสวนสอบสวน ภาค 7 พ.ต.ท.ปรีดา อิมเจริญ สว.งาน 3 กก.สส.ภ. 7 สืบทราบว่า มีผู้หญิงพาลูกสาว กับเพื่อนของลูกเร่ขายบริการทางเพศ จึงวางแผนให้สายโทรศัพท์ ติดต่อล่อซื้อบริการ โดยนัดหมายกันที่ห้องเลขที่ 314 และ 315 โรงแรมคันทรี่ ถนนสายตลิ่งชัน-สุพรรณบุรี หมู่ 3 ต.ท่าระหัด อ.เมืองสุพรรณบุรี

ก่อนถึงเวลานัดหมาย ตำรวจได้กระจายกำลังประจำจุดต่างๆ พร้อมถ่ายสำเนาธนบัตรสำหรับจ่ายเป็นเงินค่าตัวสาวบริการ 2 คน คนละ 2,000 บาท รวม 4,000 บาท ไว้เป็นหลักฐาน พร้อมเตรียมเทปบันทึกการสนทนาระหว่างสายล่อซื้อกับกลุ่มผู้ค้าบริการ จนกระทั่งถึงเวลานัดหมาย ได้มีผู้หญิงที่เป็นแม่เฒ่า ทราบชื่อนางตุ๊ก (นามสมมติ) อายุ 40 ปี บ้านอยู่ ต.กฤษณา อ.บางปลาหมอ จ.สุพรรณบุรี ได้พา น.ส.เอ (นามสมมติ) อายุ 18 ปี ลูกสาว และ น.ส.บี (นามสมมติ) อายุ 15 ปี เพื่อนของ น.ส.เอ

มาส่งให้สายล่อซื้อที่ห้องหมายเลข 314 และ 315 ของโรงแรมคันทรี หลังจากจ่ายเงินค่าตัวเป็นที่เรียบร้อย สายล่อซื้อได้ส่งสัญญาณให้ตำรวจโจมเข้าจับกุม น.ส.เอ และ น.ส.บี ได้ภายในห้อง ขณะที่กำลังอีกส่วนเข้าจับกุมนางตุ๊กได้ที่ล็อบบี้ของโรงแรม

ทันทีที่ถูกจับกุม นางตุ๊กให้การปฏิเสธว่าไม่ได้เป็นคนพาลูกสาวกับเพื่อนลูกสาวมาค้าบริการ แต่ทั้งคู่มาชายบริการกันเอง ตำรวจจึงนำเทปบันทึกการสนทนาที่เปิดให้ฟัง ในที่สุดก็ยอมจำนนรับสารภาพว่า พาลูกสาวกับเพื่อนของลูกมาขายบริการทางเพศหลายครั้งเป็นเวลานานนับปี โดยเงินค่าตัวของลูกสาวจะให้เต็มทั้ง 2,000 บาท ส่วนเงินค่าตัวของ น.ส.บี จะหัก 1,000 บาท เป็นค่านายหน้าคิดต่อลูกค้า เมื่อตำรวจสอบถามถึงลูกค้าที่เคยซื้อบริการ นางตุ๊กปิดปากเงียบไม่ยอมบอกอ้างว่าไม่รู้จักและจำไม่ได้แล้ว แต่ส่วนใหญ่เป็นพวกเสี้ยมกระเป๋านัก

ด้าน น.ส.เอเปิดปากรับสารภาพว่า ดิฉันบ้ามา นานนับปีและไม่ได้เรียนหนังสือ ที่ให้แม่พามาขายบริการ ทำไปด้วยความเต็มใจ ไม่ได้ถูกบีบบังคับแต่อย่างใด ต้องการหาเงินไปซื้อยาบ้าเสพ และหาซื้อเสื้อผ้าสวยๆ ใส่ไปเที่ยวเตร่ ส่วน น.ส.จอยให้การว่า ไม่ได้เรียนหนังสือเช่นกัน สาเหตุที่ออกมาขายบริการทางเพศเพื่อต้องการ หาเงินซื้อเสื้อผ้าและไปเที่ยวกับเพื่อนๆ หลังสอบสวนเบื้องต้นควบคุมตัวผู้ต้องหาทั้ง 3 คน ส่ง สภ.อ.เมืองสุพรรณบุรี ดำเนินคดีต่อไป

ฆ่าโหด-ทุบสาวถูกคลุมหัวศพหมกพงหญ้า

เมื่อเวลา 08.00 น.วันที่ 1 พ.ย. พ.ต.ต.สุวฤทธิ์ พร้อมมูล สารวัตรเวร สภ.ย่อยปฐมพร อ.เมืองชุมพร ได้รับแจ้งเหตุพบศพหญิงสาว ถูกฆาตกรรม ในพงหญ้าริมถนนสายสระชัย-ถ้ำแก้ว หมู่ 12 บ้านสระชัย ต.บ้านนา อ.เมืองชุมพร จึงรายงานให้ พ.ต.อ.นรินทร์ บุญวิทย์ ผกก. พ.ต.ท.โชคดี รักรงค์ สวป. พ.ต.ต.เกรียงไกร โพธิ์ประสิทธิ์ สว.สส. นำกำลังตำรวจ พร้อมแพทย์เวร รพ.ชุมพรเขตอุดมศักดิ์ และหน่วยกู้ภัยสายชลชุมพร ไปตรวจสอบ

ที่เกิดเหตุอยู่ใกล้สวนทุเรียนของชาวบ้าน เป็นคูน้ำแห้งขุดมีพงหญ้าขึ้นปกคลุม พบศพเหยื่อฆาตกรรมเป็นหญิงสาวไม่ทราบชื่อ อายุประมาณ 20-30 ปี รูปร่างอวบ ผิวขาว ตัดผมชอยสั้นย้อมสีแดง หูทั้ง 2 ข้างเจาะข้างละ 2 รู ใส่ตุ้มหูทองทั้ง 4 รู นี้วนางซ้ายสวมแหวนทอง 1 วง นุ่งกระโปรงสั้นสีน้ำตาลอ่อน สวมเสื้อแขนกุดสีขาว คอระบายลายลูกไม้ ถูกถลกขึ้นมาได้รวมม ในเสื้อชั้นในมีเงินสด 140 บาทซุกอยู่ นอกนั้นไม่พบหลักฐานว่าเป็นใครมาจากไหน สภาพศพนอนคุดคู้คว่ำหน้า หุ่นที่บังแดดรถยนต์สีดำ มีถุงพลาสติกหุ้มหัวของห้างสรรพสินค้า สีฟ้า ที่ถุงเขียนด้วยปากกาเมจิกว่า “ภาวิณี วงศ์ตันสกุล” และเขียนหมายเลข 607 1 ตัว 30 บาท คลุมศีรษะอยู่ แพทย์ชันสูตรศพเบื้องต้นพบบาดแผลถูกทุบด้วยของแข็งที่ใบหน้าและเท้าเดิมไม่ได้ หน้าผากยุบ และ

โทหนักแก้มขวามีแผลแตกลึกถึงกระดูก เลือดแห้งเกรอะกรัง คาดว่าเสียชีวิตมาแล้วไม่ต่ำกว่า 10 ชั่วโมง ยังไม่สามารถตรวจหาร่องรอยการถูกข่มขืนได้ จึงนำศพส่ง รพ.ชุมพรเขตอุดมศักดิ์ ชั้นสูตรอย่างละเอียดอีกครั้ง ส่วนบริเวณโดยรอบที่เกิดเหตุไม่มีร่องรอยการต่อสู้

จากการสอบถามชาวบ้านละแวกใกล้เคียงทราบว่า เมื่อตอนหัวค่ำวันที่ 31 ต.ค. ขณะฝนตกหนัก ได้ยินเสียงรถยนต์ขับเข้าไปจอดบริเวณจุดที่พบศพ สักพักก็รีบร้อนขับออกไป แต่ไม่มีใครสนใจ จนรุ่งเช้าจึงไปพบศพถูกฆาตกรรมโหด จากสภาพการแต่งกายของเหยื่อสาว ตำรวจสันนิษฐานว่า น่าจะเป็นนักร้องหรือหญิงบริการตามห้องอาหารหรือคาเฟ่ในตัวเมืองชุมพร ถูกคนร้ายฆ่าจากที่อื่นแล้ว นำศพใส่รถมาทิ้งอำพรางคดี โดยชนวนเหตุมาจากเรื่องชู้สาวหรือไม่ก็ความโกรธแค้นส่วนตัว จะได้สืบสวนติดตามหาญาติผู้ตายมาสอบปากคำเป็นแนวทางคดีคลายคดีต่อไป

ทักษิณรับปากรับซื้อหนี้เฝ้าจากชาวบ้าน

ผู้สื่อข่าวรายงานว่า ภายหลังการประชุม ครม. เมื่อวันที่ 1 พ.ย. พ.ต.ท.ทักษิณได้เดินทางไปพบกับกลุ่มผู้ชุมนุม เครือข่ายหนี้สินชานาแห่งประเทศไทยภาคเหนือ จำนวนประมาณ 3,000 คน ที่มาชุมนุมเรียกร้องให้นายกฯเร่งแก้ไขปัญหาหนี้สินเกษตรกร ที่บริเวณหน้าวัดคีรีวงศ์ โดย พ.ต.ท.ทักษิณได้กล่าวกับกลุ่มผู้ชุมนุมว่า สิ่งทีพูดไว้ไม่เคยลืม วันที่ 2 พ.ย.นี้ จะเป็นประธานประชุมคณะกรรมการกองทุนฟื้นฟูและพัฒนาเกษตรกรด้วยตัวเอง และจะมีมติดีๆออกมา โดยเฉพาะผู้ที่มีปัญหาถูกฟ้องร้อง ถูกยึดทรัพย์ รัฐบาลจะถอนออกมาหมด และเราจะซื้อหนี้ออกมาให้หมด ได้สั่งการใน ครม.ให้ไปจัดการเรื่องหนี้ในระบบของคนจนทั้งหมด

“ไม่ต้องห่วง ตราบดีที่นั่นเป็นนายกฯอยู่ตรงนี้มีอะไรบอกให้รู้ถ้าอะไรช่วยได้ช่วยทันที อะไรทำให้คนไทยยืนบนลำแข้งตัวเองได้ ไม่ให้คนจนถูกข่มเหงรังแก วันนี้ได้บอกใน ครม.ว่า ไอ้พวกที่ให้กู้เงิน 2 หมื่น ทำสัญญา 2 แสน แล้วไปยึดที่นาชาวบ้าน ให้ไปช่วยทำให้มันจนลงหน่อยได้ไหม ประเภทนี้ผมไม่ชอบ ประเภทที่โกงชาวบ้านตาฉ่ำๆ ที่เที่ยวไปยึดไร่นาเขา ขอให้พวกเรากลับไปทำมาหากิน รัฐบาลจะดูแลให้” นายกฯกล่าว

นางแบบผู้ดี 'ลูกตาล' โดนตุ๋นเข้าคอนโดฯ สูญเกือบแสน

เมื่อเวลา 13.00 น. วันที่ 8 พ.ย. น.ส.ทิพย์วรรณ หรือลูกตาล-ชโลมจิต จันทร์เกตุ อายุ 33 ปี นักแสดงและนางแบบ แนวปลุกใจเสือป่า เดินทางเข้าพบ ร.ต.อ.สวึก นุชเจริญผล ร้อยเวรกลุ่มงานสอบสวน บก.ป. เพื่อแจ้งความดำเนินคดี กับนายพีระศิลป์ วาณิชกุล อายุ 31 ปี ในข้อหาฉ้อโกง หลังจากได้ทำ

สัญญาเช่าห้องพักเลขที่ 5/28 ชั้น 18 ปิยะสาทรคอนโดมิเนียม ซอยสวนพลู แขวงและเขตสาทร กทม. โดยจ่ายเงินไปจำนวน 75,000 บาท แต่ไม่สามารถเข้าไปพักอาศัยได้โดย น.ส.ชโลมจิตเผยถึง ต้นเหตุของการถูกโกงในครั้งนี้ว่า ได้อ่านพบประกาศให้เช่าห้องพักปิยะสาทร คอนโดมิเนียม ใน นิตยสารอาคารและที่ดินฉบับหนึ่ง แล้วมีความสนใจที่จะเช่าพัก เนื่องจากบ้านที่ซื้อไว้ยังตกแต่งไม่เสร็จ ตั้งใจว่าจะพักอาศัยประมาณ 6 เดือน จึงโทรศัพท์ ติดต่อกับนายพีระศิลป์ ซึ่งเป็นนายหน้าให้ เช่าคอนโดมิเนียมดังกล่าว และนัดไปดูห้องพักร่วมตกลงทำสัญญาเช่า เมื่อวันที่ 7 ต.ค.ที่ผ่านมา ในสัญญาระบุค่าเช่าเดือนละ 25,000 บาท จ่ายค่ามัดจำล่วงหน้า 3 เดือน เป็นเงิน 75,000 บาท เมื่อทำ สัญญาแล้วจึงได้รับกุญแจห้องและคีย์การ์ดสำหรับเข้าคอนโดมิเนียม นางแบบสาวกล่าวว่า ต่อมา เมื่อวันที่ 13 ต.ค.ที่ผ่านมา ได้ขนข้าวของเครื่องใช้ไปที่ห้องดังกล่าว แต่ ปรากฏว่า มีนายโสภณ โล ผู้จัดการโครงการและสถาปนิกอาวุโส บริษัทบิลเดอร์คอนซัลแตนต์ จำกัด และบริษัท บิลเดอร์เรีย ดิไซน์ จำกัด เข้ามาภายในห้องแจ้งว่าเขาเป็นเจ้าของห้อง โดยนำหลักฐานต่างๆมายืนยัน ตนจึงเล่า เรื่องที่นายพีระศิลป์ลงโฆษณาให้เช่าห้อง และได้จ่ายเงินค่าเช่าเกือบแสนบาทให้ฟัง แต่นายโสภณ ไม่สนใจ อ้างว่าไม่มีส่วนทราบเรื่องที่เกิดขึ้นแต่อย่างใด จึงจำใจต้องเก็บข้าวของกลับออกมา และ พยายามติดต่อนายพีระศิลป์ แต่ไม่สามารถติดต่อได้จึงเชื่อว่าถูกหลอก น.ส.ชโลมจิตเผยอีกว่า ได้ ตรวจสอบกับห้องเช่าที่อยู่ติดกันทราบว่า มีชาวต่างชาติอีกรายที่หลงเชื่อจ่ายเงินให้นายพีระศิลป์ไป แล้วเช่นเดียวกัน พนักงานในคอนโดมิเนียมหลายคนบอกว่า เคยเห็นนายพีระศิลป์ เข้ามาที่ คอนโดมิเนียมบ่อยครั้ง แต่ไม่ทราบว่า เป็นเจ้าของห้องพักหรือไม่ เชื่อว่าน่าจะมีส่วนรู้เห็นกับนายพี ระศิลป์ด้วย เนื่องจากกุญแจห้องพักและคีย์การ์ดเพื่อเข้าคอนโดมิเนียมที่ได้มาไม่น่าจะทำปลอม ขึ้นมาได้ จึงไม่อยากให้มีใครหลงเชื่ออีก เพราะยังมีการลงโฆษณาให้เช่าห้องพักในนิตยสารอยู่ เช่นเดิม ด้าน ร.ต.อ.สวิกกกล่าวว่า เบื้องต้นตำรวจได้รับเรื่องไว้แล้ว ส่วนจะมีการพิจารณาส่งให้ สน. พุ่งมหาเมฆ เจ้าของท้องที่เกิดเหตุรับไปดำเนินการหรือไม่ นั้น จะรายงานให้ผู้บังคับบัญชาทราบ และพิจารณาสั่งการตามลำดับขั้นต่อไป

"น้องเบนซ์ จูเนียร์" แท็กชีเมา ชนสนั่นทางด่วน

ลูกทุ่งเด็กชื่อดัง "น้องเบนซ์ จูเนียร์" เจ้าของเสียงเพลง "สู้ไว้" ประสบอุบัติเหตุรถตู้ชนกับแท็กชี บนทางด่วน โชคดีไม่เป็นอะไรมาก แต่ต้องเข้าโรงพยาบาลรอให้หมอตรวจเช็กอาการทางสมองอีก ที่ เคยก่อนเกิดเหตุนอนหลับมาในรถตู้ มีพ่อเป็นคนขับ กำลังจะไปเข้าห้องอัดเสียงทำอัลบั้มเพลงชุด ใหม่ เพราะห้ร้ายเจอแท็กชีเมาขับรถมาชน ตำรวจพาตัวโซเฟอร์ไปสน.ทางด่วน 2 แจ้งข้อหาขับรถ ประมาท กับเมาแล้วขับ แม่นักร้องเด็กยืนยันให้ดำเนินคดีถึงที่สุดเพื่อไม่ให้เป็นเยี่ยงอย่าง ด้าน

“น้องมายด์”เด็กคอยใจดีรู้ข่าวรีบเดินทางมาเยี่ยมให้กำลังใจถึงร.พ.

เหตุการณ์ลูกทุ่งเด็ก “น้องเบนซ์ จูเนียร์” ประสบอุบัติเหตุรถตู้เกี่ยวชนกับแท็กซี่ครั้งนี้ เมื่อเวลา 02.30 น. วันที่ 20 พ.ย. พ.ต.ท.ธัญญารินทร์ บริรักษ์ สารวัตรเวร สน.ทางด่วน 2 รับแจ้งเหตุรถเกี่ยวชนกันที่บริเวณทางด่วน 2 ขาเข้ากรุงเทพฯ ช่วงบริเวณสถานีขนส่งสายเหนือ (หมอชิตใหม่) แขวงจอมพล เขตจตุจักร รุดไปตรวจสอบพบรถตู้โตโยต้า ไฮเอช สีบรอนซ์ฟ้า ทะเบียน สก-8203 กรุงเทพมหานคร ในสภาพถูกชนด้านขวาบูบเป็นรอยยาว จอดอยู่ชิดขอบของทางด่วน ใกล้กับพบรถแท็กซี่น้ำเงิน-แดง ยี่ห้อโตโยต้า ลิโม ทะเบียน ทพ-8144 กรุงเทพมหานคร จอดชิดเลนขวา กันชนหน้ามีร่องรอยการชน ด้านซ้ายบูบทั้งแถบ ล้อด้านขวาแยกแตก ล้อแม็กช็อค

สอบสวนทราบว่ารถตู้โตโยต้าคันดังกล่าวมีนายสหวัด ฤกษ์ณ โสภา อายุ 50 ปี เป็นคนขับ มีนางพรทิพย์ ฤกษ์ณ โสภา อายุ 44 ปี ภรรยาและค.ญ.สุภจิรา ฤกษ์ณ โสภา หรือน้องเบนซ์ จูเนียร์ บุตรสาว อายุ 8 ขวบ นักร้องลูกทุ่งเด็กชื่อดัง เจ้าของอัลบั้มเพลงดัง"สู้ไว้"นั่งมาในรถด้วย ส่วนรถแท็กซี่มีนายอภิรักษ์ บุญปก อายุ 31 ปี อยู่บ้านเลขที่ 141/1 ม.1 ต.สามขา อ.โพธาราย จ.ร้อยเอ็ด เป็นคนขับ นายสหวัดกล่าวว่า ก่อนเกิดเหตุตนขับรถพาน้องเบนซ์ไปเรียนเทควันโดที่เมืองทองธานี จากนั้นได้แวะทานอาหารเพื่อเดินทางไปห้องอัดเสียงของบริษัทที่อปโลไน้โดมอนด์ ย่านพระโขนง ระหว่างขับรถมาตามทางด่วนแจ้งวัฒนะเพื่อไปลงที่อโงจรงค์ เมื่อมาถึงที่เกิดเหตุเป็นบริเวณทางโค้งหน้าหมอชิตใหม่ ตนขับมาเลนกลาง แต่จู่ๆก็มีรถแท็กซี่ขับเข้ามาชนอย่างรุนแรงจนรถของตนเสียหลักโคลงไปโคลงมา น้องเบนซ์ซึ่งนอนหลับอยู่บนเตียงด้านหลังรถ หัวกระแทกกับพนักพิงอย่างแรง หลังเหตุการณ์ตนได้ลงไปตกลงกับคนขับรถแท็กซี่ ซึ่งอยู่ในอาการมีนเมาและพยายามบ่ายเบี่ยงไม่รับผิดชอบ ตนจึงแจ้งเจ้าหน้าที่ตำรวจให้มาดำเนินการ

ด้านนางพรทิพย์กล่าวว่า หลังจากที่น่าตัวคู่กรณีมาที่สน.ทางด่วน 2 ตนและพ่อของน้องเบนซ์พยายามที่จะตกลงกับคนขับรถแท็กซี่ แต่ไม่เป็นผล แถมยังแสดงท่าทางก้าวร้าว ไม่ยอมรับผิด และไม่ยอมให้ตำรวจตรวจวัดแอลกอฮอล์ ทั้งที่เป็นฝ่ายขับชนรถตนเอง ทำให้น้องเบนซ์ซึ่งนอนอยู่เตียงด้านหลังกลิ้งลงมาจนหัวกระแทกพนักพิง ซึ่งคงต้องนำไปเอกซเรย์ที่โรงพยาบาลเพื่อตรวจอาการว่ารุนแรงหรือไม่

“รถแท็กซี่ขับมาชนข้างที่น้องเบนซ์นอนอยู่พอดี ทำให้ลูกสาวหัวกระแทกพนักพิงจนสะดุ้งตื่น ลูกจึงขึ้นมาบอกว่าเจ็บที่หัวด้านซ้าย ปวดตาซ้ายด้วยและมีนหัว จากนั้นก็ร้องไห้ ทำให้รู้สึกสงสารลูกและคงต้องดำเนินคดีกับคู่กรณีให้ถึงที่สุดให้เป็นตัวอย่างสำหรับผู้ที่เมาแล้วขับ เพราะเมาแล้วขับเป็นอันตรายกับทุกคน ไม่ใช่เฉพาะน้องเบนซ์เท่านั้น” นางพรทิพย์กล่าว

ด้านน้องเบนซ์กล่าวด้วยน้ำเสียงอ่อนเพลียว่า น้องเบนซ์รู้สึกว่ามีเสียงดังมากระแทกกับรถจนตกจากเตียง คิดว่ารถต้องมาชนแน่ แต่ไม่คิดว่าจะชนด้านที่เบนซ์นอนอยู่ หลังจากที่ดินแล้วรู้สึกปวดและ

เจ็บที่หัวด้านซ้าย

ผู้สื่อข่าวรายงานว่า นายอภิรักษ์ โชเฟอร์แท็กซี่ซึ่งถูกคุมตัวมาที่สน.ทางด่วน 2 ได้พยายามดื่มน้ำและอมสออล์เพื่อกลบเกลื่อนกลิ่นสุรา และเมื่อเจ้าหน้าที่ตำรวจนำเครื่องตรวจวัดแอลกอฮอล์มาตรวจวัด นายอภิรักษ์ก็พยายามอ้างว่าป่วยเป็นหวัด เจ้าหน้าที่ตำรวจจึงระบุว่าหากไม่ป่วยจะทำการเจาะเลือด นายอภิรักษ์จึงยอมเป่า ผลออกมาพบว่าปริมาณแอลกอฮอล์ 95 มิลลิกรัมเปอร์เซ็นต์ ถือว่าเกินกว่าที่กฎหมายกำหนด เจ้าหน้าที่ตำรวจจึงแจ้งข้อหาขับรถโดยประมาทเป็นเหตุให้ผู้อื่นได้รับบาดเจ็บ และข้อหาขับรถในขณะเมาสุรา

พ.ต.ท.ธัญญารินทร์กล่าวว่า สอบสวนผู้ต้องหาทราบว่าขับรถมาจากประชาชนจะไปหาเพื่อน โดยจะลงทางด่วนที่พญาไท แต่เมื่อมาถึงที่เกิดเหตุปรากฏว่ารถเสียหลักไปชนกับขอบทางด่วนจนยางแตก และเสียหลักไปชนกับรถตู้ของน้องเบนซ์ ซึ่งเจ้าหน้าที่จะดำเนินคดีตามข้อกล่าวหาต่อไป ต่อมาเวลา 17.30 น. ด.ญ.บุญพัชรี แสงทองวีรกุล หรือน้องมายด์ วัย 10 ขวบ นักร้องเด็กชื่อดัง เจ้าของเพลง เด็กคอยใจดี พร้อมด้วยแม่และพ่อ เดินทางมาเยี่ยมน้องเบนซ์ จูเนียร์ ซึ่งเข้าพักรักษาตัว และรอดูอาการ ที่ห้อง 1022 โรงพยาบาลนนทเวช ย่านงามวงศ์วาน โดยน้องมายด์ได้ตรงเข้าไปกอดน้องเบนซ์ พร้อมสอบถามถึงอาการบาดเจ็บ ทั้งยังได้นำนมสดที่เตรียมมาให้ให้น้องเบนซ์ด้วย นางพรทิพย์ มารดาของน้องเบนซ์กล่าวว่า หลังจากเข้าพักรักษาตัวที่ร.พ.ตั้งแต่เมื่อคืน ตอนนี้หลังจากที่หมอเอกซเรย์แล้วผลปรากฏว่าไม่เป็นอะไรมาก แต่มีรอยช้ำเล็กน้อย พอหมอให้ยา ก็ปวดคุดได้ปกติ แต่พอหมดฤทธิ์ยา ก็จะมีอาการปวดหัว หมอจึงให้นอนที่โรงพยาบาลก่อนเพื่อตรวจเช็คอาการทางสมองเพราะกลัวกระทบกระเทือน ถ้าหากเข้าวันที่ 21 พ.ย. อาการปกติก็จะเอกซเรย์อีกครั้ง ถ้าไม่มีอะไรก็จะกลับบ้านได้

เมื่อถามถึงผลกระทบของงาน นางพรทิพย์กล่าวว่า ตอนนี้ต้องยกเลิกคิวอัดเสียงออกไปก่อน ส่วนงานโชว์ตัวก็ต้องยกเลิกไปบางส่วน คิดว่าถ้าวันที่ 21 พ.ย.นี้ไม่มีอะไรก็กลับบ้านและทำงานได้ปกติ ส่วนเรื่องคดีจะให้เจ้าหน้าที่ตำรวจดำเนินการไปตามกฎหมาย เราไม่ได้เรียกร้องค่าเสียหาย แต่อยากจะให้ดำเนินการเพราะถ้ารู้ว่าเขาก็ไม่ควรจะขับ แรกๆ เขาพยายามจ่ายค่าเสียหายให้ แต่เราก็ไม่รับ อยากให้เป็นข้อเตือนใจสำหรับการเมาแล้วขับ ไม่อยากจะให้คิดว่าจ่ายเงินแล้วเรื่องก็จบไป แล้วคนขับก็กลับไปทำอีก

สำหรับน้องเบนซ์ จูเนียร์ เป็นนักร้องสังกัดค่ายท็อปไลน์ ไดมอนด์ มีผลงานอัลบั้มชุดแรกชื่อ "สู้ไว้" มีเพลงโด่งดังหลายเพลง อาทิ คุณครูที่รัก ดึกตอก ขอเวลาลูกบ้าง ขณะนี้กำลังอยู่ในระหว่างกำลังอัดเทปชุดที่สองคือชุด "จักรยานโซ่หย่อน"

ภาคผนวก ข

ผลการตัดคำและการแทนที่คำหายขาดด้วยเครื่องหมายดอกจัน ของตัวอย่างข้อความจากข่าวหนังสือพิมพ์ที่ใช้ทดสอบ ประสิทธิภาพของโปรแกรมในงานวิจัย

ผลการตัดคำ ' / เข้าทุกข์ / ' / โผล่ / เพียบ / แท้ / ซี่ / หิ้น / ! / โคน / ซี่ตัว / - / หลาย / คดี /
ความ / คีบ / หน้า / คดี / แท้ / ซี่ / มหา / ภัย / กรณี / น.ส. / อารี / รัตน์ / พวง / ทอง / อายุ /
29 / ปี / สาว / พ / ริ / ตดี / บาร์ / ร้าน / ซพ / รา / โน / ซอย / สุขุม / วิ / ท / 26 / เรียก /
รถ / แท้ / ซี่ / เขียว / เหลือง / ยี่ / หื้อ / มิ / ต / ชู / บี / ซิ / ทะ / เบียน / มค / 4973 /
กรุงเทพมหานคร / จาก / ปาก / ซอย / ลาด /พร้าว / ว / 95 / ให้ / ไป / ส่ง / ศูนย์ / การ / ค้า / สยาม
/ ดิสคัฟเวอ / รี / ย่าน / ปทุม / วัน / แต่ / เมื่อ / รถ / วิ่ง / ถึง / สะพาน / หัว / ซ้าง / เขต / ราช /
เทวี / กทม . / . / โซเฟอร์ / แท้ / ซี่ / ออก / ลาย / โจร / จอด / รถ / ทำ / ที่ / ขอ / ลง / ไป / ปี / สสาวะ /
เสร็จ / แล้ว / ชัก / มีด / จี้ / จับ / มัด / มือ / โผล่ / หลัง / ก่อน / ปลด / ทรัพย์ / สิ้น / บัง / คับ / ให้ / นั่ง /
เบาะ / หน้า / รถ / แล้ว / ขับ / มุ่ง / หน้า / ไป / ตาม / ถนน / เพชร /บุรี / แต่ / เมื่อ / ถึง / แยก / อรุ /
พงษ์ / รถ / ติด / สัญญาณ / ไฟ / แดง / ผู้ / เสีย / หาย / ตัด / ลิ / น / ใจ / เปิด / ประตู / รถ / แล้ว / กลิ้ง / ตัว /
ลง / จาก / รถ / หนี / เอา / ตัว / รอด / มา / ได้ / หวาด / หวิด / ส่วน / คน / ร้าย / ขับ / รถ / หลบ / หนี
/ ไป / ต่อ / มา / เข้า / หน้า / ที่ / ตำ / รวจ / สน . / . / พญา / ไท / ติดตาม / จับ /กุม / โซเฟอร์ / โจร / ได้ /
รับ / สาร / ภาพ / ว่า / ก่อ / คดี / มา / แล้ว / โจร / โชน / โดย / เลือ / ก / เหยื่อ / ผู้ / โดย / สาร / ผู้ / หนี /
หน้า / คา / ดิ / ปลด / ทรัพย์ / แล้ว / พา / ไป / ข่ม / ขืน / มา / แล้ว / หลาย / ราย / นั้น / เมื่อ / เวลา /
11.30 / น. / วัน / ที่ / 31 / ต.ค. / . / พล / .ต.ต. / ปรา / โมช / ปทุม / วงศ์ / ผ / บก / .น.1 /
พ.ต.อ. / กา / รุ / ณ / กลิ่น / ประ / ยู / ร / ผก / ก. / สน . / . / พญา / ไท / พ.ต.ท. / สมิง / รอด
/ รัต / ษะ / สว / .สส. / . / ได้ / ควบคุม / ตัว / นาย / ณรงค์ / หรือ / บัง / ชู / ป / ปั้น / หยี / อายุ
/ 40 / ปี / โซเฟอร์ / แท้ / ซี่ / ภัย / สังค / ม / ผู้ / ต้อง / หา / ไป / ทำ / แผน / ประ / ทูษ / กรรม /
ประกอบ / คำ / รับ / สาร / ภาพ / บริ / เวณ / เจริ / สะพาน / หัว / ซ้าง / ถนน / พญา / ไท / แขวง /
ถนน / เพชร /บุรี / เขต / ราช / เทวี / กทม . / . / จุด / ที่ / นาย / ณรงค์ / จอด / รถ / ทำ / ที่ / ไป /
ปี / สสาวะ / เสร็จ / แล้ว / ชัก / มีด / จี้ / จับ / มัด / ผู้ / เสีย / หาย / พร้อม / กับ / ปลด / ทรัพย์ / สิ้น /

ส่วน/จุด/ที่/ 2/ บริเวณ/แยก/อุรุ/พงษ์/ จุด/ที่/ผู้เสียหาย/เปิด/ประตู/รถ/แล้ว/ทิ้ง
 /ตัว/กลิ้ง/ลง/จาก/รถ/แท็ก/ซี่/ ขณะ/จอด/ติด/สัญญาณ/ไฟ/แดง/ ใช้/เวลา/ทำ/
 แผน/ทั้ง/สอง/แห่ง/ ประมาณ/ครึ่ง/ชั่วโมง/จึง/เสร็จ/สิ้น/ ต่อ/มา/ได้/มี/ น.ส./
 ธิดา/ เส็ก/สงบ/ อายุ/ 27/ ปี/ พนักงาน/ชาย/จิ/เวล/รี/ย่าน/สี่/พระยา/ อยู่/
 บ้าน/เลข/ที่/ 52/203/ หมู่/ บ้าน/กฤษณา/ ถนน/กรุงเทพ/กรีฑา/ แขวง/และ/เขต
 /สะพาน/สูง/ กทม./ ผู้เสียหาย/เหยื่อ/แท็ก/ซี่/ มหา/ภัย/ เข้า/พบ/ พนักงาน
 สอบสวน/เจ้าของ/คดี/ โดย/นำ/บันทึก/ประจำ/วัน/ที่/แจ้ง/ความ/ไว้/ สน./ประเวศ
 / เหตุ/เกิด/เมื่อ/เวลา/ 08.00/ น./ วัน/ที่/ 10/ ต.ค./ ที่/ผ่าน/มา/ เพื่อ/ชี้ตัว
 /คนร้าย/เพิ่มเติม/ เหยื่อ/สาว/ผู้เสียหาย/ให้/ราย/ละเอียด/ว่า/ เรียก/รถ/แท็ก/ซี่/
 เขียว/เหลือง/จาก/หน้า/หมู่/บ้าน/กฤษณา/ให้/ไป/ส่ง/ย่าน/สี่/พระยา/ แต่/เมื่อ/รถ
 /ผ่าน/บริเวณ/ถนน/คู/ขนาน/มอเตอร์/เวย์/ ใกล้/สุเหร่า/หัว/หมาก/ใหม่/ แขวง/
 และ/เขต/สวน/หลวง/ กทม./ โชเฟอร์/จอด/ชั่ว/ได้/จอด/รถ/อ้าง/ว่า/ขอ/ไป
 ตรวจ/เช็ก/วาล์ว/แก๊ส/ด้าน/หลัง/รถ/ แต่/กลับ/เปิด/ประตู/ชัก/มิด/ปลาย/แหลม/
 จี/บังคับ/ปลด/ทรัพย์สิน/ได้/โทรศัพท์/มือ/ถือ/ยี่ห้อ/ฟิลิ/ป/ รูน/จี/ 790/ และ/
 เงินสด/ 1,500/ บาท/ พร้อม/กับ/ใช้/เชือก/มัด/มือ/ไพล่หลัง/ แต่/หลังจาก/ขับ/
 รถ/ไป/ได้/ไม่/ไกล/ ตน/แก้/มัด/มือ/จน/เป็น/อิสรภาพ/ เสร็จ/แล้ว/พยายาม/
 แหย่/มิด/จาก/คนร้าย/ แต่/โชเฟอร์/โจร/ จอด/รถ/แล้ว/แหย่/มิด/กลับคืน/ไป/ได้/
 สาว/เหยื่อ/แท็ก/ซี่/ โจร/กรุง/ให้/ราย/ละเอียด/ต่อว่า/ พยายาม/ทำ/ใจ/ดี/ผู้/เสื่อ/
 วิงวอน/ร้องขอ/ชีวิต/ ว่า/ได้/ทรัพย์สิน/ไป/แล้ว/อย่า/ทำ/อะไร/ ทำให้/โชเฟอร์/แท็ก
 /ซี่/ใจอ่อน/ขวาง/มิด/ทิ้ง/ แต่/ก็/ยัง/มัด/มือ/พา/ขึ้น/รถ/ขับ/ไป/ต่อ/ โดย/ขับ/
 อ้อม/เข้าไป/ใน/ซอย/คา/รา/บา/ว/ ถนน/กรุงเทพ/กรีฑา/ ซึ่ง/เป็น/เส้นทาง/
 เปลี่ยว/ เสร็จ/แล้ว/จอด/รถ/เพื่อ/จะ/ทำ/มิดิมีร้าย/ จึง/อาศัย/จังหวะ/ที่/คนร้าย/ไม่
 ท้น/ตั้งตัว/ รวบรวม/สติ/พละ/กำลัง/เปิด/ประตู/รถ/วิ่ง/หนี/อย่าง/ไม่/คิด/ชีวิต/
 โดย/ไม่/ทัน/ได้/จำ/หมายเลข/ทะเบียน/รถ/ ทราบ/แต่/ว่า/เป็น/แท็ก/ซี่/เขียว/เหลือง/
 กระทั่ง/เมื่อ/เห็น/หน้า/ของ/คนร้าย/ทาง/หนังสือพิมพ์/ จึง/จำ/ได้/ว่า/เป็น/คน/
 เดียว/กับ/ที่/ก่อ/เหตุ/ ต่อ/มา/เจ้าหน้าที่/ได้/นำ/ตัว/นาย/ณรงค์/ออก/มา/ให้/
 เหยื่อ/สาว/ราย/นี้/ชี้ตัว/ น.ส./ธิดา/ยืนยัน/ว่า/จำ/หน้า/ได้/อย่าง/แม่นยำ/เพราะ/ว่า/
 นั่ง/อยู่/ใน/รถ/ด้วย/กัน/นาน/กว่า/ชั่วโมง/เศษ/ เจ้าหน้าที่/จึง/อายัด/ตัว/และ/แจ้ง/
 ข้อหา/ชิงทรัพย์/ใน/เวลา/กลางคืน/โดย/มี/อาวุธ/ โดย/ใช้/ยาน/พาหนะ/และ/
 หน่วงหนี่ยว/ หรือ/กักขัง/ผู้/อื่น/ หรือ/กระทำ/ด้วย/ประการ/ใด/ให้/ผู้/อื่น/

ปราศจาก / เสรีภาพ / ใน / ร่างกาย / เพิ่ม / อีก / คดี / โดย / ตำ / สุด / ทาง / สก.อ. / เมือง / สมุทรปราการ / กำลัง / ขอ / อนุมัติ / หมายจับ / มา / อาศัย / ตัว / เพิ่ม / อีก / ราย / ด้าน / พล . ต.ต. / ปราโมช / ปทุม / วงศ์ / ผ / บก . น . / 1 / เปิด / เผย / ว่า / ใน / วัน / พุธ / นี้ / จะ / นำ / ตัว / ผู้ / ต้อง / หา / ไป / ผาก / ชัง / และ / คิด / ว่า / คง / ต้อง / นำ / ตัว / กลับ / มา / ควบคุม / ต่อ / เพื่อ / ขยาย / ผล / และ / ให้ / ผู้ / เสีย / หาย / ชี้ / ตัว / ยืนยัน / เพราะ / น่า / จะ / มี / ผู้ / ที่ / ตก / เป็น /เหยื่อ / อีก / หลาย / ราย / สำหรับ / รถ / แท็กซี่ / คัน / ที่ / ใช้ / ก่อ / เหตุ / ผู้ / ต้อง / หา / เข้า / มา / จาก / อยู่ / เจ้า / แดง / ใน / ซอย / พัฒนาการ / 64 / โดย / เข้า / เหม / 24 / ชั่วโมง / เป็น / รถ / รุ่น / เก่า / เจ้าของ / อยู่ / ตก / ลง / ให้ / ไป / ชำ / ระ / ค่า / เข้า / อา / ทิ / ตย์ / ละ / ครั้ง / โดย / คิด / วัน / ละ / 800 / บาท / วัน / อา / ทิ / ตย์ / ขับ / ฟรี / 1 / วัน / วัน / เดียว / กัน / พ.ต.ท.กิต / พัทธ์ / เทียง / น้อย / สว. / สส. / สน. / ลาด /พร้าว / ว / ได้ / ติด / ต่อ / น.ส. / นิ / นา / (/ นาม / สม / มติ) / อายุ / 20 / ปี / นัก / ศึกษา / ชั้น / ปี / 1 / มหาวิทยาลัย / อัส / สัม / ชัญ / หรือ / เอ / แบ / ค / ลูก / สาว / ของ / สมาชิก / สภา / จังหวัด / สุ / รา / ช / ฎ / ร์ / ธานี / และ / เป็น / หลาน / สาว / นาย / สุ / เทพ / เทือก / สุ / บรรณ / เลข / ชาติ / การ / พร / รค / ประชา / ชี / ป / ตย์ / ผู้ / เสีย / หาย / อีก / ราย / ให้ / ยืนยัน / ตัว / โฆ / เฟอร์ / แท็กซี่ / มหา / ภัย / ราย / นี้ / หลังจาก / เมื่อ / คืน / วัน / ที่ / 12 / ต.ค. / ที่ / ผ่าน / มา / เวลา / ประมาณ / ติ / สอง / น.ส. / นิ / นา / ตก / เป็น / เหยื่อ / โดย / เรียก / รถ / แท็กซี่ / จาก / เอ / แบ / ค / คอน / โด / หลัง / มหาวิทยาลัย / ราม / คำ / แห / ง / ให้ / ไป / ส่ง / ใน / ซอย / ลาด /พร้าว / ว / 35 / เพื่อ / หา / เพื่อน / แต่ / ปรากฏ / ว่า / เมื่อ / ผ่าน / หน้า / ร้าน / อาหาร / แม่ / ลา / ปลา / เผา / กลาง / ซอย / ลาด /พร้าว / ว / 130 / แขวง / คลอง / จั่น / เขต / บาง / กะ / ปี / คน / ขับ / รถ / แท็กซี่ / ได้ / จอด / รถ / อ้าง / ว่า / ขอ / ลง / ไป / ทำ / ธุ / ระ / จาก / นั้น / เดิน / อ้อม / หลัง / รถ / ไป / เปิด / ประตู / หลัง / ใช้ / มือ / จิก / หัว / แล้ว / กระ / ซาก / โทรศัพท์ / มือ / ถี / ้อ / กับ / กระ / เป้า / สด / าย / ค์ / มี / เงิน / สด / 3,300 / บาท / ก่อน / ที่ / เหยื่อ / สาว / จะ / สะ / บัด / ตัว / เปิด / ประตู / รถ / อีก / ด้าน / วิ่ง / หลบ / หนี / ได้ / หวาด / หวิด / โดย / ผู้ / เสีย / หาย / บอก / ว่า / เห็น / รูป / ของ / คน / ร้าย / แล้ว / ยืนยัน / ว่า / เป็น / คน / เดียว / กัน / แต่ / ตอน / ก่อ / เหตุ / ผม / ยาว / กว่า / ใน / ขณะ / นี้ / ทาง / ตำ / รวจ / จึง / ประ / สาน / ให้ / เข้า / พบ / พนักงาน / สอบ / สวน / สน. / ลาด /พร้าว / ว / เพื่อ / สอบ / ปาก / คำ / เพิ่ม / เดิม / ก่อน / ขอ / อนุมัติ / หมาย / จับ / และ / อา / ยัด / ตัว / ผู้ / ต้อง / หา / ราย / นี้ / ต่อ / ไป / ส่วน / คดี / แท็กซี่ / ี่ / ราว / น้ำมัน / แล้ว / จุด / ไฟ / เผา / สาว / ไม่ / ทราบ / ชื่อ / อย่าง / โหด / เขี้ย / ม / ก่อน / ไป / สิ้น / ใจ / ที่ / โรง / พยา / บาล / เหตุ / เกิด / ใน / ซอย / พระ / ราม / สอง / 38 / แยก / 3 / ถนน / พระ / ราม / สอง / แขวง / และ / เขต / จอม / ทอง / ผัง / ธน / ๗ / โดย / ตำ / รวจ / ยัง / สืบ / สน / ชื่อ / “ / จำ / ปี / มี / ชัย / ” / ที่ / ผู้ / ตาย / บอก / ก่อน / หม / ด / ลม / หาย / ใจ / เนื่องจาก / มี / อยู่ /

จำนวน / มาก / และ / ยัง / มี / ชีวิต / อยู่ / นั้น / ความ / คืบ / หน้า / เมื่อ / เวลา / 11.00 / น. / วัน / ที่ / 31 / ต.ค. / . / ที่ / กก. / สส. / บก. / น.8 / พ.ต.อ. / จี / ร / ศักดิ์ / ขำ / คง / ผก / ก.สส.น.8 / ได้ / ประชุม / ชุด / สืบสวน / เพื่อ / คลี่คลาย / คดี / จาก / นั้น / เปิดเผย / ว่า / ได้ / เรียก / ฝ่าย / สืบสวน / ที่ / ออก / ไป / หา / ข่าว / กลับ / มา / หรือ / เกี่ยว / กับ / ประเด็น / ข้อมูล / ที่ / ได้ / มา / ใน /เบื้องต้น / โดย / เฉพาะ / ชื่อ / “ / จำปี / มี / ชัย / ” / อายุ / 35 / ปี / ชาว / จ. / ร้อยเอ็ด / ทาง / เจ้าหน้าที่ / ตรวจ / สอบ / ราย / ชื่อ / และ / นามสกุล / อย่าง / ละเอียดย / รวมทั้ง / คำ / ที่ / โกส / เสียง / แล้ว / ปรากฏ / ว่า / มี / อยู่ / เป็น / ร้อย / ชื่อ / ซึ่ง / จาก / การ / สุ่มตัวอย่าง / ทุก / คน / ยัง / มี / ชีวิต / อยู่ / ทั้งสิ้น / ทำให้ / การ / ตาม / หา / ว่า / ผู้ / ตาย / เป็น / ใคร / เริ่ม / ปรึกษา / ส่วน / ที่ / พยาน / เห็น / ว่า / สาว / ผู้ / ตาย / ลง / มา / จาก / รถ / สี / ส้ม / เล็ก / หรือ / รถ / กระ / ป้อ / โดย / ไล่ / เสื้อ / สี / ฉุกเฉิน / ยี่ห้อ / บอ / ดี / โกล / ฟ / มี / รูป / มือ / ที่ / หน้าอก / ก่อน / เรียก / รถ / แท็กซี่ / ซี่ / หาย / ไป / ใน / ซอย / ดัง / กล่าว / เจ้าหน้าที่ / ฝ่าย / สืบสวน / ตรวจ / สอบ / แล้ว / พบ / ว่า / ผู้หญิง / คน / นั้น / ยัง / มี / ชีวิต / อยู่ / โดย / พัก / อยู่ / การ / เคหะ / ชุมชน / พระ / ราม / 2 / ทำให้ / ต้อง / เริ่ม / ต้น / ค้น / หา / กัน / ใหม่ / ผก / ก.สส.น.8 / กล่าว / ด้วยว่า / จาก / การ / วิเคราะห์ / สภาพ / ศพ / ใน /เบื้องต้น / เขี่ย / สาว / สูง / ประมาณ / 150 / ซม. / . / เศษ / หน้าตา / ดู / แล้ว / ไม่ / น่า / จะ / ถึง / 35 / ปี / ส่วน / ชื่อ / ที่ / ผู้ / ตาย / บอ / ก่อน / ตาย / ว่า / “ / จำปี / มี / ชัย / ” / อายุ / 35 / ปี / ชาว / จ. / ร้อยเอ็ด / หลังจาก / เจ้าหน้าที่ / ตรวจ / สอบ / ชื่อ / ของ / ผู้หญิง / ไป / แล้ว / ไม่ / น่า / จะ / ไข / ได้ / หัน / มา / ตรวจ / สอบ / ราย / ชื่อ / ของ / ผู้ชาย / บ้าง / เพราะ / อาจ / จะ / เป็น / ชื่อ / ของ / โฆเฟอร์ / แท็กซี่ / ที่ / ก่อ / เหตุ / ก็ / เป็น / ได้ / นอกจาก / นี้ / ยัง / ให้ / ชุด / สืบสวน / ตรวจ / สอบ / หญิง / ชาย / บริการ / ทาง / เพศ / ใน / ละแวก / เกิด / เหตุ / ย่าน / ถนน / พระ / ราม / 2 / วงเวียน / ใหญ่ / พระ / ประแดง / ว่า / มี / สาว / ชาย / บริการ / คน / ไค / หายตัว / ไป / บ้าง / สำหรับ / ข้อมูล / ที่ / พยาน / ระบุ / ว่า / เห็น / รถ / แท็กซี่ / เขี่ย / เหลือง / ออก / มา / จาก / ซอย / เกิด / เหตุ / ตรวจ / สอบ / แล้ว / พบ / ว่า / มี / อยู่ / 2 / ยี่ห้อ / คือ / โต / โย / ต้า / และ / นิ / ส / สัน / แต่ / ชุด / สืบสวน / ประมวล / จาก / คำ / บอ / เล่า / ของ / พยาน / แล้ว / เชื่อ / ว่า / น่า / จะ / เป็น / รถ / นิ / ส / สัน / ซา / ลูน / ที่ / ม่าน / ผ้า / สี / ฟ้ำ / ติด / ด้าน / หลัง / มาก / กว่า / ขณะ / นี้ / อยู่ / ระหว่าง / ตรวจ / สอบ / ตาม / อยู่ / รถ / แท็กซี่ / ต่าง ๆ / ย่าน / บาง / ปะ / กอก / บาง / ปะ / แก้ว / และ / ราษฎร์ / บุรณะ / ให้ / ชัดเจน / นอกจาก / นี้ / กำลัง / รอ / ผล / การ / ตรวจ / พิสูจน์ / ของ / แพทย์ / รพ. / ศิ / ริ / ราช / โดย / เจ้าหน้าที่ / ฝ่าย / สืบสวน / ได้ / ทำ / หนังสือ / ขอ / ให้ / ตรวจ / ศพ / อย่าง / ละเอียดย / โดย / เฉพาะ / ใน / กระเพาะ / อาหาร / และ / เลือด / ให้ / มาก / ที่สุด / เท่า / ที่ / จะ / ทำ / ได้ /

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน 'เจ้าทุกข์' โผล่เพียบ แท้ก็ซี่*! โคนซี่ตัว-หลายคดี ความตีหน้าคดีแท้กซี่มหาภัย กรณี น.ส.อารีรัตน์ พวงทอง อายุ 29 ปี สาวพ*ตดีบาร์ ร้านชพราโน ซอย สุขุมวิท 26 เรียกกรดแท้กซี่เขียวเหลือง ยี่ห้อมิตชูบิชิ ทะเบียน มค 4973 กรุงเทพมหานคร จากปากซอยลาดพร้าว 95 ให้ไปส่งศูนย์การค้า สยามดิศคัฟเวอร์ ย่านปทุมวัน แต่เมื่อรถวิ่งถึงสะพานหัวช้าง เขตราชเทวี กทม. โชเฟอร์แท้กซี่ออกฉายโจร จอดรถทำทีขอลงไปปีสสาวะ เสร็จแล้วชักมีดจี้จับมัดมือไพล่หลัง ก่อนปลดทรัพย์สิน บังคับให้นั่งเบาะหน้ารถ แล้วขับมุ่งหน้าไปตามถนนเพชรบุรี แต่เมื่อถึงแยกอรุณพงษ์รถติดสัญญาณไฟแดง ผู้เสียหายตัดสินใจเปิดประตูรถแล้วถลันตัวลงจากรถ หนีเอาตัวรอดมาได้หวุดหวิด ส่วนคนร้ายขับรถหลบหนีไป ต่อมาเจ้าหน้าที่ตำรวจ สน.พญาไท ติดตามจับกุมโชเฟอร์โจรได้ รับสารภาพว่าก่อคดีมาแล้วโชกโชน โดยเลือกเหยื่อผู้โดยสาร ผู้หญิงหน้าตาดี ปลดทรัพย์แล้วพาไปข่มขืนมาแล้วหลายรายนั้น เมื่อเวลา 11.30 น. วันที่ 31 ต.ค. พล.ต.ต.ปราโมช ปทุมวงศ์ ผบก.น.1 พ.ต.อ.การุณ กลิ่นประยูร ผกก.สน. พญาไท พ.ต.ท.สมิง รอดรัตนะ สว.สส. ได้ควบคุมตัวนายณรงค์ หรือบังซูป ปันหยี อายุ 40 ปี โชเฟอร์แท้กซี่ภัยสังคมผู้ต้องหา ไปทำแผน*กรรมประกอบคำรับสารภาพบริเวณเชิงสะพานหัวช้าง ถนนพญาไท แขวงถนนเพชรบุรี เขตราชเทวี กทม. จุดที่นายณรงค์จอดรถทำทีไปปีสสาวะ เสร็จแล้วชักมีดจี้จับมัดผู้เสียหาย พร้อมกับปลดทรัพย์สิน ส่วนจุดที่ 2 บริเวณแยกอรุณพงษ์ จุดที่ผู้เสียหายเปิดประตูรถแล้วทิ้งตัวถลันลงจากรถแท้กซี่ ขณะจอดติดสัญญาณไฟแดง ใช้เวลาทำแผนทั้งสองแห่ง ประมาณครึ่งชั่วโมงจึงเสร็จสิ้น ต่อมาได้มี น.ส.ธิดา เคิกสงบ อายุ 27 ปี พนักงานขายจิวเวลรี่ย่านสี่พระยา อยู่บ้านเลขที่ 52203 หมู่ บ้านกฤษณา ถนนกรุงเทพกรีฑา แขวงและเขตสะพานสูง กทม. ผู้เสียหายเหยื่อแท้กซี่มหาภัย เข้าพบ พนักงานสอบสวนเจ้าของคดี โดยนำบันทึกประจำวันที่แจ้งความไว้ สน. ประเวศ เหตุเกิดเมื่อเวลา 08.00 น. วันที่ 10 ต.ค. ที่ผ่านมา เพื่อชี้ตัวคนร้ายเพิ่มเติม เหยื่อสาวผู้เสียหายให้รายละเอียดว่า เรียกกรดแท้กซี่เขียวเหลืองจากหน้าหมู่บ้านกฤษณาให้ไปส่งย่านสี่พระยา แต่เมื่อรถผ่านบริเวณถนนคูขนนมอเดอร์เวย์ ใกล้สุเหร่าหัวหมากใหม่ แขวงและเขตสวนหลวง กทม. โชเฟอร์ยอด*ได้จอดรถอ้างว่าขอไปตรวจเช็ควาล์วแก๊สด้านหลังรถ แต่กลับเปิดประตูชักมีดปลายแหลมจี้บังคับปลดทรัพย์สินได้โทรศัพท์มือถือยี่ห้อฟิลิป รุ่นจี 790 และเงินสด 1,500 บาท พร้อมกับใช้เชือกมัดมือไพล่หลัง แต่หลังจากขับรถไปได้ไม่กี่ไกล ตนแก้มัดมือจนเป็นอิสระภาพ เสร็จแล้วพยายามแย่งมีดจากคนร้าย แต่โชเฟอร์โจร จอดรถแล้วแย่งมีดกลับคืนไปได้ สาวเหยื่อแท้กซี่โจรกรุงให้รายละเอียดต่อว่า พยายามทำใจดีสู้เสือ วิงวอนร้องขอชีวิต ว่าได้ทรัพย์สินไปแล้วอย่าทำอะไร ทำให้โชเฟอร์แท้กซี่ใจอ่อนขวางมีดทิ้ง แต่ก็ยังมีมือพาวขึ้นรถขับไปต่อ โดยขับอ้อมเข้าไปในซอยการาบาว ถนนกรุงเทพกรีฑา ซึ่งเป็นเส้นทางเปลี่ยว เสร็จแล้วจอดรถเพื่อจะทำมีดมีร้าย จึงอาศัยจังหวะที่คนร้ายไม่ทันตั้งตัว รวบรวมสติพละกำลังเปิดประตูรถวิ่งหนีอย่างไม่คิดชีวิต โดยไม่

ทันได้จำหน่ายเลขทะเบียนรถ ทราบแต่ว่าเป็นแท็กซี่เขียวเหลือง กระทบเมื่อเห็นหน้าของคนร้ายทางหนังสือพิมพ์ จึงจำได้ว่าเป็นคนเดียวกับที่ก่อเหตุ ต่อมาเจ้าหน้าที่ได้นำตัวนายณรงค์ออกมาให้เหยื่อสาวรายนี้ชี้ตัว น.ส.ธิดายืนยันว่าจำหน้าได้อย่างแม่นยำเพราะว่านั่งอยู่ในรถด้วยกันนานกว่าชั่วโมงเศษ เจ้าหน้าที่จึงอายัดตัวและแจ้งข้อหาชิงทรัพย์ในเวลากลางคืนโดยมีอาวุธ โดยใช้ยานพาหนะและหน่วงเหนี่ยว หรือกักขังผู้อื่น หรือกระทำความผิดประการใดให้ผู้อื่นปราศจากเสรีภาพในร่างกาย เพิ่มอีกคดี โดยล่าสุดทาง สภ.อ.เมืองสมุทรปราการ กำลังขออนุมัติหมายจับมาอายัดตัวเพิ่มอีกราย ด้าน พล.ต.ต.ปราโมช ปทุมวงศ์ ผบก.น. 1 เปิดเผยว่า ในวันพรุ่งนี้จะนำตัวผู้ต้องหาไปฝากขังและคิดว่าคงต้องนำตัวกลับมาควบคุม ต่อเพื่อขยายผลและให้ผู้เสียหายชี้ตัวยืนยัน เพราะน่าจะมีผู้ที่ตกเป็นเหยื่ออีกหลายราย สำหรับรถแท็กซี่คันที่ใช้ก่อเหตุ ผู้ต้องหาเข้ามาจากอุ้งเงางง ในซอยพัฒนาการ 64 โดยเช่าเหมา 24 ชั่วโมง เป็นรถรุ่นเก่า เจ้าของอยู่ตลิ่งใต้ไปชำระค่าเช่าอาทิตย์ละครั้ง โดยคิดวันละ 800 บาท วันอาทิตย์ขับฟรี 1 วัน วันเดียวกัน พ.ต.ท.กิตติพัทธ์ เทียงน้อย สว.สส.สน. ลาดพร้าว ได้ติดต่อ น.ส.นينا (นามสมมติ) อายุ 20 ปี นักศึกษาชั้นปี 1 มหาวิทยาลัยอัสสัมชัญ หรือเอแบค ลูกสาวของสมาชิกสภาจังหวัดสุราษฎร์ธานี และเป็นหลานสาวนายสุเทพ เทือกสุบรรณ เลขาธิการพรรคประชาธิปัตย์ ผู้เสียหายอีกราย ให้ยืนยันตัวไซเฟอร์แท็กซี่หมายเลขรายนี้ หลังจากเมื่อคืนวันที่ 12 ต.ค. ที่ผ่านมา เวลาประมาณตีสอง น.ส.นินาตกเป็นเหยื่อ โดยเรียกรถแท็กซี่จากเอแบคคอนโด หลังมหาวิทยาลัยรามคำแหง ให้ไปส่งในซอยลาดพร้าว 35 เพื่อหาเพื่อน แต่ปรากฏว่าเมื่อผ่านหน้าร้านอาหารแม่ลา ปลาเผา กลางซอยลาดพร้าว 130 แขวงคลองจั่น เขตบางกะปิ คนขับรถแท็กซี่ได้จอดรถ อ้างว่าขอลงไปทำธุระ จากนั้นเดินอ้อมหลังรถไปเปิดประตูหลัง ใช้มือจิกหัวแล้วกระชากโทรศัพท์มือถือถือกับกระเป๋าสตางค์มีเงินสด 3,300 บาท ก่อนที่เหยื่อสาวจะสะบัดตัว เปิดประตูรถอีกด้านวิ่งหลบหนีได้หวุดหวิด โดยผู้เสียหายบอกว่า เห็นรูปของคนร้ายแล้ว ยืนยันว่าเป็นคนเดียวกัน แต่ตอนก่อเหตุผมยาวกว่าในขณะนี้ ทางตำรวจจึงประสานให้เข้าพบพนักงานสอบสวน สน.ลาดพร้าว เพื่อสอบปากคำเพิ่มเติม ก่อนขออนุมัติหมายจับและอายัดตัวผู้ต้องหารายนี้ต่อไป ส่วนคดีแท็กซี่รูดน้ำมันแล้วจุดไฟเผาสาวไม่ทราบชื่ออย่างไหนก่อนไปสิ้นใจที่โรงพยาบาล เหตุเกิดในซอยพระรามสอง 38 แยก 3 ถนนพระรามสอง แขวงและเขตจอมทอง ฝั่งธนฯ โดยตำรวจยังสืบสนชื่อ “จำปี มีชัย” ที่ผู้ตายบอกก่อนหมดลมหายใจ เนื่องจากมีอยู่จำนวนมากและยังมีชีวิตอยู่นั้น ความคืบหน้าเมื่อเวลา 11.00 น. วันที่ 31 ต.ค. ที่ กก.สส.บก.น.8 พ.ต.อ.จิรศักดิ์ จำดง ผกก.สส.น.8 ได้ ประชุมชุดสืบสวนเพื่อคลี่คลายคดี จากนั้นเปิดเผยว่า ได้ เรียกฝ่ายสืบสวนที่ออกไปหาข่าวกลับมาหารือเกี่ยวกับประเด็นข้อมูลที่ได้มาในเบื้องต้น โดยเฉพาะชื่อ “จำปี มีชัย” อายุ 35 ปี ชาว จ.ร้อยเอ็ด ทางเจ้าหน้าที่ตรวจสอบรายชื่อและนามสกุลอย่างละเอียด รวมทั้งคำที่ใกล้เคียงแล้วปรากฏว่ามีอยู่เป็นร้อยชื่อ ซึ่งจากการสุ่มตัวอย่างทุกคนยังมีชีวิตอยู่ทั้งสิ้น ทำให้การตามหาว่าผู้ตาย

เป็นใครเริ่มรีบหรี ส่วนที่พยานเห็นว่าสาวผู้ตายลงมาจากรถ ลี้อเล็ก หรือรถกะป้อ โดยใส่เสื้อสี
 ฉุกเฉิน ยี่ห้อบอดี้โกลฟ มีรูปมือที่หน้าอก ก่อนเรียกรถแท็กซี่หายไปนชอยดังกล่าว เจ้าหน้าที่ฝ่าย
 สืบสวนตรวจสอบแล้วพบว่าผู้หญิงคนนั้นยังมีชีวิตอยู่ โดยพักอยู่การเคหะชุมชนพระราม 2 ทำให้
 ต้องเริ่มต้นค้นหาคนใหม่ ผกก.สส.น.8 กล่าวด้วยว่า จากการวิเคราะห์สภาพศพในเบื้องต้น เขี่ย
 สาวสูงประมาณ 150 ซม.เศษ หน้าตาดูแล้วไม่น่าจะถึง 35 ปี ส่วนชื่อที่ผู้ตายบอกก่อนตายว่า “จำปี มี
 ชัย” อายุ 35 ปี ชาว จ.ร้อยเอ็ด หลังจากเจ้าหน้าที่ตรวจสอบชื่อของผู้หญิงไปแล้วไม่น่าจะใช่ ได้หัน
 มาตรวจ สอบรายชื่อของผู้ชายบ้าง เพราะอาจจะเป็นชื่อของโซเฟอร์ แท็กซี่ที่ก่อเหตุก็เป็นได้
 นอกจากนี้ยังให้ชุดสืบสวนตรวจ สอบหญิงขายบริการทางเพศในละแวกเกิดเหตุ ย่านถนนพระราม
 2 วงเวียนใหญ่ พระประแดง ว่ามีสาวขายบริการคนใดหายตัวไปบ้าง สำหรับข้อมูลที่พยานระบุว่า
 เห็นรถแท็กซี่เขียวเหลืองออกมาจากซอยเกิดเหตุ ตรวจสอบแล้วพบว่า มีอยู่ 2 ยี่ห้อ คือ โตโยต้า และ
 นิสสัน แต่ชุดสืบสวนประมวลจากคำบอกเล่าของพยานแล้วเชื่อว่าน่าจะเป็นรถนิสสัน ซาลูน ที่มาน
 ผ้าสีฟ้าติดด้านหลังมากกว่า ขณะนี้อยู่ระหว่างตรวจสอบตามอุ้งรถแท็กซี่ต่างๆ ย่านบางปะกอก บาง
 ปะแก้ว และราษฎร์บูรณะให้ชัดเจน นอกจากนี้กำลังรอผลการตรวจ พิสูจน์ของแพทย์ รพ.ศิ*ราช
 โดยเจ้าหน้าที่ฝ่ายสืบสวนได้ทำหนังสือขอให้ตรวจศพอย่างละเอียด โดยเฉพาะในกระเพาะอาหาร
 และเลือดให้มากที่สุดเท่าที่จะทำได้

ผลการตัดคำ ศาล/จำ/คุก/แม่เล้า/ 2/ ปี/ ส่ง/สาว/ค้า/กาม/มา/เลย/
 ที่/ศาลอาญา/กรุงเทพ/ใต้/ วัน/ที่/ 1/ พ.ย./ ศาล/อ่าน/คำพิพากษา/ใน/คดี/นาง/
 สา/ริ/ศา/ เนื่อง/สม/ศรี/ อายุ/ 34/ ปี/ บ้าน/เดิม/อยู่/ที่/ ต./ขาม/เฒ่า/พัฒนา
 / อ./กัณฑ์/วิชัย/ จ./มหาสารคาม/ เป็น/จำเลย/ฐาน/เป็น/ธุระ/จัดหา/หญิง/เพื่อ/
 สนอง/ความ/ใคร่/ของ/ผู้/อื่น/ คดี/นี้/จำเลย/ให้การ/รับ/สารภาพ/มา/ตลอด/ และ
 /ยื่น/หลักทรัพย์/ 4/ แสน/บาท/ เพื่อ/ขอ/ประกัน/ตัว/ระหว่าง/พิจารณา/แต่/ศาล/
 ไม่/อนุญาต/ คดี/นี้/โจทก์/ฟ้อง/ว่า/ ดำรง/กอง/ส/วิศ/ดิ/ภาพ/เด็ก/และ/เยาวชน
 /สืบ/ทราบ/ว่า/ มี/คน/ไทย/ทำตัว/เป็น/เอ/เจนต์/ส่ง/หญิง/สาว/หน้าตา/ดี/ ไป/
 ทำ/งาน/ขาย/บริการ/ทาง/เพศ/ ที่/ร้าน/อาหาร/ชื่อ/ลี/ซี/ง/คา/รา/โอ/เกะ/ อยู่/
 ใน/รัฐ/ยะ/โฮร์/ ประเทศ/มาเลเซีย/ โดย/ให้/หญิง/ไทย/ขายตัว/วัน/ละ/ 3/ รอบ
 / ได้/ค่า/ตอบแทน/รอบ/ละ/ 1,000/ บาท/ แต่/มี/ข้อแม้/ว่า/ต้อง/ทำ/งาน/ครบ/
 2/ ปี/ ดำรง/สค./จึง/ประสาน/ไป/ยัง/กอง/ตำรวจ/รถไฟ/ วาง/แผน/จับกุม/
 กระทั่ง/วัน/ที่/ 9/ ธ.ค./ 2547/ เวลา/ 20.00/ น./ ดำรง/สามารถ/ดัก/จับกุม
 /นาง/สา/ริ/ศา/ได้/บน/ขบวน/รถไฟ/สาย/กรุงเทพ/ฯ-/หาด/ใหญ่/ ตรวจ/ค้น/ตัว/

พบ/ตัว/รถไฟ/ 2/ ใบ/ หนังสือ/ เดินทาง/ 2/ เล่ม/ โทรศัพท์/มือ/ถือ/ และ/
บัญชี/เงินสด/ของ/ ธนาคารพาณิชย์/ นาง/สา/ริ/ศา/รับ/สารภาพ/ว่า/กำลัง/จะ/พา/
หญิง/สาว/ไป/ส่ง/ที่/ประเทศ/มาเลเซีย/จริง/ ศาล/พิเคราะห์/แล้ว/เห็น/ว่า/โจทก์/
มี/พยาน/มา/สืบ/ฟัง/ได้/ว่า/ นาง/สา/ริ/ศา/เป็น/หนึ่ง/ใน/ขบวนการ/ส่ง/หญิง/
สาว/คน/ไทย/ไป/ขาย/บริการ/ที่/ประเทศ/มาเลเซีย/จริง/ จึง/มี/ความ/ผิด/ฐาน/
เป็น/ธุระ/จัดหา/หญิง/เพื่อ/สนอง/ความ/ใคร่/ผู้/อื่น/ พิพากษา/ลงโทษ/จำ/คุก/ 4/
ปี/ คำ/รับ/สารภาพ/เป็น/ประโยชน์/คง/รับ/โทษ/จำ/คุก/ 2/ ปี/ หลัง/ฟัง/คำ/
พิพากษา/นาง/สา/ริ/ศา/ได้/ถูก/คุม/ตัว/ไป/ขัง/ที่/ทัณฑสถาน/หญิง/ แม้/พยายาม/
ยื่น/คำ/ร้อง/ขอ/ประกัน/ตัว/ แต่/ศาล/ไม่/อนุญาต/ โดย/ให้/เหตุผล/ว่า/จำเลย/มี/
พฤติการณ์/ร้ายแรง/เป็น/ภัย/ต่อ/สังคม/

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน ศาลจำคุก* 2 ปี ส่งสาวค้ากามมาเลย์

ที่ศาลอาญากรุงเทพใต้ วันที่ 1 พ.ย. ศาลอ่านคำพิพากษาในคดีนางสา*ศา เนื่องสมศรี อายุ 34 ปี บ้าน
เดิมอยู่ที่ ต.ขามเต่าพัฒนา อ.กันทรวิชัย จ.มหาสารคาม เป็นจำเลยฐานเป็นธุระจัดหาหญิงเพื่อสนอง
ความใคร่ของผู้อื่น คดีนี้จำเลยให้การรับสารภาพมาตลอด และยื่นหลักทรัพย์ 4 แสนบาท เพื่อขอ
ประกันตัวระหว่างพิจารณาแต่ศาลไม่อนุญาต คดีนี้โจทก์ฟ้องว่า ตำรวจกองสวัสดิภาพเด็กและ
เยาวชนสืบทราบว่ามีคนไทยทำตัวเป็นเอเย่นต์ส่งหญิงสาวหน้าตาดี ไปทำงานขายบริการทางเพศ ที่
ร้านอาหารชื่อลีซิงคาราโอเกะ อยู่ในรัฐยะโฮร์ ประเทศมาเลเซีย โดยให้หญิงไทยขายตัววันละ 3
รอบ ได้ค่าตอบแทนรอบละ 1,000 บาท แต่มีข้อแม้ว่าต้องทำงานครบ 2 ปี ตำรวจสด.จึงประสานไป
ยังกองตำรวจรถไฟ วางแผนจับกุม กระทั่งวันที่ 9 ธ.ค. 2547 เวลา 20.00 น. ตำรวจสามารถดักจับกุม
นางสา*ศาได้บนขบวนรถไฟสายกรุงเทพฯ-หาดใหญ่ ตรวจค้นตัวพบตัวรถไฟ 2 ใบ หนังสือ
เดินทาง 2 เล่ม โทรศัพท์มือถือ และบัญชีเงินสดของ ธนาคารพาณิชย์ นางสา*ศาได้รับสารภาพว่ากำลัง
จะพาหญิงสาวไปส่งที่ประเทศมาเลเซียจริง ศาลพิเคราะห์แล้วเห็นว่าโจทก์มีพยานมาสืบฟังได้ว่า
นางสา*ศาเป็นหนึ่งในขบวนการส่งหญิงสาวคนไทยไปขายบริการที่ประเทศมาเลเซียจริง จึงมี
ความผิดฐานเป็นธุระจัดหาหญิงเพื่อสนองความใคร่ผู้อื่น พิพากษาลงโทษจำคุก 4 ปี คำรับสารภาพ
เป็นประโยชน์คงรับโทษจำคุก 2 ปี หลังฟังคำพิพากษานางสา*ศาได้ถูกคุมตัวไปขังที่ทัณฑสถาน
หญิง แม้พยายามยื่นคำร้องขอประกันตัว แต่ศาลไม่อนุญาต โดยให้เหตุผลว่าจำเลยมีพฤติการณ์
ร้ายแรงเป็นภัยต่อสังคม

ผลการตัดคำ ล่อ/ซื้อ/จับ/คา/โรงแรม/แม่/พา/ลูก/สาว/วัย/ 15/ ค้า/กาม/
เมื่อ/เวลา/ 19.00/ น./ วัน/ที่/ 1/ พ.ย./ พ.ศ.อ.เร/วัช/ กลิ่น/เก/ษร/ รอง/
ผ/บก./กลุ่ม/งาน/สืบสวน/สอบสวน/ภาค/ 7/ พ.ต.อ./สม/ชาย/ รัก/เสนาะ/ ผก
/ก./กลุ่ม/งาน/สืบสวน/สอบสวน/ภาค/ 7/ พ.ต.ท./ปรีดา/ อิ่ม/เจริญ/ สว./.
งาน/ 3/ กก./สส.ภ./ 7/ สืบ/ทราบ/ว่า/มี/ผู้หญิง/พา/ลูก/สาว/ กับ/เพื่อน/
ของ/ลูก/เร/ชาย/บริการ/ทาง/เพศ/ จึง/วาง/แผน/ให้/สาย/โทรศัพท์/ ติดต่อ/ล่อ/
ซื้อ/บริการ/ โดย/นัดหมาย/กัน/ที่/ห้อง/เลข/ที่/ 314/ และ/ 315/ โรงแรม/คัน/
ท/รี/ ถนน/สาย/ตลิ่ง/ชั้น/-/สุพรรณบุรี/ หมู่/ 3/ ต./ท่า/ระหัด/ อ./เมือง/
สุพรรณบุรี/ ก่อน/ถึง/เวลา/นัดหมาย/ ดำรวจ/ได้/กระจาย/กำลัง/ประจำ/จุด/ต่าง/ๆ/
พร้อม/ถ่าย/สำเนา/ธนบัตร/สำหรับ/จ่าย/เป็น/เงิน/ค่าตัว/สาว/บริการ/ 2/ คน/ คน
ละ/ 2,000/ บาท/ รวม/ 4,000/ บาท/ ไว้/เป็น/หลักฐาน/ พร้อม/เตรียม/เท/ป/
บันทึก/การ/สนทนา/ระหว่าง/สาย/ล่อ/ซื้อ/กับ/กลุ่ม/ผู้/ค้า/บริการ/ จนกระทั่ง/ถึง/
เวลา/นัดหมาย/ ได้/มี/ผู้หญิง/ที่/เป็น/แม่/เล้า/ ทราบ/ชื่อ/นาง/ตุ้/ก/ (/นาม/
สมมติ/)/ อายุ/ 40/ ปี/ บ้าน/อยู่/ ต.กฤษ/นา/ อ./บาง/ปลา/ม้า/ จ./
สุพรรณบุรี/ ได้/พา/ น.ส./เอ/ (/นาม/สมมติ/)/ อายุ/ 18/ ปี/ ลูก/สาว/ และ
/ น.ส.บี/ (/นาม/สมมติ/)/ อายุ/ 15/ ปี/ เพื่อน/ของ/ น.ส./เอ/ มา/ส่ง/ให้/
สาย/ล่อ/ซื้อ/ที่/ห้อง/หมายเลข/ 314/ และ/ 315/ ของ/ โรงแรม/คัน/ท/รี/
หลังจาก/จ่าย/เงิน/ค่าตัว/เป็น/ที่/เรียบร้อย/ สาย/ล่อ/ซื้อ/ได้/ส่ง/สัญญาณ/ให้/
ตำรวจ/จู่โจม/เข้า/จับกุม/ น.ส./เอ/ และ/ น.ส.บี/ ได้/ภายใน/ห้อง/ ขณะ/ที่/
กำลัง/อีก/ส่วน/เข้า/จับกุม/นาง/ตุ้/ก/ได้ที่/ลิ/อบ/บี้/ของ/โรงแรม/ทันที/ที่/ถูก/
จับกุม/นาง/ตุ้/ก/ให้การ/ปฏิเสธ/ว่า/ไม่ได้/เป็น/คน/พา/ลูก/สาว/กับ/เพื่อน/ลูก/
สาว/มา/ค้า/บริการ/ แต่/ทั้ง/คู่/มา/ขาย/บริการ/กันเอง/ ตำรวจ/จึง/นำ/เท/ป/
บันทึก/การ/สนทนา/มา/เปิด/ให้/ฟัง/ ใน/ที่สุด/ก็/ยอม/จำนน/รับ/สารภาพ/ว่า/
พา/ลูก/สาว/กับ/เพื่อน/ของ/ลูก/มา/ขาย/บริการ/ทาง/เพศ/หลาย/ครั้ง/เป็น/เวลา/
นาน/นับ/ปี/ โดย/เงิน/ค่าตัว/ของ/ลูก/สาว/จะ/ให้/เต็ม/ทั้ง/ 2,000/ บาท/ ส่วน
/เงิน/ค่าตัว/ของ/ น.ส.บี/ จะ/หัก/ 1,000/ บาท/ เป็น/ค่า/นายหน้า/ติดต่อ/ลูก/ค้า/
เมื่อ/ตำรวจ/สอบ/ถาม/ถึง/ลูก/ค้า/ที่/เคย/ซื้อ/บริการ/ นาง/ตุ้/ก/ปี/ค/ปาก/เสียง/ไม่
/ยอม/บอก/อ้าง/ว่า/ไม่/รู้จัก/และ/จำ/ไม่/ได้/แล้ว/ แต่/ส่วน/ใหญ่/เป็น/พวก/
เล็/ยก/ระ/เป้า/หนัก/ด้าน/ น.ส./เอ/เปิด/ปาก/รับ/สารภาพ/ว่า/ คิด/ยา/บ้า/มา/
นาน/นับ/ปี/และ/ไม่/ได้/เรียน/หนังสือ/ ที่/ให้/แม่/พา/มา/ขาย/บริการ/ ทำ/ไป

/ ด้วย / ความ / เต็มใจ / ไม่ / ได้ / ถูก / บีบบังคับ / แต่ / อย่าง / ใด / ต้องการ / หาเงิน / ไป / ซื้อ / ยา / บ้า / เสพ / และ / หา / ซื้อ / เสื้อ / ผ้า / สวย ๆ / ไป / ไป / เทียว / เตรี / ส่วน / น.ส. / จอ / ย / ให้การ / ว่า / ไม่ / ได้ / เรียน / หนังสือ / เช่น / กัน / สาเหตุ / ที่ / ออก / มา / ขาย / บริการ / ทาง / เพศ / เพื่อ / ต้องการ / หาเงิน / ซื้อ / เสื้อ / ผ้า / และ / ไป / เทียว / กับ / เพื่อน ๆ / หลัง / สอบสวน / เบื้องต้น / ควบคุม / ตัวผู้ / ต้องหา / ทั้ง / 3 / คน / ส่ง / สภ.อ. / เมือง / สุพรรณบุรี / ดำเนินคดี / ต่อ / ไป /

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน ล่อซื้อจับคาโรงแรม แม่พาลูกสาววัย 15 คำกาม
 เมื่อเวลา 19.00 น. วันที่ 1 พ.ย. พ.ต.อ.เรวัช กลิ่นเกษร รอง ผบก.กลุ่มงานสืบสวนสอบสวนภาค 7 พ.ต.อ.สมชาย รักเสนาะ ผกก.กลุ่มงานสืบสวนสอบสวน ภาค 7 พ.ต.ท.ปรีดา อิ่มเจริญ สว.งาน 3 กก.สส.ภ. 7 สืบทราบว่ามีผู้หญิงพาลูกสาว กับเพื่อนของลูกเร่ขายบริการทางเพศ จึงวางแผนให้สายโทรศัพท์ ติดต่อล่อซื้อบริการ โดยนัดหมายกันที่ห้องเลขที่ 314 และ 315 โรงแรมคันทรี่ ถนนสายตลิ่งชัน-สุพรรณบุรี หมู่ 3 ต.ท่าระหัด อ.เมืองสุพรรณบุรี ก่อนถึงเวลานัดหมาย ดำรวจได้กระจายกำลังประจำจุดต่างๆ พร้อมถ่ายสำเนาธนบัตรสำหรับจ่ายเป็นเงินค่าตัวสาวบริการ 2 คน คนละ 2,000 บาท รวม 4,000 บาท ไว้เป็นหลักฐาน พร้อมเตรียมเทปบันทึกการสนทนาระหว่างสายล่อซื้อ กับกลุ่มผู้ค้าบริการ จนกระทั่งถึงเวลานัดหมาย ได้มีผู้หญิงที่เป็น* ทราบชื่อนางตุ๊ก (นามสมมติ) อายุ 40 ปี บ้านอยู่ ต.กฤษณา อ.บางปลาม้า จ.สุพรรณบุรี ได้พา น.ส.เอ (นามสมมติ) อายุ 18 ปี ลูกสาว และ น.ส.บี (นามสมมติ) อายุ 15 ปี เพื่อนของ น.ส.เอ มาส่งให้สายล่อซื้อที่ห้องหมายเลข 314 และ 315 ของโรงแรมคันทรี่ หลังจากจ่ายเงินค่าตัวเป็นที่เรียบร้อย สายล่อซื้อได้ส่งสัญญาณให้ตำรวจดูโจมเข้าจับกุม น.ส.เอ และ น.ส.บี ได้ภายในห้อง ขณะที่กำลังอีกส่วนเข้าจับกุมนางตุ๊กได้ที่ลิโอบบี้ของโรงแรม ทันทีที่ถูกจับกุม นางตุ๊กให้การปฏิเสธว่าไม่ได้ เป็นคนพาลูกสาวกับเพื่อนลูกสาวมาค้าบริการ แต่ทั้งคู่มารขายบริการกันเอง ตำรวจจึงนำเทปบันทึกการสนทนา มาเปิดให้ฟัง ในที่สุดก็ยอมจำนนรับสารภาพว่า พาลูกสาวกับเพื่อนของลูกมาขายบริการทางเพศหลายครั้งเป็นเวลานานนับปี โดยเงินค่าตัวของลูกสาวจะให้เต็มทั้ง 2,000 บาท ส่วนเงินค่าตัวของ น.ส.บี จะหัก 1,000 บาท เป็นค่านายหน้าติดต่อลูกค้า เมื่อตำรวจสอบถามถึงลูกค้าที่เคยซื้อบริการ นางตุ๊กปิดปากเงียบไม่ยอมบอก อ้างว่าไม่รู้จักและจำไม่ได้แล้ว แต่ส่วนใหญ่เป็นพวกเสี่ยกระเป่าหนัก ด้าน น.ส.เอเปิดปากรับสารภาพว่า ดิฉันขี้มา นานนับปีและไม่ได้เรียนหนังสือ ที่ให้แม่พามาขายบริการ ทำไปด้วยความเต็มใจ ไม่ได้ถูกบีบบังคับแต่อย่างใด ต้องการหาเงินไปซื้อยาบ้าเสพ และหาซื้อเสื้อผ้าสวยๆ ไปไป เทียวเตร่ ส่วน น.ส.จอยให้การว่า ไม่ได้เรียนหนังสือเช่นกัน สาเหตุที่ออกมาขายบริการทางเพศเพื่อต้องการ หาเงินซื้อเสื้อผ้าและไปเที่ยวกับเพื่อนๆ หลังสอบสวนเบื้องต้นควบคุมตัวผู้ต้องหาทั้ง 3 คน ส่ง สภ.อ.เมืองสุพรรณบุรี ดำเนินคดีต่อไป

ผลการตัดคำ ฆ่า / โหด / - / ทบ / สาว / ถู / คลุม / หัว / ศพ / หมก / พง / หล้า /
เมื่อ / เวลา / 08.00 / น. / วัน / ที่ / 1 / พ.ย. / . / พ.ต.ต. / สุว / ฤทธิ์ / พร้อมมูล / สารวัตร /
เวร / สก. / ย่อย / ปฐม / พร / อ. / เมือง / ชุมพร / ได้ / รับ / แจ้ง / เหตุ / พบ / ศพ / หญิง / สาว /
ถูก / ฆาตกรรม / ใน / พง / หล้า / ริม / ถนน / สาย / สระ / ยักษ์ / - / ถ้ำ / แก้ว / หมู่ / 12 /
บ้าน / สระ / ยักษ์ / ต. / บ้าน / นา / อ. / เมือง / ชุมพร / จึง / รายงาน / ให้ / พ.ต.อ. / นรินทร์ /
บุษย / วิ / ทย์ / ผก / ก. / . / พ.ต.ท. / โชค / ดี / รัก / รงค์ / สว / ป. / . / พ.ต.ต. / เกรียง / ไกร /
โพธิ์ / ประสิทธิ์ / สว. / สส. / . / นำ / กำลัง / ดำรง / พร้อม / แพทย์ / เวร / รพ. / . / ชุมพร /
เขต / อุดม / ศักดิ์ / และ / หน่วย / กู้ภัย / สาย / ชล / ชุมพร / ไป / ตรวจ / สอบ / ที่ / เกิด / เหตุ /
อยู่ / ใกล้ / สวน / ทูเรียน / ของ / ชาว / บ้าน / เป็น / กู / น้ำ / แห่ง / ขอด / มี / พง / หล้า / ขึ้น / ปก
คลุม / พบ / ศพ / เขื่อ / ฆาตกรรม / เป็น / หญิง / สาว / ไม่ / ทราบ / ชื่อ / อายุ / ประมาณ /
20-30 / ปี / รูปร่าง / อวบ / ผิว / ขาว / ตัด / ผม / ซอย / สั้น / ย้อม / สี / แดง / หู / ทั้ง / 2 /
ข้าง / เจาะ / ข้าง / ละ / 2 / รู / ใส่ / ตุ่มหู / ทอง / ทั้ง / 4 / รู / นิ้ว / นาง / ซ้าย / สวม /
แหวน / ทอง / 1 / วง / นุ่ง / กระโปรง / สั้น / สี / น้ำตาล / อ่อน / สวม / เสื้อ / แขน / กุด / สี /
ขาว / คอ / ระบาย / ลาย / ลูกไม้ / ถูก / ถลก / ขึ้น / มา / ได้ / ราว / นม / ใน / เสื้อ / ขึ้น / ใน / มี
/ เงินสด / 140 / บาท / ซุก / อยู่ / นอก / นั้น / ไม่ / พบ / หลักฐาน / ว่า / เป็น / ใคร / มา / จาก /
ไหน / สภาพ / ศพ / นอน / กุด / คู้ / คอ / หน้า / หุ่น / ที่ / บัง / แดด / รยยนต์ / สี / ดำ / มี / ถู /
พลาสติก / หู / หิ้ว / ของ / ห้าง / สรรพ / สินค้า / สินค้า / ฟ้าย / ที่ / ถู / เขียน / ด้วย / ปากกา / เม /
จิก / ว่า / “ / ภา / วิ / ณี / วงศ์ / รัตน / สกุล / ” / และ / เขียน / หมายเลข / 607 / 1 / ตัว /
30 / บาท / คลุม / ศีรษะ / อยู่ / แพทย์ / ชันสูตร / ศพ / เบื้องต้น / พบ / บาดแผล / ถูก / ทบ /
ด้วย / ของแข็ง / ที่ / โบกหน้า / จน / เละ / แทะ / จำ / เค้า / เดิม / ไม่ / ได้ / หน้าผาก / ยุบ / และ /
โหนกแก้ม / ขวา / มี / แผล / แตก / ลึก / ถึง / กระดูก / เลือด / แห่ง / เกราะกรัง / คาด / ว่า / เสีย
/ ชีวิต / มา / แล้ว / ไม่ / ต่ำ / กว่า / 10 / ชั่วโมง / ยัง / ไม่ / สามารถ / ตรวจ / หา / ร่องรอย / การ
/ ถูก / ข่ม / ขึ้น / ได้ / จึง / นำ / ศพ / ส่ง / รพ. / . / ชุมพร / เขต / อุดม / ศักดิ์ / ชันสูตร / อย่าง /
ละเอียด / อีก / ครั้ง / ส่วน / บริเวณ / โดย / รอบ / ที่ / เกิด / เหตุ / ไม่ / มี / ร่องรอย / การ / ต่อสู้ /
จาก / การ / สอบ / ถาม / ชาว / บ้าน / ละแวก / ใกล้ / เคียง / ทราบ / ว่า / เมื่อ / ตอน / หัวค่ำ / วัน /
ที่ / 31 / ต.ค. / . / ขณะ / ฝน / ตกหนัก / ได้ยิน / เสียง / รยยนต์ / ขับ / เข้าไป / จอด / บริเวณ /
จุด / ที่ / พบ / ศพ / สัก / พัก / ก็ / รีบ / ร้อน / ขับ / ออก / ไป / แต่ / ไม่ / มี / ใคร / สนใจ / จน /
รุ่งเช้า / จึง / ไป / พบ / ศพ / ถูก / ฆาตกรรม / โหด / จาก / สภาพ / การ / แต่ง / กาย / ของ / เขื่อ /
สาว / ดำรง / สันนิษฐาน / ว่า / น่า / จะ / เป็น / นัก / ร้อง / หรือ / หญิง / บริการ / ตาม / ห้อง /
อาหาร / หรือ / คา / เฟ่ / ใน / ตัวเมือง / ชุมพร / ถูก / คนร้าย / ฆ่า / จาก / ที่ / อื่น / แล้ว / นำ / ศพ /

ใส่ / รถ / มา / ทิ้ง / อำพราง / คดี / โดย / ชนวน / เหตุ / มา / จาก / เรื่อง / ชู้สาว / หรือ / ไม่ / ก็ / ความ / โกรธ / แค้น / ส่วนตัว / จะ / ได้ / สืบสวน / ติดตาม / หา / ญาติ / ผู้ / ตาย / มา / สอบ / ปากคำ / เป็น / แนวทาง / คลี่คลาย / คดี / ต่อ / ไป /

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน มา*-ทูปสาวถูกคลุมหัวศพหมกพงหญ้า

เมื่อเวลา 08.00 น.วันที่ 1 พ.ย. พ.ต.ต.สุวฤทธิ์ พร้อมมูล สารวัตรเวร สภ.ย่อยปฐมพร อ.เมืองชุมพร ได้รับแจ้งเหตุพบศพหญิงสาว ถูกฆาตกรรม ในพงหญ้าริมถนนสายสระยักษ-ถ้ำแก้ว หมู่ 12 บ้านสระยักษ ต.บ้านนา อ.เมืองชุมพร จึงรายงานให้ พ.ต.อ.นรินทร์ บุญวิทย์ ผกก. พ.ต.ท.โชคดี รักรงค์ สวป. พ.ต.ต.เกรียงไกร โพธิ์ประสิทธิ์ สว.สส. นำกำลังตำรวจ พร้อมแพทย์เวร รพ.ชุมพรเขตอุดมศักดิ์ และหน่วยกู้ภัยสายชลชุมพร ไปตรวจสอบ ที่เกิดเหตุอยู่ใกล้สวนทุเรียนของชาวบ้าน เป็นคูน้ำแห้งขอมมีพงหญ้าขึ้นปกคลุม พบศพเหยื่อฆาตกรรมเป็นหญิงสาวไม่ทราบชื่อ อายุประมาณ 20-30 ปี รูปร่างอวบ ผิวขาว ตัดผมซอยสั้นย้อมสีแดง หูทั้ง 2 ข้างเจาะข้างละ 2 รู ใส่ตุ้มหูทองทั้ง 4 รู นิ้วนางซ้ายสวมแหวนทอง 1 วง นุ่งกระโปรงสั้นสีน้ำตาลอ่อน สวมเสื้อแขนกุดสีขาว คอระบายลายลูกไม้ ถูกถลกขึ้นมาได้ราวม ในเสื้อชั้นในมีเงินสด 140 บาทซุกอยู่ นอกนั้นไม่พบหลักฐานว่าเป็นใครมาจากไหน สภาพศพนอนคุดคู้คว่ำหน้าหันที่บังแดดรถยนต์สีดำ มีถุงพลาสติกหุ้มหัวของข้างสรรพสินค้า สีนฟ้า ที่ถุงเขียนด้วยปากกาเมจิกว่า “ภาวิณี วงศ์รัตนสกุล” และเขียนหมายเลข 607 1 ตัว 30 บาท คลุมศีรษะอยู่ แพทย์ชันสูตรศพเบื้องต้นพบบาดแผลถูกทุบด้วยของแข็งที่ใบหน้าจนและแทบจำเค้าเดิมไม่ได้ หน้าผากยุบ และโหนกแก้มขวามีแผลแตกลึกถึงกระดูก เลือดแห้งเกราะกรัง คาดว่าเสียชีวิตมาแล้วไม่ต่ำกว่า 10 ชั่วโมง ยังไม่สามารถตรวจหาร่องรอยการถูกข่มขืนได้ จึงนำศพส่ง รพ.ชุมพรเขตอุดมศักดิ์ ชันสูตรอย่างละเอียดอีกครั้ง ส่วนบริเวณโดยรอบที่เกิดเหตุไม่มีร่องรอยการต่อสู้ จากการสอบถามชาวบ้านละแวกใกล้เคียงทราบว่า เมื่อตอนหัวค่ำวันที่ 31 ต.ค. ขณะฝนตกหนัก ได้ยินเสียงรถยนต์ขับเข้าไปจอดบริเวณจุดที่พบศพ สักพักก็รีบร้อนขับออกไป แต่ไม่มีใครสนใจ จนรุ่งเช้าจึงไปพบศพถูกฆาตกรรม* จากสภาพการแต่งกายของเหยื่อสาว ตำรวจสันนิษฐานว่า น่าจะเป็นนักร้องหรือหญิงบริการตามห้องอาหารหรือคาเฟ่ในตัวเมืองชุมพร ถูกคนร้ายฆ่าจากที่อื่นแล้วนำศพใส่รถมาทิ้งอำพรางคดี โดยชนวนเหตุมาจากเรื่องชู้สาวหรือไม่ก็ความโกรธแค้นส่วนตัว จะได้สืบสวนติดตามหาญาติผู้ตายมาสอบปากคำเป็นแนวทางคลี่คลายคดีต่อไป

ผลการตัดคำ ทักมึณ / รับปาก / รับ / ซื่อ / หนี / เน่า / จาก / ชาว / บ้าน /

ผู้สื่อข่าว / รายงาน / ว่า / ภายหลัง / การ / ประชุม / ค / รม / . / เมื่อ / วัน / ที่ / 1 / พ.ย. / . / พ.ต.ท. / ทักมึณ / ได้ / เดินทาง / ไป / พบ / กับ / กลุ่ม / ผู้ / ชุมนุม / เครือ / ข่าย / หนีสิน / ชาว / นา / แห่ง / ประเทศ / ไทย / ภาค / เหนือ / จำนวน / ประมาณ / 3,000 / คน / ที่มา / ชุมนุม / เรียก

/ ร้อง / ให้ / นายก / ๑ / เร่ง / แก้ไข / ปัญหา / หนี้สิน / เกษตรกร / ที่ / บริเวณ / หน้า / วัด / คิริ / วงศ์ / โดย / พ.ต.ท. / ทักษิณ / ได้ / กล่าว / กับ / กลุ่ม / ผู้ / ชุมชุม / ว่า / สิ่ง / ที่ / พุด / ไว้ / ไม่ / เคย / ลืม / วัน / ที่ / 2 / พ.ย. / นี้ / จะ / เป็น / ประชาน / ประชุม / คณะ / กรรมการ / อง / ทุน / พื้นฟู / และ / พัฒนา / เกษตรกร / ด้วย / ตัว / เอง / และ / จะ / มี / มติ / ดี / ๆ / ออก / มา / โดย / เฉพาะ / ผู้ / ที่ / มี / ปัญหา / ถูก / ฟ้อง / ร้อง / ถูก / ยึด / ทรัพย์ / รัฐบาล / จะ / ถอน / ออก / มา / หมด / และ / เรา / จะ / ชื้อ / หนี้ / ออก / มา / ให้ / หมด / ได้ / ตั้ง / การ / ใน / ๑ / ๒ / ๓ / ๔ / ๕ / ให้ / ไป / จัดการ / เรื่อง / หนี้ / นอก / ระบบ / ของ / คน / จน / ทั้งหมด / “ / ไม่ / ต้อง / ห่วง / ๓ / ๔ / ๕ / ไต / ที่ / นั่ง / เป็น / นายก / ๑ / อยู่ / ตรง / นี้ / มี / อะไร / บอก / ให้ / รู้ / ถ้า / อะไร / ช่วย / ได้ / ช่วย /ทันที / อะไร / ทำให้ / คน / ไทย / ยืน / บน / ลำ / แขนง / ตัว / เอง / ได้ / ไม่ / ให้ / คน / จน / ถูก / ข่ม / เหวง / รังแก / วัน / นี้ / ได้ / บอก / ใน / ๑ / ๒ / ๓ / ๔ / ๕ / ว่า / ไอ้ / พวก / ที่ / ให้ / กู้ / เงิน / 2 / หมื่น / ทำ / สัญญา / 2 / แสน / แล้ว / ไป / ยึด / ที่ / นา / ชาว / บ้าน / ให้ / ไป / ช่วย / ทำให้ / มัน / จน / ลง / หน้อย / ได้ / ใหม่ / ประเภท / นี้ / ผม / ไม่ / ชอบ / ประเภท / ที่ / โกง / ชาว / บ้าน / ตา / คำ / ๆ / ที่ / เทียว / ไป / ยึด / ไร่ / นา / เขา / ขอ / ให้ / พวก / เรา / กลับ / ไป / ทำ / มา /หากิน / รัฐบาล / จะ / ดูแล / ให้ / ” / นายก / ๑ / กล่าว /

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน ทักษิณรับปากรับซื้อหนี้เน่าจากชาวบ้าน ผู้สื่อข่าวรายงานว่า ภายหลังจากประชุม ครม. เมื่อวันที่ 1 พ.ย. พ.ต.ท.ทักษิณได้เดินทางไปพบกับกลุ่มผู้ชุมนุม เครือข่ายหนี้สินชาวนาแห่งประเทศไทยภาคเหนือ จำนวนประมาณ 3,000 คน ที่มาชุมนุมเรียกร้องให้นายกฯเร่งแก้ไขปัญหาหนี้สินเกษตรกร ที่บริเวณหน้าวัดคิริวงศ์ โดย พ.ต.ท.ทักษิณได้กล่าวกับกลุ่มผู้ชุมนุมว่า สิ่งที่เราพูดไว้ไม่เคยลืม วันที่ 2 พ.ย. นี้ จะเป็นประชานประชุม คณะกรรมการกองทุนฟื้นฟูและพัฒนาเกษตรกรด้วยตัวเอง และจะมีมติดี ๆ ออกมา โดยเฉพาะผู้ที่มีปัญหาถูกฟ้องร้อง ถูกยึดทรัพย์ รัฐบาลจะถอนออกมาหมด และเราจะซื้อหนี้ ออกมาให้หมด ได้ตั้ง การใน ครม. ให้ไปจัดการเรื่องหนี้ นอก ระบบ ของ คน จน ทั้งหมด “ ไม่ ต้อง ห่วง ๓ ๔ ๕ ไต ที่ นั่ง เป็น นายก ๑ อยู่ ตรง นี้ มี อะไร บอก ให้ รู้ ถ้า อะไร ช่วย ได้ ช่วยทันที อะไร ทำให้ คน ไทย ยืน บน ลำ แขนง ตัว เอง ได้ ไม่ ให้ คน จน ถูก ข่ม เหวง รังแก วัน นี้ ได้ บอก ใน ครม. ว่า * พวก ที่ ให้ กู้ เงิน 2 หมื่น ทำ สัญญา 2 แสน แล้ว ไป ยึด ที่ นา ชาว บ้าน ให้ ไป ช่วย ทำให้ มัน จน ลง หน้อย ได้ ใหม่ ประเภท นี้ ผม ไม่ ชอบ ประเภท ที่ โกง ชาว บ้าน ตา คำ ๆ ที่ เทียว ไป ยึด ไร่ นา เขา ขอ ให้ พวก เรา กลับ ไป ทำ มา หากิน รัฐบาล จะ ดูแล ให้ ” นายก ๑ กล่าว

ผลการตัดคำ นาง/แบบ/ผู้ด / ' / ลูก / ตาล / ' / โคน / ต้น / เซ่า / คอน / โด / ๗ / สูญ / เกือบ / แสน /

เมื่อ / เวลา / 13.00 / น. / วัน / ที่ / 8 / พ.ย. / น.ส. / ทิพย์ / วรรณ / หรือ / ลูก / ตาล / - / ชโลม / จิต / จันทร / เกตุ / อายุ / 33 / ปี / นัก / แสดง / และ / นาง / แบบ / แนว / ปลุกใจ / เสือป่า / เดินทาง / เข้า / พบ / ร.ต.อ.ส / วิ / ก / นุช / เจริญ / ผล / ร้อย / เวร / กลุ่ม / งาน / สอบสวน / บก.ป. / เพื่อ / แจ้งความ / ดำเนินคดี / กับ / นาย / พิ / ระ / ศิลป์ / วาณิช / กุล / อายุ / 31 / ปี / ใน / ข้อหา / น้อ โกง / หลังจาก / ได้ / ทำ / สัญญา / เซ่า / ห้อง / พัก / เลข / ที่ / 5/28 / ชั้น / 18 / ปี / ยะ / สาทร / คอน / โด / มิ / เนียม / ซอย / สวน / พลู / แขวง / และ / เขต / สาทร / กทม. / โดย / จ่าย / เงิน / ไป / จำนวน / 75,000 / บาท / แต่ / ไม่ / สามารถ / เข้าไป / พัก / อาศัย / ได้ / โดย / น.ส. / ชโลม / จิต / เผย / ถึง / ต้นเหตุ / ของ / การ / ถูก / โกง / ใน / ครั้ง / นี้ / ว่า / ได้ / อ่าน / พบ / ประกาศ / ให้ / เซ่า / ห้อง / พัก / ปี / ยะ / สาทร / คอน / โด / มิ / เนียม / ใน / นิตยสาร / อาคาร / และ / ที่ดิน / ฉบับ / หนึ่ง / แล้ว / มี / ความ / สนใจ / ที่ / จะ / เซ่า / พัก / เนื่องจาก / บ้าน / ที่ / ซื่อ / ไว้ / ยัง / ตกแต่ง / ไม่ / เสร็จ / ตั้งใจ / ว่า / จะ / พัก / อาศัย / ประมาณ / 6 / เดือน / จึง / โทรศัพท์ / ติดต่อ / กับ / นาย / พิ / ระ / ศิลป์ / ซึ่ง / เป็น / นายหน้า / ให้ / เซ่า / คอน / โด / มิ / เนียม / ดัง / กล่าว / และ / นัด / ไป / ดู / ห้อง / พัก / พร้อม / ตกง / ทำ / สัญญา / เซ่า / เมื่อ / วัน / ที่ / 7 / ต.ค. / ที่ / ผ่าน / มา / ใน / สัญญา / ระบุ / ค่า / เซ่า / เดือน / ละ / 25,000 / บาท / จ่าย / ค่า / มัดจำ / ล่วงหน้า / 3 / เดือน / เป็น / เงิน / 75,000 / บาท / เมื่อ / ทำ / สัญญา / แล้ว / จึง / ได้ / รับ / กุญแจ / ห้อง / และ / คีย์การ์ด / สำหรับ / เข้าค่อ / น / โด / มิ / เนียม / นาง / แบบ / สวาก / ล่า / ว / ว่า / ต่อ / มา / เมื่อ / วัน / ที่ / 13 / ต.ค. / ที่ / ผ่าน / มา / ได้ / ขน / ข้าวของ / เครื่อง / ใช้ / ไป / ที่ / ห้อง / ดัง / กล่าว / แต่ / ปรากฏ / ว่า / มี / นาย / โสภณ / โล / ผู้จัดการ / โครงการ / และ / สถาปนิก / อาวุโส / บริษัท / บี / ล / เด / อร์ / คอน / ซัลแดนส์ / จำกัด / และ / บริษัท / บี / ล / เด / อร์เรี / ย / ดี / ไซน์ / จำกัด / เข้า / มาท / ย / ใน / ห้อง / แจ้ง / ว่า / เขา / เป็น / เจ้าของ / ห้อง / โดย / นำ / หลักฐาน / ต่าง / ๗ / มา / ยืนยัน / ตน / จึง / เล่า / เรื่อง / ที่ / นาย / พิ / ระ / ศิลป์ / ลง / โฆษณา / ให้ / เซ่า / ห้อง / และ / ได้ / จ่าย / เงิน / ค่า / เซ่า / เกือบ / แสน / บาท / ให้ / ฟัง / แต่ / นาย / โสภณ / ไม่ / สนใจ / อ้าง / ว่า / ไม่ / มี / ส่วน / ทราบ / เรื่อง / ที่ / เกิด / ขึ้น / แต่ / อย่าง /ใด / จึง / จำ / ใจ / ต้อง / เก็บข้าว / ของ / กลับ / ออก / มา / และ / พยายาม / ติดต่อ / นาย / พิ / ระ / ศิลป์ / แต่ / ไม่ / สามารถ / ติดต่อ / ได้ / จึง / เชื่อ / ว่า / ถูก / หลอก / น.ส. / ชโลม / จิต / เผย / อีก / ว่า / ได้ / ตรวจ / สอบ / กับ / ห้อง / เซ่า / ที่ / อยู่ / ติด / กัน / ทราบ / ว่า / มี / ชาว / ต่าง / ชาติ / อีก / ราย / ที่ / หลง / เชื้อ / จ่าย / เงิน / ให้ / นาย / พิ / ระ / ศิลป์ / ไป / แล้ว / เช่น / เดียวกัน / พนักงาน / ใน /

คอน / โด / มิ / เนียม / หลาย / คน / บอก / ว่า / เคย / เห็น / นาย / พิ / ระ / ศิลป์ / เข้า / มา / ที่ / คอน / โด / มิ / เนียม / บ่อย / ครั้ง / แต่ / ไม่ / ทราบ / ว่า / เป็น / เจ้าของ / ห้อง / พัก / หรือ / ไม่ / เชื่อ / ว่า / น่า / จะ / มี / ส่วน / รู้ / เห็น / กับ / นาย / พิ / ระ / ศิลป์ / ด้วย / เนื่องจาก / ฤๅญ / แจก / ห้อง / พัก / และ / คีย์การ์ด / เพื่อ / เข้า / คอ / น / โด / มิ / เนียม / ที่ / ได้ / มา / ไม่ / น่า / จะ / ทำ / ปลอม / ขึ้น / มา / ได้ / จึง / ไม่ / ยาก / ให้ / มี / ใคร / หลง / เชื่อ / อีก / เพราะ / ยัง / มี / การ / ลง / โฆษณา / ให้ / เข้า / ห้อง / พัก / ใน / นิตยสาร / อยู่ / เช่น / เดิม / ด้าน / ร.ต.อ.ส / วิ / กก / ล่า / ว / ว่า / เบื้องต้น / ดำรง / ได้ / รับ / เรื่อง / ไว้ / แล้ว / ส่วน / จะ / มี / การ / พิจารณา / ส่ง / ให้ / สน . / / ท่ง / มหาเมฆ / เจ้าของ / ท้องที่ / เกิด / เหตุ / รับ / ไป / ดำเนิน / การ / หรือ / ไม่ / นั้น / จะ / รายงาน / ให้ / ผู้ / บังคับบัญชา / ทราบ / และ / พิจารณา / ส่ง / การ / ตาม / ลำดับ / ชั้น / ต่อ / ไป / ผลการแทนที่คำหายด้วยเครื่องหมายดอกจัน นางแบบนู้ด 'ลูกตาล' โคนตุนเช่าคอนโดฯ สูญเกือบแสน

เมื่อเวลา 13.00 น. วันที่ 8 พ.ย. น.ส.ทิพย์วรรณ หรือลูกตาล-ชโลมจิต จันทร์เกตุ อายุ 33 ปี นักแสดง และนางแบบ แนวปลุกใจเสือป่า เดินทางเข้าพบ ร.ต.อ.สวิก นุชเจริญผล ร้อยเวรกลุ่มงานสอบสวน บก.ป. เพื่อแจ้งความดำเนินคดี กับนายพีระศิลป์ วาณิชกุล อายุ 31 ปี ในข้อหาฉ้อโกง หลังจากได้ทำสัญญาเช่าห้องพักเลขที่ 528 ชั้น 18 ปิยะสาทรคอนโดมิเนียม ซอยสวนพลู แขวงและเขตสาทร กทม. โดยจ่ายเงินไปจำนวน 75,000 บาท แต่ไม่สามารถเข้าไปพักอาศัยได้ โดย น.ส.ชโลมจิตเผยถึง ต้นเหตุของการถูกโกงในครั้งนี้ว่า ได้อ่านพบประกาศให้เช่าห้องพักปิยะสาทร คอนโดมิเนียม ใน นิตยสารอาคารและที่ดินฉบับหนึ่ง แล้วมีความสนใจที่จะเช่าพัก เนื่องจากบ้านที่ซื้อไว้ยังตกแต่งไม่เสร็จ ตั้งใจว่าจะพักอาศัยประมาณ 6 เดือน จึงโทรศัพท์ ติดต่อกับนายพีระศิลป์ ซึ่งเป็นนายหน้าให้เช่าคอนโดมิเนียมดังกล่าว และนัดไปดูห้องพักรวมตกลงทำสัญญาเช่า เมื่อวันที่ 7 ต.ค.ที่ผ่านมา ในสัญญาระบุค่าเช่าเดือนละ 25,000 บาท จ่ายค่ามัดจำล่วงหน้า 3 เดือน เป็นเงิน 75,000 บาท เมื่อทำสัญญาแล้วจึงได้รับฤๅญแจห้องและคีย์การ์ดสำหรับเช่าคอนโดมิเนียม นางแบบสาวกล่าวว่า ต่อมา เมื่อวันที่ 13 ต.ค.ที่ผ่านมา ได้ขนข้าวของเครื่องใช้ไปที่ห้องดังกล่าว แต่ปรากฏว่า มีนายโสภณ โล ผู้จัดการโครงการและสถาปนิกอาวุโส บริษัทบิลเดอร์คอนซัลแตนส์ จำกัด และบริษัท บิลเดอร์เรียลตี้ จำกัด เข้ามาภายในห้องแจ้งว่าเขาเป็นเจ้าของห้อง โดยนำหลักฐานต่างๆมายืนยัน ตนจึงเล่าเรื่องที่นายพีระศิลป์ลงโฆษณาให้เช่าห้อง และได้จ่ายเงินค่าเช่าเกือบแสนบาทให้ฟัง แต่นายโสภณไม่สนใจ อ้างว่าไม่มีส่วนทราบเรื่องที่เกิดขึ้นแต่อย่างใด จึงจำใจต้องเก็บข้าวของกลับออกมา และพยายามติดต่อนายพีระศิลป์ แต่ไม่สามารถติดต่อได้จึงเชื่อว่าถูกหลอก น.ส.ชโลมจิตเผยอีกว่า ได้ตรวจสอบกับห้องเช่าที่อยู่ติดกันทราบว่า มีชาวต่างชาติอีกรายที่หลงเชื่อจ่ายเงินให้นายพีระศิลป์ไปแล้วเช่นเดียวกัน พนักงานในคอนโดมิเนียมหลายคนบอกว่า เคยเห็นนายพีระศิลป์ เข้ามาที่

คอนโดมิเนียมบ่อยครั้ง แต่ไม่ทราบว่าเป็นเจ้าของห้องพักหรือไม่ เชื่อว่าน่าจะมีส่วนรู้เห็นกับนายพีระศิลป์ด้วย เนื่องจากกุญแจห้องพักและคีย์การ์ดเพื่อเข้าคอนโดมิเนียมที่ได้มาไม่น่าจะทำปลอมขึ้นมาได้ จึงไม่ยอมให้มีใครหลงเชื่ออีก เพราะยังมีการลงโฆษณาให้เช่าห้องพักในนิตยสารอยู่ เช่นเดิม ด้าน ร.ต.อ.สวิกกล่าวว่า เบื้องต้นตำรวจได้รับเรื่องไว้แล้ว ส่วนจะมีการพิจารณาส่งให้ สน. พุ่งมหาเมฆ เจ้าของห้องที่เกิดเหตุรับไปดำเนินการหรือไม่ นั้น จะรายงานให้ผู้บังคับบัญชาทราบ และพิจารณาสั่งการตามลำดับขั้นต่อไป

ผลการตัดคำ " / น่อง / เบนซ์ / จู / เนียร์ / " / แท็ก / ซี่ / เม้า / ชน / สนั่น / ทางด่วน / ลูกทุ่ง / เด็ก / ซื่อ / คัง / " / น่อง / เบนซ์ / จู / เนียร์ " / เจ้าของ / เสียง / เพลง / " / สู้ / ไว้ย / " / ประสบ / อุบัติเหตุ / รถ / ตู้ / ชน / กับ / แท็ก / ซี่ / บน / ทางด่วน / โชค / ดี / ไม่ / เป็น / อะไร / มาก / แต่ / ต้อง / เข้า / โรง / พยาบาล / รอ / ให้ / หมอ / ตรวจ / เช็ก / อาการ / ทาง / สมอ / อิก / ที / เผย / ก่อน / เกิด / เหตุ / นอน / หลับ / มา / ใน / รถ / ตู้ / มี / พ่อ / เป็น / คน / ขับ / กำลัง / จะ / ไป / เข้า / ห้อง / อัคคีภัย / ทำ / อัลบั้ม / เพลง / ชุด / ใหม่ / เกราะ / ร้าย / เจอ / แท็ก / ซี่ / ซี่ / เม้า / ขับ / รถ / มา / ชน / ตำรวจ / พา / ตัว / โชเฟอร์ / ไป / สน . / ทางด่วน / 2 / แจ้ง / ขั้ / หา / ขับ / รถ / ประมาท / กับ / เม้า / แล้ว / ขับ / แม่ / นักร้อง / เด็ก / ยืน ย้น / ให้ / ดำเนินคดี / ถึง / ที่ / สุด / เพื่อ / ไม่ / ให้ / เป็น / เยี่ยง / อย่าง / ด้าน / " / น่อง / มา / ยต์ " / เด็ก / ดอย / ใจดี / สู้ / ข่า / วั / เดิน / ทาง / มา / เยี่ยม / ให้ / กำลัง / ใจ / ถึง / ร.พ. . / เหตุการณ์ / ลูกทุ่ง / เด็ก / " / น่อง / เบนซ์ / จู / เนียร์ " / ประสบ / อุบัติเหตุ / รถ / ตู้ / เฉี่ยว / ชน / กับ / แท็ก / ซี่ / ครั้ง / นี้ / เมื่อ / เวลา / 02.30 / น. / วัน / ที่ / 20 / พ.ย. . / พ.ต.ท. / รัชฎา / ญา / ริน / ทรัพย์ / บริรักษ์ / สารวัตร / เวร / สน . / ทางด่วน / 2 / รับ / แจ้ง / เหตุ / รถ / เฉี่ยว / ชน / กัน / ที่ / บริเวณ / ทางด่วน / 2 / ขา / เข้า / กรุงเทพ / ๑ / ช่วง / บริเวณ / สถานี / ขนส่ง / สาย / เหนือ / (/ หมอ / ซิต / ใหม่ /) / แขวง / จอมพล / เขต / จตุ / จักร / รุด / ไป / ตรวจ / สอบ / พบ / รถ / ตู้ / โต / โย / ต้า / ไฮ / เอ / ช / ลี / บรอนซ์ / ไฟ / ทะเบียน / สก-8203 / กรุงเทพมหานคร / ใน / สภาพ / ถูก / ชน / ด้าน / ขวา / ยวบ / เป็น / รอย / ยาว / จอด / อยู่ / ซิด / ขอบ / ของ / ทางด่วน / ใกล้ / กัน / พบ / รถ / แท็ก / ซี่ / นำเงิน / - / แดง / ยี่ / ห้อ / โต / โย / ต้า / ลี / โม / ทะเบียน / ทพ-8144 / กรุงเทพมหานคร / จอด / ซิด / เล่น / ขวา / กัน / ชน / หน้า / มี / ร่องรอย / การ / ชน / ด้าน / ซ้าย / ยวบ / ทั้ง / แถบ / ล้อ / ด้าน / ขวา / ยาง / แตก / ล้อ / แม็กซ์ / คด / สอบสวน / ทราบ / ว่า / รถ / ตู้ / โต / โย / ต้า / คัน / คัง / คัง / มี / นาย / สห / วัศ / กฤษณ / โสภ / อายุ / 50 / ปี / เป็น / คน / ขับ / มี / นาง / พร / ทิพย์ / กฤษณ / โสภ / อายุ / 44 / ปี / ภรรยา / และ / ด.ญ. / สุข / จิ / รา / กฤษณ / โสภ / หรือ / น่อง / เบนซ์ / จู / เนียร์ / บุตร / สาว / อายุ / 8 /

ขวบ / นัก / ร้อง / ลูกทุ่ง / เด็ก / ชื่อ / ดั่ง / เจ้าของ / อัลบั้ม / เพลง / ดัง / " / ผู้ / โวย / " / นั่ง / มา
 / ใน / รถ / ด้วย / ส่วน / รถ / แท็ก / ซี / มี / นาย / อภิรักษ์ / บุญ / ปก / อายุ / 31 / ปี / อยู่ /
 บ้าน / เลข / ที่ / 141/1 / ม.1 / ต. / สามขา / อ. / โพน / ทราย / จ. / ร้อยเอ็ด / เป็น / คน /
 ขับ / นาย / สห / วัส / กล่าว / ว่า / ก่อน / เกิด / เหตุ / คน / ขับ / รถ / พา / น้อง / เบนซ์ / ไป /
 เรียน / เท / ลวัน / โด / ที่ / เมือง / ทอง / ธาณี / จาก / นั้น / ได้ / แวะ / ทาน / อาหาร / เพื่อ /
 เดินทาง / ไป / ห้อง / อัดเสียง / ของ / บริษัท / ที่ / อป / ไลน์ / ไค / มอ / นด์ / ย่าน / พระ / โขนง /
 ระหว่าง / ขับ / รถ / มา / ตาม / ทางด่วน / แจ้ง / วัฒนะ / เพื่อ / ไป / ลง / ที่ / อจ / ณรงค์ / เมื่อ /
 มา / ถึง / ที่ / เกิด / เหตุ / เป็น / บริเวณ / ทาง / โค้ง / หน้า / หมอ / ชิต / ใหม่ / คน / ขับ / มา / เล่น /
 กลาง / แต่ / จู่ ๆ / ก็ / มี / รถ / แท็ก / ซี / ขับ / เข้า / มา / ชน / อย่าง / รุนแรง / จน / รถ / ของ / คน
 / เสียหลัก / โคลง / ไป / โคลง / มา / น้อง / เบนซ์ / ซึ่ง / นอน / หลับ / อยู่ / บน / เตียง / ด้าน / หลัง
 / รถ / หัว / กระแทก / กับ / พนัก / พิง / อย่าง / แรง / หลัง / เหตุการณ์ / คน / ได้ / ลง / ไป / ตก
 ลง / กับ / คน / ขับ / รถ / แท็ก / ซี / ซึ่ง / อยู่ / ใน / อากาศ / มึนเมา / และ / พยายาม / บ่ายเบี่ยง /
 ไม่ / รับผิดชอบ / คน / จึง / แจ้ง / เจ้าหน้าที่ / ดำรวจ / ให้ / มา / ดำเนิน / การ / ด้าน / นาง / พร /
 ทิพย์ / กล่าว / ว่า / หลังจาก / ที่ / นำ / ตัว / คู่กรณี / มา / ที่ / สน . / ทางด่วน / 2 / คน / และ /
 พ่อ / ของ / น้อง / เบนซ์ / พยายาม / ที่ / จะ / ตกลง / กับ / คน / ขับ / รถ / แท็ก / ซี / แต่ / ไม่ / เป็น
 / ผล / แถม / ยัง / แสดง / ท่าทาง / ก้าวร้าว / ไม่ / ยอม / รับผิด / และ / ไม่ / ยอม / ให้ / ดำรวจ /
 ตรวจ / วัด / แอลกอฮอล์ / ทั้ง / ที่ / เป็น / ฝ่าย / ขับ / ชน / รถ / คน / เอง / ทำให้ / น้อง / เบนซ์ / ซึ่ง
 / นอน / อยู่ / เตียง / ด้าน / หลัง / กลิ้ง / ลง / มา / จน / หัว / กระแทก / พนัก / พิง / ซึ่ง / คง / ต้อง /
 นำ / ไป / เอกซเรย์ / ที่ / โรง / พยาบาล / เพื่อ / ตรวจ / อาการ / ว่า / รุนแรง / หรือ / ไม่ / “ / รถ /
 แท็ก / ซี / ขับ / มา / ชน / ข้าง / ที่ / น้อง / เบนซ์ / นอน / อยู่ / พอ / ดี / ทำให้ / ลูก / สาว / หัว /
 กระแทก / พนัก / พิง / จน / สะคุ้ง / ตื่น / ลูก / ขึ้น / มา / บอก / ว่า / เจ็บ / ที่ / หัว / ด้าน / ซ้าย /
 ปวด / ตา / ซ้าย / ด้วย / และ / มี / หัว / จาก / นั้น / ก็ / ร้อง / ให้ / ทำให้ / รู้ / สึก / สงสาร / ลูก /
 และ / คง / ต้อง / ดำเนินคดี / กับ / คู่กรณี / ให้ / ถึง / ที่ / สุด / ให้ / เป็น / ตัวอย่าง / สำหรับ / ผู้ / ที่ /
 เมา / แล้ว / ขับ / เพราะ / เมา / แล้ว / ขับ / เป็น / อันตราย / กับ / ทุก / คน / ไม่ / ใช่ว่า / เฉพาะ /
 น้อง / เบนซ์ / เท่านั้น / ” / นาง / พร / ทิพย์ / กล่าว / ด้าน / น้อง / เบนซ์ / กล่าว / ด้วย / น้ำเสียง
 / อ่อนเปลี้ย / ว่า / น้อง / เบนซ์ / รู้ / สึก / ว่า / มี / เสียง / ดัง / มา / กระแทก / กับ / รถ / จน / ตก / จาก
 / เตียง / คิด / ว่า / รถ / ต้อง / มา / ชน / แน่ / แต่ / ไม่ / คิด / ว่า / จะ / ชน / ด้าน / ที่ / เบนซ์ /
 นอน / อยู่ / หลังจาก / ที่ / ตื่น / แล้ว / รู้ / สึก / ปวด / และ / เจ็บ / ที่ / หัว / ด้าน / ซ้าย / ผู้ / สื่อ / ข่าว /
 รายงาน / ว่า / นาย / อภิรักษ์ / โขเฟอร์ / แท็ก / ซี / ซึ่ง / ถูก / คুম / ตัว / มา / ที่ / สน . / ทางด่วน /
 2 / ได้ / พยายาม / คืบ / น้ำ / และ / อม / ฮอลด์ / เพื่อ / กลบ / เคลื่อน / กลิ่น / สุรา / และ / เมื่อ /

เจ้าหน้าที่ / ดำรวจ / นำ / เครื่อง / ตรวจ / วัด / แอลกอฮอล์ / มา / ตรวจ / วัด / นาย / อภิรักษ์ / กี่ /
 พยายาม / อ้าง / ว่า / ป่วย / เป็น / หวัด / เจ้าหน้าที่ / ดำรวจ / จึง / ระบุ / ว่า / หาก / ไม่ / เป่า / จะ /
 ทำ / การ / เจาะ / เลือด / นาย / อภิรักษ์ / จึง / ยอม / เป่า / ผล / ออก / มา / พบ / ว่า / มี / ปริมาณ
 / แอลกอฮอล์ / 95 / มิลลิกรัม / เปอร์เซ็นต์ / ถือว่า / เกิน / กว่า / ที่ / กฎหมาย / กำหนด /
 เจ้าหน้าที่ / ดำรวจ / จึง / แจ้ง / ข้อหา / ขับ / รถ / โดย / ประมาท / เป็น / เหตุ / ให้ / ผู้ / อื่น / ได้ / รับ
 / บาดเจ็บ / และ / ข้อหา / ขับ / รถ / ใน / ขณะ / เมา / สุรา / พ.ต.ท. / ชาญ / ญา / ริน / ทรัพย์ / กล่าว /
 ว่า / สอบสวน / ผู้ / ต้องหา / ทราบ / ว่า / ขับ / รถ / มา / จาก / ประชา / ชน / จะ / ไป / หา / เพื่อน /
 โดย / จะ / ลง / ทาง / ค่วน / ที่ / พญา / ไท / แต่ / เมื่อ / มา / ถึง / ที่ / เกิด / เหตุ / ปราบ / ว่า / รถ / เสีย
 หลัก / ไป / ชน / กับ / ขอบ / ทาง / ค่วน / จน / ยาง / แตก / และ / เสีย / หลัก / ไป / ชน / กับ / รถ / ตู้ /
 ของ / น้อง / เบนซ์ / ซึ่ง / เจ้าหน้าที่ / จะ / ดำเนิน / คดี / ตาม / ข้อ / กล่าวหา / ต่อ / ไป / ต่อ / มา /
 เวลา / 17.30 / น. / ค.ญ. / บุญ / พิช / ริ / แสง / ทอง / วีร / กุล / หรือ / น้อง / มา / ยต์ /
 วัย / 10 / ขวบ / นัก / ร้อง / เด็ก / ชื่อ / ดั่ง / เจ้า / ของ / เพลง / เด็ก / คอย / ใจ / ดี / พร้อม / ด้วย
 / แม่ / และ / พ่อ / เดินทาง / มา / เยี่ยม / น้อง / เบนซ์ / จู / เนียร์ / ซึ่ง / เข้า / พัก / รักษา / ตัว /
 และ / รอ / ดู / อาการ / ที่ / ห้อง / 1022 / โรง / พยาบาล / น / นท / เวช / ย่าน / งาม / วงศ์ / วาน
 / โดย / น้อง / มา / ยต์ / ได้ / ตรง / เข้า / ไป / กอด / น้อง / เบนซ์ / พร้อม / สอบ / ถาม / ถึง /
 อาการ / บาดเจ็บ / ทั้ง / ยัง / ได้ / นำ / นม / สด / ที่ / เตรียม / มา / ให้ / น้อง / เบนซ์ / ด้วย / นาง /
 พร / ทิพย์ / มารดา / ของ / น้อง / เบนซ์ / กล่าว / ว่า / หลังจาก / เข้า / พัก / รักษา / ตัว / ที่ / ร.พ. /
 ตั้งแต่ / เมื่อ / คืบ / ตอน / นี้ / หลังจาก / ที่ / หมอ / เอกชเรย์ / แล้ว / ผล / ปราบ / ว่า / ไม่ / เป็น /
 อะไร / มาก / แต่ / มี / รอย / ข้ำ / เล็ก / น้อย / พอ / หมอ / ให้ / ยา / ก็ / พุด / คุย / ได้ / ปกติ / แต่
 / พอ / หมอ / ฤทธิ์ / ยา / ก็ / จะ / มี / อาการ / ปวด / หัว / หมอ / จึง / ให้ / นอน / ที่ / โรง /
 พยาบาล / ก่อน / เพื่อ / ตรวจ / เช็ก / อาการ / ทาง / สมอ / เพราะ / กลัว / กระทบ / กระ / เทียน / ถ้า /
 หาก / เข้า / วัน / ที่ / 21 / พ.ย. / อาการ / ปกติ / ก็ / จะ / เอกชเรย์ / อีก / ครั้ง / ถ้า / ไม่ / มี /
 อะไร / ก็ / จะ / กลับ / บ้าน / ได้ / เมื่อ / ถาม / ถึง / ผล / กระ / เทบ / ของ / งาน / นาง / พร / ทิพย์ /
 กล่าว / ว่า / ตอน / นี้ / ต้อง / ยก / เลิก / คิว / อัลดเซีย / ออก / ไป / ก่อน / ส่วน / งาน / โชว์ / ตัว / ก็
 / ต้อง / ยก / เลิก / ไป / บาง / งาน / คิด / ว่า / ถ้า / วัน / ที่ / 21 / พ.ย. / นี้ / ไม่ / มี / อะไร / ก็ /
 กลับ / บ้าน / และ / ทำ / งาน / ได้ / ปกติ / ส่วน / เรื่อง / คดี / จะ / ให้ / เจ้าหน้าที่ / ดำรวจ / ดำเนิน
 / การ / ไป / ตาม / กฎหมาย / เรา / ไม่ / ได้ / เรียก / ร้อง / ค่า / เสีย / หาย / แต่ / อยาก / จะ / ให้ /
 ดำเนิน / การ / เพราะ / ถ้า / รู้ / ว่า / เมา / ก็ / ไม่ / ควร / จะ / ขับ / แรก / ๆ / เขา / พยายาม / จ่าย /
 ค่า / เสีย / หาย / ให้ / แต่ / เรา / ก็ / ไม่ / รับ / อยาก / ให้ / เป็น / ข้อ / เตือน / ใจ / สำหรับ / การ / เมา /
 แล้ว / ขับ / ไม่ / อยาก / จะ / ให้ / คิด / ว่า / จ่าย / เงิน / แล้ว / เรื่อง / ก็ / จบ / ไป / แล้ว / คน / ขับ

/ ก็ / กลับ / ไป / ทำ / อีก / สำหรับ / น้อง / เบนซ์ / จู / เนียร์ / เป็น / นัก / ร้อง / สังกัด /ค่าย / ที่ /
 / อป / ไลน์ / ได / มอ / นด์ / มี / ผล / งาน / อัลบั้ม / ชุด / แรก / ชื่อ / " / สู้ / โวย / " / มี / เพลง /
 โด่งดัง / หลาย / เพลง / อาทิ / คุณ / ครู / ที่รัก / ดี / ค็อก / ขอ / เวลา / ลูก / บ้าง / ขณะ / นี้ /
 / กำลัง / อยู่ / ใน / ระหว่าง / กำลัง / อัค / เท / ป / ชุด / ที่ / สอง / คือ / ชุด / " / จักรยาน / โช / หย่อน /
 / " /

ผลการแทนที่คำหยาบด้วยเครื่องหมายดอกจัน "น้องเบนซ์ จูเนียร์" แท้ก็ชี้มา ชนสนั่นทางด่วน
 ลูกทุ่งเด็กชื่อดัง "น้องเบนซ์ จูเนียร์" เจ้าของเสียงเพลง "สู้*" ประสบอุบัติเหตุรถตู้ชนกับแท็กซี่บน
 ทางด่วน โชคดีไม่เป็นอะไรมาก แต่ต้องเข้าโรงพยาบาลรอให้หมอตรวจเช็กอาการทางสมองอีกที
 เผยก่อนเกิดเหตุนอนหลับมาในรถตู้ มีพ่อเป็นคนขับ กำลังจะไปเข้าห้องอัดเสียงทำอัลบั้มเพลงชุด
 ใหม่ เพราะห้*เจอแท็กซี่ชี้มาขับรถมาชน ตำรวจพาตัวไซเฟอร์ไปสน.ทางด่วน 2 แจ้งขอหาขับรถ
 ประมาท กับเมาแล้วขับ แม่ขับรถเด็กยืนยันให้ดำเนินคดีถึงที่สุดเพื่อไม่ให้เป็นเยี่ยงอย่าง ด้าน
 "น้องมายด์" เด็กลอยใจดีรู้ข่าวรีบเดินทางมาเยี่ยมให้กำลังใจถึงร.พ. เหตุการณ์ลูกทุ่งเด็ก "น้องเบนซ์
 จูเนียร์" ประสบอุบัติเหตุรถตู้เฉี่ยวชนกับแท็กซี่ครั้งนี้ เมื่อเวลา 02.30 น. วันที่ 20 พ.ย. พ.ต.ท.ธัญญา
 รินทร์ บริรักษ์ สารวัตรเวร สน.ทางด่วน 2 รับแจ้งเหตุรถเฉี่ยวชนกันที่บริเวณทางด่วน 2 ขาเข้า
 กรุงเทพฯ ช่วงบริเวณสถานีขนส่งสายเหนือ (หมอชิตใหม่) แขวงจอมพล เขตจตุจักร รุดไป
 ตรวจสอบพบรถตู้โตโยต้า ไฮเอช สีบรอนซ์ฟ้า ทะเบียน สก-8203 กรุงเทพมหานคร ในสภาพถูกชน
 ด้านขวายุบเป็นรอยยาว จอดอยู่ชิดขอบของทางด่วน ใกล้กันพบรถแท็กซี่น้ำเงิน-แดง ยี่ห้อโตโยต้า
 ลิโม ทะเบียน ทพ-8144 กรุงเทพมหานคร จอดชิดเลนขวา กันชนหน้ามีร่องรอยการชน ด้านซ้ายยุบ
 ทั้งแถบ ล้อด้านขวาแยกแตก ล้อแม็กซ์คด สอบสวนทราบว่ารถตู้โตโยต้าคันดังกล่าวมีนายสหัส
 ฤกษ์ โสภา อายุ 50 ปี เป็นคนขับ มีนางพรทิพย์ ฤกษ์ โสภา อายุ 44 ปี ภรรยาและด.ญ.สุกจิรา
 ฤกษ์ โสภา หรือน้องเบนซ์ จูเนียร์ บุตรสาว อายุ 8 ขวบ นักร้องลูกทุ่งเด็กชื่อดัง เจ้าของอัลบั้มเพลง
 ดัง "สู้*" นั่งมาในรถด้วย ส่วนรถแท็กซี่มีนายอภิรักษ์ บุญปก อายุ 31 ปี อยู่บ้านเลขที่ 1411 ม.1 ต.
 สามขา อ.โพธาราย จ.ร้อยเอ็ด เป็นคนขับ นายสหัสกล่าวว่ ก่อนเกิดเหตุคนขับรถพาน้องเบนซ์
 ไปเรียนเทควันโดที่เมืองทองธานี จากนั้นได้แวะทานอาหารเพื่อเดินทางไปห้องอัดเสียงของ
 บริษัทที่อป/ไลน์โดมอนด์ ย่านพระโขนง ระหว่างขับรถมาตามทางด่วนแจ้งวัฒนะเพื่อไปลงที่อจ
 ณรงค์ เมื่อมาถึงที่เกิดเหตุเป็นบริเวณทางโค้งหน้าหมอชิตใหม่ ตนขับมาเลนกลาง แต่จู่ๆก็มีรถ
 แท็กซี่ขับเข้ามาชนอย่างรุนแรงจนรถของตนเสียหลักโคลงไปโคลงมา น้องเบนซ์ซึ่งนอนหลับอยู่
 บนเตียงด้านหลังรถ หัวกระแทกกับพนักพิงอย่างแรง หลังเหตุการณ์ตนได้ลงไปตกลงกับคนขับรถ
 แท็กซี่ ซึ่งอยู่ในอาการมึนเมาและพยายามบ่ายเบี่ยงไม่รับผิดชอบ ตนจึงแจ้งเจ้าหน้าที่ตำรวจให้มา
 ดำเนินการ ด้านนางพรทิพย์กล่าวว่า หลังจากที่นำตัวคู่กรณีมาที่สน.ทางด่วน 2 ตนและพ่อของน้อง

เบนซ์พยายามที่จะตกลงกับคนขับรถแท็กซี่ แต่ไม่เป็นผล แลมยังแสดงท่าทางก้าวร้าว ไม่ยอมรับผิดชอบ และไม่ยอมให้ตำรวจตรวจวัดแอลกอฮอล์ ทั้งที่เป็นฝ่ายขับชนรถตนเอง ทำให้น้องเบนซ์ซึ่งนอนอยู่เตียงด้านหลังกลิ้งลงมาจนหัวกระแทกพนักพิง ซึ่งคงต้องนำไปเอกซเรย์ที่โรงพยาบาลเพื่อตรวจอาการว่ารุนแรงหรือไม่ “รถแท็กซี่ขับมาชนข้างที่น้องเบนซ์นอนอยู่พอดี ทำให้ลูกสาวหัวกระแทกพนักพิงจนสะดุ้งตื่น ลูกขึ้นมาบอกว่าเจ็บที่หัวด้านซ้าย ปวดตาซ้ายด้วยและมีเนื้องอก จากนั้นก็ร้องไห้ ทำให้รู้สึกสงสารลูกและคงต้องดำเนินคดีกับคู่กรณีให้ถึงที่สุดให้เป็นตัวอย่างสำหรับผู้ที่เมาแล้วขับ เพราะเมาแล้วขับเป็นอันตรายกับทุกคน ไม่ใช่เฉพาะน้องเบนซ์เท่านั้น” นางพรทิพย์ กล่าว ด้านน้องเบนซ์กล่าวด้วยน้ำเสียงอ่อนเพลียว่า น้องเบนซ์รู้สึกว่ามีเสียงดังมากกระแทกกับรถจนตกจากเตียง คิดว่ารถต้องมาชนแน่ แต่ไม่คิดว่าจะชนด้านที่เบนซ์นอนอยู่ หลังจากตื่นแล้วรู้สึกปวดและเจ็บที่หัวด้านซ้าย ผู้สื่อข่าวรายงานว่า นายอภิรักษ์ โชเฟอร์แท็กซี่ซึ่งถูกคุมตัวมาที่สน.ทางด่วน 2 ได้พยายามดื่มน้ำและอมสอกลีเพื่อคลายเกลือกลิ่นสุรา และเมื่อเจ้าหน้าที่ตำรวจนำเครื่องตรวจวัดแอลกอฮอล์มาตรวจวัด นายอภิรักษ์ก็พยายามอ้างว่าป่วยเป็นหวัด เจ้าหน้าที่ตำรวจจึงระบุว่าหากไม่เป่าจะทำการเจาะเลือด นายอภิรักษ์จึงยอมเป่า ผลออกมาพบว่าปริมาณแอลกอฮอล์ 95 มิลลิกรัมเปอร์เซ็นต์ ถือว่าเกินกว่าที่กฎหมายกำหนด เจ้าหน้าที่ตำรวจจึงแจ้งข้อหาขับรถโดยประมาทเป็นเหตุให้ผู้อื่นได้รับบาดเจ็บ และข้อหาขับรถในขณะที่เมาสุรา พ.ต.ท.ธัญญารินทร์กล่าวว่า สอบสวนผู้ต้องหาทราบว่าขับรถมาจากประชาชนจะไปหาเพื่อนโดยจะลงทางด่วนที่พญาไท แต่เมื่อมาถึงที่เกิดเหตุปรากฏว่ารถเสียหลักไปชนกับขอบทางด่วนจนยางแตก และเสียหลักไปชนกับรถตู้ของน้องเบนซ์ ซึ่งเจ้าหน้าที่จะดำเนินคดีตามข้อกล่าวหาต่อไป ต่อมาเวลา 17.30 น. ด.ญ.บุญย์พัชรี แสงทองวีรกุล หรือน้องมายด์ วัย 10 ขวบ นักร้องเด็กชื่อดังเจ้าของเพลง เด็กคอยใจดี พร้อมด้วยแม่และพ่อ เดินทางมาเยี่ยม น้องเบนซ์ ภูเนียร์ ซึ่งเข้าพักรักษาตัวและรอดูอาการ ที่ห้อง 1022 โรงพยาบาลนนทเวช ย่านงามวงศ์วาน โดยน้องมายด์ได้ตรงเข้าไปกอดน้องเบนซ์ พร้อมสอบถามถึงอาการบาดเจ็บ ทั้งยังได้นำนมสดที่เตรียมมาให้น้องเบนซ์ด้วย นางพรทิพย์ มารดาของน้องเบนซ์กล่าวว่า หลังจากเข้าพักรักษาตัวที่ร.พ.ตั้งแต่วันที่ 21 พ.ย. ตอนนี้หลังจากที่หมอเอกซเรย์แล้วผลปรากฏว่าไม่เป็นอะไรมาก แต่มีรอยช้ำเล็กน้อย พอหมอให้ยาาก็พุดคุยได้ปกติ แต่พอหมดฤทธิ์ยา ก็จะมีอาการปวดหัว หมอจึงให้นอนที่โรงพยาบาลก่อนเพื่อตรวจเช็คอาการทางสมองเพราะกลัวกระทบกระเทือน ถ้าหากเข้าวันที่ 21 พ.ย. อาการปกติก็จะเอกซเรย์อีกครั้ง ถ้าไม่มีอะไรก็จะกลับบ้านได้ เมื่อถามถึงผลกระทบของงาน นางพรทิพย์กล่าวว่า ตอนนี้ต้องยกเลิกคิวอัดเสียงออกไปก่อน ส่วนงานโชว์ตัวก็ต้องยกเลิกไปบางส่วน คิดว่าถ้าวันที่ 21 พ.ย.นี้ไม่มีอะไรก็กลับบ้านและทำงานได้ปกติ ส่วนเรื่องคดีจะให้เจ้าหน้าที่ตำรวจดำเนินการไปตามกฎหมาย เราไม่ได้เรียกร้องค่าเสียหาย แต่อยากจะให้ดำเนินการเพราะถ้ารู้ว่าเมาก็ไม่ควรที่จะขับ แรกๆ เขาพยายามจ่ายค่าเสียหายให้ แต่เราก็กไม่รับ อยากให้เป็นข้อเตือนใจสำหรับการ

เมาแล้วขับ ไม่อยากจะให้คิดว่าจ่ายเงินแล้วเรื่องก็จบไป แล้วคนขับก็กลับไปทำอีก สำหรับน้องเบนซ์ จูเนียร์ เป็นนักร้องสังกัดค่ายท็อปไลน์ ไดมอนด์ มีผลงานอัลบั้มชุดแรกชื่อ"สู้*" มีเพลงโด่งดังหลายเพลง อาทิ คุณครูที่รัก ดึกตอก ขอเวลาลูกบ้าง ขณะนี้กำลังอยู่ในระหว่างกำลังอัดเทปชุดที่สองคือชุด"จักรยานโซ่หย่อน"

ประวัติผู้เขียน

ชื่อ นามสกุล

นายวุฒิชัย วิเชียรไชย

ประวัติการศึกษา

วิทยาศาสตรบัณฑิต (ระบบสารสนเทศเพื่อการจัดการ)
มหาวิทยาลัยมหาสารคาม ปีการศึกษา 2545