

# การสรุปเอกสารเชิงความหมายโดยใช้ออนโทโลยี

## Semantic Text Summarization Using Ontology

อรรณพ อูไรเรืองพันธุ์<sup>1</sup>สมจิตร อาจอินทร์<sup>2</sup>

ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น 40002

อีเมล : deknai\_kung@hotmail.com<sup>1</sup>, somjit@kku.ac.th<sup>2</sup>

### บทคัดย่อ

ข่าวสารสารสนเทศในรูปแบบอิเล็กทรอนิกส์มีเพิ่มมากขึ้นเรื่อยๆ ในปัจจุบัน ซึ่งทำให้ต้องเสียเวลาในการค้นหาคัดเลือกข้อมูลให้ตรงกับความต้องการของผู้ใช้ จึงได้มีงานวิจัยและแนวคิดเกี่ยวกับการสรุปใจความอัตโนมัติ (Text Summarization) มาช่วยในการแก้ปัญหาการคัดเลือกข้อมูล แต่เป็นการอ้างอิงคำหลัก (Keyword-Base) ซึ่งผลลัพธ์ที่ได้ยังคงมีความถูกต้องและประสิทธิภาพไม่เพียงพอ งานวิจัยนี้จึงได้นำเทคโนโลยีด้านความคล้ายคลึงกันเชิงความหมาย (Semantic Similarity) การพัฒนาฐานข้อมูลคำศัพท์เวิร์ดเน็ต (WordNet) การจัดหมวดหมู่ของกลุ่มคำเหมือนโดยใช้หลักการ Lexical Chain มาประยุกต์ใช้เพื่อเพิ่มประสิทธิภาพและความถูกต้องมากขึ้น และผลการทดลองจากงานวิจัยนี้ได้ผลลัพธ์ที่ค่อนข้างมีประสิทธิภาพ โดยวัดจากค่าความถูกต้อง (Precision) ค่าความครบถ้วน (Recall) และค่า F-measure ซึ่งมีค่าเท่ากับ 91.61%, 81.92% และ 84.90% ตามลำดับ

**คำสำคัญ:** การสรุปใจความอัตโนมัติ ฐานข้อมูลคำศัพท์เวิร์ดเน็ต การจัดหมวดหมู่คำหลัก ความคล้ายคลึงกันเชิงความหมาย

### Abstract

In the present, as the amount of electronic information is getting more and more, it consumes a lot of time in searching and selecting the needed information. Therefore, the researches and the concepts related to Text Summarization were studied to solve this

problem. However, it was just a Keyword-Base which is not effective enough. For this reason, this research used Semantic Similarity technology, WordNet, categorizing the similar meaning words by using Lexical Chain in order to increase effectiveness and correctness. The results of the study were shown that the program is effective regarding on the values of Precision, Recall and F-measure which were 91.61%, 81.92% and 84.90% respectively.

**Keyword:** Text Summarization, WordNet, Lexical Chain, Semantic Similarity

### 1. บทนำ

ปัจจุบันเทคโนโลยีสารสนเทศได้มีอิทธิพลต่อการดำเนินชีวิตเราเป็นอย่างมาก ซึ่งส่งผลให้มีข้อมูลข่าวสารเอกสารอิเล็กทรอนิกส์เพิ่มมากขึ้นเรื่อยๆ ทำให้สร้างปัญหาในการพิจารณาคัดเลือกข้อมูลสารสนเทศซึ่งบางครั้งอาจได้ข้อมูลที่ไม่ตรงกับความต้องการของผู้ใช้ และเมื่อข้อมูลยิ่งมากยิ่งขึ้นยิ่งประสบปัญหาในการใช้ข้อมูล เพราะเอกสารส่วนใหญ่อยู่กระจัดกระจายไม่เป็นหมวดหมู่ ทำให้ผู้ใช้งานต้องเสียเวลาในการคัดเลือกข้อมูลเพื่อให้ได้ตามต้องการ นับตั้งแต่มีงานวิจัยที่เสนอแนวคิดการสรุปใจความอัตโนมัติ (Automatic Summarization) [1] ทำให้ช่วยลดเวลาในการค้นหาและคัดเลือกข้อมูลที่ต้องการได้ แต่ก็ยังพบปัญหาเนื่องจากการสรุปใจความอัตโนมัติในยุคแรกๆ นั้นเป็นการใช้คำหลักในฐานข้อมูล (Keyword-Base) เท่านั้น ทำให้การพิจารณาถึงความหมายของเนื้อหา (Contents) ที่ได้มีประสิทธิภาพไม่เพียงพอ

ต่อมาได้มีการศึกษาวิจัยเกี่ยวกับเทคโนโลยีการวัดค่าความคล้ายคลึงเชิงความหมาย (Semantic Similarity) [2] และมีการพัฒนาฐานข้อมูลอภิธานศัพท์หรือเวิร์ดเน็ต (WordNet) [3], [4] ซึ่งมีการอธิบายโครงสร้างและความสัมพันธ์ระหว่างกลุ่มคำคล้ายคลึงกันไว้ด้วยกัน จึงมีการคิดค้นวิจัยด้านเทคนิคการสรุปใจความอัตโนมัติกันอย่างกว้างขวางเพื่อให้ได้ประสิทธิภาพและความถูกต้องมากที่สุด

งานวิจัยนี้มุ่งเน้นเพื่อการออกแบบแนวคิดวิธีการสรุปใจความของเอกสารที่มีประสิทธิภาพและถูกต้อง โดยมีการใช้เทคโนโลยีทางด้านเชิงความหมาย (Semantic) เข้ามาช่วยในการคัดเลือกสกัดคำหลัก โดยอาศัยฐานข้อมูลคำศัพท์เวิร์ดเน็ตในการพิจารณาด้วย แล้วทำการจัดหมวดหมู่ของคำโดยใช้เทคนิคการสร้าง Lexical Chain มาวิเคราะห์หาความสัมพันธ์ของคำหลักที่สกัดจากเอกสาร และคัดเลือกผลลัพธ์จากลำดับความสำคัญของกลุ่มคำโดยใช้ค่าถ่วงน้ำหนักเชิงความหมายประกอบ

## 2. วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

### 2.1 การสรุปใจความสำคัญ (Text Summarization)

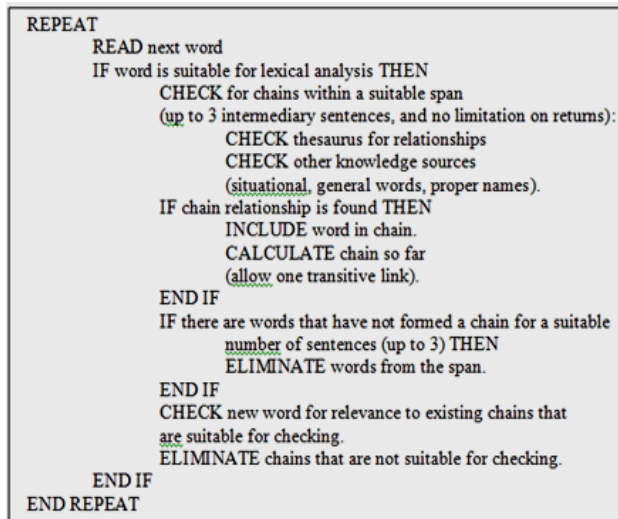
การสรุปใจความสำคัญ [1] เป็นเทคนิคการสรุปเนื้อหาของเอกสารโดยมีขั้นตอน 3 กระบวนการ คือ ขั้นตอนการวิเคราะห์คำ ขั้นตอนการหาประเด็นสำคัญ และขั้นตอนการสังเคราะห์ประโยค โดยสามารถแบ่งวิธีการสรุปใจความสำคัญตามวิธีการสร้างได้ 2 วิธี คือ วิธีการดึงคำจากต้นฉบับ (Extraction) ซึ่งเป็นการสร้างใจความสำคัญโดยใช้เทคนิคทางด้านสถิติ และ Keyword-Based ตัดคำจากเอกสารต้นฉบับ ผลที่ได้ยังมีประสิทธิภาพและความถูกต้องไม่เพียงพอ มีความกำกวมของคำแต่สามารถเพิ่มปริมาณ Domain ได้ [5] และอีกวิธีคือวิธีการสรุปแบบภาพรวม (Abstraction) ซึ่งเป็นการวิเคราะห์คำจากต้นฉบับด้วยหลักการทางภาษาศาสตร์ (Semantic) โดยเป็นการถอดความหรือแปลความหมายจากต้นฉบับ ผลที่ได้มีประสิทธิภาพและถูกต้องมากกว่าวิธีการ Abstraction แต่มีการจำกัดปริมาณ Domain งานวิจัยนี้ได้นำแนวคิดวิธีการ Abstraction มาประยุกต์ใช้เพื่อลดปัญหาความกำกวมของคำและให้ได้ผลลัพธ์ในเชิงความหมาย มีประสิทธิภาพและถูกต้องมากที่สุด

### 2.2 ฐานข้อมูลคำศัพท์เวิร์ดเน็ต (WordNet)

WordNet [4], [3] เป็นฐานข้อมูลคำศัพท์ออนไลน์ภาษาอังกฤษที่ออกแบบตามทฤษฎีทางภาษาศาสตร์ ซึ่งถูกพัฒนาโดย Princeton University ตั้งแต่ปี ค.ศ. 1985 โดยมีวัตถุประสงค์เพื่อผสมผสานระหว่างพจนานุกรมคำศัพท์ (Dictionary) กับพจนานุกรมอภิธานศัพท์หรือคำพ้องความหมาย/คำตรงข้าม (Thesaurus) เข้าด้วยกัน การเก็บข้อมูลคำศัพท์จะแบ่งตามชนิดของคำ เช่น คำนาม คำกริยา คำคุณศัพท์ คำวิเศษณ์ เป็นต้น และจับกลุ่มคำศัพท์ที่มีลักษณะความหมายคล้ายคลึงกันเข้าด้วยกัน เรียกว่า Synset [6] ซึ่งแต่ละกลุ่ม Synset จะแทนค่าของ 1 แนวคิด (Sense) และ Synset จะเชื่อมโยงไปยัง Synset อื่นๆ ด้วยโครงสร้างแบบลำดับชั้น (Hierarchical Structure) โดยมีความสัมพันธ์แบบ Synonym, Antonym, Hypernym, Hyponym, Meronym, Holonym, Troponym, Entailment, Coordinated เป็นต้น งานวิจัยนี้ได้ใช้ WordNet เป็นฐานข้อมูล ในกระบวนการดำเนินการต่างๆ

### 2.3 การจัดหมวดหมู่ของคำ (Lexical Chain)

เป็นการระบุความสอดคล้องของคำศัพท์ในเชิงความหมาย โดยจัดกลุ่มคำศัพท์ที่มีความหมายคล้ายคลึงกันไว้ด้วยกัน [7] โดยมีการอาศัยฐานข้อมูลที่ได้นิยามความสัมพันธ์ของคำศัพท์ไว้แล้ว เช่น WordNet เป็นต้น การจัดหมวดหมู่ของกลุ่มคำแบ่งออกเป็น 2 ขั้นตอน ได้แก่ การหาคำสำคัญ (Candidate Word) และการค้นหากลุ่มคำหลัก (Lexical Chain) โดยทำการจัดหมวดหมู่กลุ่มคำที่มีความหมายใกล้เคียงกันไว้ด้วยกัน เป็นการพิจารณาจากความสัมพันธ์ระหว่างกลุ่มคำ ที่ถูกระบุเป็นแบบ Extra-Strong, Strong หรือ Medium-Strong ตามลำดับการสร้าง Lexical Chain



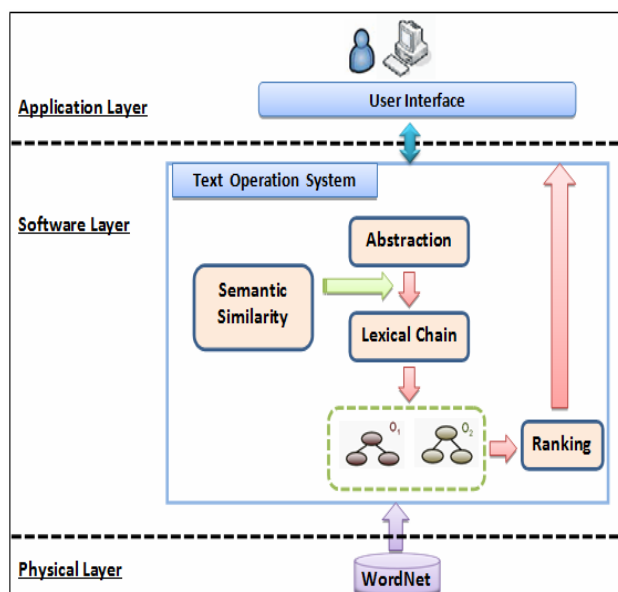
ภาพที่ 1 ขั้นตอนการสร้าง Lexical Chain [8]

### 3. วิธีการดำเนินงาน

การวิจัยนี้มีเป้าหมายเพื่อออกแบบและพัฒนาระบบการสรุปเอกสารอัตโนมัติโดยนำหลักการเชิงความหมายมาประยุกต์ใช้เพื่อให้มีประสิทธิภาพและความถูกต้อง โดยมีวิธีการดำเนินงาน แบ่งเป็น 2 ส่วน ได้แก่ ส่วนที่ 1 การออกแบบสถาปัตยกรรมระบบ และส่วนที่ 2 การพัฒนาฐานความรู้

#### 3.1 การออกแบบสถาปัตยกรรมระบบ

จากการศึกษาวรรณกรรมและงานวิจัยที่เกี่ยวข้อง และเพื่อตอบสนองต่อปัญหาและวัตถุประสงค์ของการวิจัย จึงได้ออกแบบสถาปัตยกรรมระบบ ดังนี้



ภาพที่ 2 สถาปัตยกรรมระบบ

จากสถาปัตยกรรมระบบ สามารถแบ่งระบบการทำงานออกเป็น 3 ระดับ ดังนี้

1. Application Layer เป็นส่วนของ User Interface ที่ทำหน้าที่ติดต่อกับผู้ใช้งานระบบ
2. Software Layer เป็นส่วนของการทำงานระบบ ซึ่งประกอบด้วยกระบวนการต่างๆ อาทิเช่น การสกัดคำหลัก การวัดค่าความคล้ายคลึงกันเชิงความหมาย การจัดหมวดหมู่ของคำ การคำนวณค่าน้ำหนักของกลุ่มคำ เป็นต้น
3. Physical Layer เป็นส่วนของฐานข้อมูลระบบ ในงานวิจัยนี้จะใช้ฐานข้อมูลคำศัพท์เวิร์ดเน็ตอ้างอิงในกระบวนการต่างๆ

#### 3.2 การพัฒนาฐานความรู้

เป็นการพัฒนาในส่วนของการทำงานระบบ มีองค์ประกอบหลัก 2 ส่วน ซึ่งประกอบด้วย **การสกัดความรู้** และ **การจัดกลุ่มความรู้** โดยการสร้าง Lexical Chain

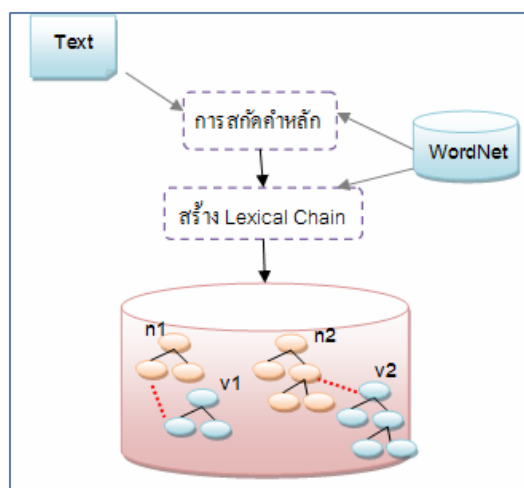
##### 3.2.1 การสกัดความรู้ (การตัดคำและสกัดคำหลัก)

การวิจัยนี้เป็นการออกแบบเพื่อใช้กับเอกสารหรือบทความภาษาอังกฤษที่เป็นไฟล์ \*.txt หรือการ Copy & Paste เท่านั้น ระบบจะทำการแยกคำและประโยคโดยใช้เครื่องหมายต่างๆ อาทิเช่น ., ?, ! เป็นต้น การตัดคำที่เป็นกริยาช่วยทั้งรูปคำเต็ม รูปคำย่อและรูปคำปฏิเสธ การตัดคำนำหน้าคำนาม (Article) และจะทำการคัดเลือกคำเฉพาะคำนาม และคำกริยาที่สื่อความหมายโดยอาศัยฐานข้อมูลคำศัพท์เวิร์ดเน็ต

ขั้นตอนการสกัดคำ จะระบุค่าความถี่ของคำหลัก (tf) และค่าน้ำหนักของคำ (W) ไว้ด้วย ซึ่งค่า W จะหาได้จากค่า  $tf/tf_{max}$  ของคำหลัก

##### 3.2.2 การจัดกลุ่มความรู้โดยการสร้าง Lexical Chain

เป็นการนำคำหลักที่สกัดได้มาจัดกลุ่มหรือสร้าง Chain ของคำศัพท์โดยอาศัยความสัมพันธ์เชิงความหมาย [7] โดยการวิจัยนี้จะอาศัยฐานข้อมูลคำศัพท์เวิร์ดเน็ตซึ่งได้นิยามความสัมพันธ์ของคำศัพท์ไว้แล้ว



ภาพที่ 3 แนวคิดการสร้าง Lexical Chain

งานวิจัยนี้ได้ออกแบบอัลกอริทึมการสร้าง Lexical Chain โดยประยุกต์จากแนวคิดของ ศรีสุตา [9]

1. กำหนดให้เซตของคำที่จะสร้าง chain เป็น  $n_1 \dots n_n$  และ  $v_1 \dots v_n$
2. เริ่มต้นด้วยการเลือกคำหลักค่านามที่ 1 ( $n_1$ ) และคำหลักคำกริยาที่ 1 ( $v_1$ ) กำหนดให้เป็นส่วนหัวเริ่มต้น chain ที่ 1 ( $Cn_1$ ) และ ( $Cv_1$ ) ของคำหลักค่านามและคำหลักคำกริยาตามลำดับ
3. เลือกคำที่ 2 ( $n_2$ ) และ ( $v_2$ )

ค้นหา chain ตั้งแต่  $Cn_1 \dots Cn_n$  และ  $Cv_1 \dots Cv_n$  โดยพิจารณาความสัมพันธ์ในลักษณะต่างๆ ที่ปรากฏในฐานความรู้เวิร์ดเน็ต ที่ทำให้  $n_2$  และ  $v_2$  มีความสัมพันธ์กับคำที่อยู่ใน chain แบบใดแบบหนึ่ง ดังต่อไปนี้

- 1) Extra-strong (Repetition)
- 2) Strong (Synonym, kindOf, Is-A, hasPart, part of, oppositeOf)
- 3) Medium-strong (ความสัมพันธ์แบบถ่ายทอด)

ถ้าพบความสัมพันธ์แบบใดแบบหนึ่ง จะทำการเพิ่ม  $n_2$  และ  $v_2$  เข้าไปใน chain ที่พบความสัมพันธ์แบบเหนียวแน่นที่สุด (mostly strong) หรือถ้าพบความสัมพันธ์ที่มีความเหนียวแน่นเท่ากัน ให้เพิ่มลงใน chain ที่ update ล่าสุด

ถ้าไม่พบทำการสร้างให้  $n_2$  และ  $v_2$  เป็นส่วนหัวของ chain ใหม่

4. ทำซ้ำขั้นตอนที่ 3 จนถึงคำ  $n_n$  และ  $v_n$

ภาพที่ 4 อัลกอริทึมการสร้าง Lexical Chain

ซึ่งมีการระบุค่าน้ำหนักความสัมพันธ์ (R) ทั้ง 3 แบบ ดังนี้

ตารางที่ 1 ค่าน้ำหนักความสัมพันธ์

ประเภทความสัมพันธ์		น้ำหนัก
Extra-strong	Repetition	1.0
Strong	Synonym	0.9
	Hypernym, Hyponym, Meronym, Holonym, troponym (Level 1)	0.6
Medium-strength	Hypernym, Hyponym, Meronym, Holonym (Level 2-5)	0.2
	Antonym, coordinate terms, entailment	0.1

หลังจากการจัดกลุ่มความรู้ ระบบจะทำการคำนวณค่าน้ำหนักของ Lexical Chain เพื่อเป็นการให้คะแนนแต่ละ Chain ว่ามีความสัมพันธ์กันของสมาชิกมากน้อยเพียงใด เพื่อหา Chain ที่จะป็นผลลัพธ์ของระบบ โดยคำนวณได้สมการ  $S_k$  ที่ได้รับปรั้งจาก Morris and Hirst [8] ดังสมการ 1

$$S_k = \left( \sum_{m=1}^{P_k+q} tf_{mk} R_{mk} W_{mk} \right) \times H_k \quad (1)$$

เมื่อ  $S_k$  คือ น้ำหนักของ Lexical Chain ที่ k  
 $tf_{mk}$  คือ ความถี่การเกิดคำหลักที่ m ที่เกิดซ้ำใน Lexical Chain ที่ k  
 $R_{mk}$  คือ น้ำหนักของความสัมพันธ์ของคำที่ m ใน Lexical Chain ที่ k  
 $W_{mk}$  คือ น้ำหนักของคำหลักที่ m ที่เกิดซ้ำใน Lexical Chain ที่ k  
 $H_k$  คือ ค่า Homogeneity index ซึ่งคำนวณได้จาก

$$H_k = 1 - \frac{P_k}{L_k} \quad (2)$$

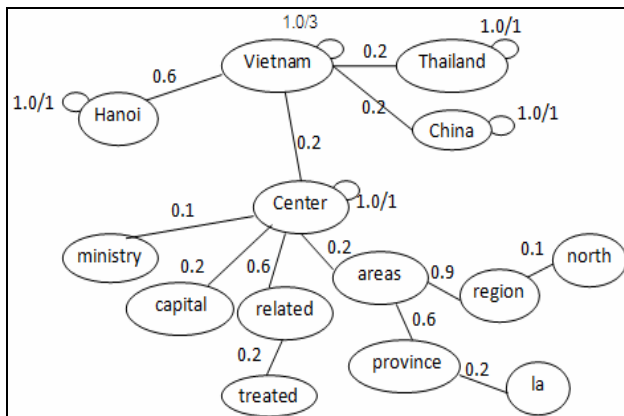
$P_k$  คือ จำนวนสมาชิกที่เกิดแตกต่างกันใน Lexical Chain ที่ k  
 $R_k$  คือ จำนวนคำหลักที่เป็นสมาชิกทั้งหมดใน Lexical Chain ที่ k

จากการคำนวณน้ำหนักของ Lexical Chain ที่ได้จากสมการข้างต้น จะทำการเรียงลำดับ Lexical Chain โดยพิจารณาจากค่าน้ำหนักมากที่สุดไปหาน้อยที่สุด

## 4. ผลการวิจัย

### 4.1 ผลการทดลอง

งานวิจัยนี้ ทำการทดลองระบบ โดยใช้ตัวอย่างที่บทความข่าวภาษาอังกฤษจากเว็บไซต์ [www.reuter.com](http://www.reuter.com) และ [www.bangkokpost.com](http://www.bangkokpost.com) ซึ่งแปลงเป็นไฟล์ .txt หรือใช้ในลักษณะการ Copy & Paste เพื่อนำเข้าข้อมูลสู่ระบบกระบวนการดำเนินงานจะเป็นไปตามวิธีการดำเนินการตามหัวข้อที่กล่าวข้างต้น ได้แก่ การสกัดคำหลัก โดยอาศัยฐานข้อมูลคำศัพท์เวิร์ดเน็ต และมีการกำหนดค่าความถี่ของคำหลัก (tf) และคำนวณค่าน้ำหนักของคำหลัก (W) ในขั้นตอนการสกัดคำหลักไว้ด้วย หลังจากนั้นทำการจัดกลุ่มความรู้ด้วยการสร้าง Lexical Chain โดยพิจารณาจากลักษณะความสัมพันธ์เชิงความหมายระหว่างคำหลักโดยอาศัยฐานข้อมูลคำศัพท์เวิร์ดเน็ต ตามอัลกอริทึมที่ออกแบบ และระบุค่าน้ำหนักความสัมพันธ์ระหว่างคำ (R) ด้วย จากการทดลองทำให้เกิดกลุ่มคำหลักหรือ Lexical Chain หลายกลุ่ม โดยแต่ละกลุ่มคำหลักจะมีคำที่มีความหมายในแนวคิดเดียวกันเชื่อมโยงสัมพันธ์กัน



ภาพที่ 5 ตัวอย่างการจัดกลุ่มคำหลักของ Lexical Chain<sub>0</sub>

การคำนวณค่าน้ำหนักของ Lexical Chain<sub>0</sub> ตามสมการ  $S_k$  โดยจะได้ค่าน้ำหนักเท่ากับ 2.34 และทำการคำนวณค่าน้ำหนักของ Lexical Chain อื่นๆ ที่ทำได้ในตัวอย่างบทความข่าวด้วย หลังจากนั้น ระบบได้จัดเรียงลำดับค่าน้ำหนักของ Lexical Chain จะแสดงผลหน้าจอเพียง 3 อันดับแรก และผลการทดลองที่ได้ คือ Lexical Chain<sub>2</sub>, Lexical Chain<sub>0</sub> และ Lexical Chain<sub>3</sub>

ตารางที่ 2 ผลการทดลอง

Chain	$S_k$	String
2	3.18857142857143	people floods inundated flood flooding washed agriculture tide
0	2.34	vietnam thailand china hanoi center region provinces related capital areas treated la north ministry
3	1.63	landslides including swept

### 4.2 การทดสอบวัดประสิทธิภาพด้วยค่าเอฟเมเชอร์

การทดสอบเพื่อประเมินประสิทธิภาพการจัดกลุ่มคำหลักในการวิจัยนี้ จะใช้การวัดค่าความถูกต้อง (Precision) ค่าความครบถ้วน (Recall) และค่าเอฟเมเชอร์ (F-measure) [10] เพื่อทดสอบความถูกต้องจากการสุ่มข้อมูลการจัดกลุ่มคำหลักจำนวน 12 กลุ่มคำหลักที่จัดกลุ่มได้ในตัวอย่างเอกสารบทความข่าว

ค่าเอฟเมเชอร์ คือ ค่าการวัดประสิทธิภาพพื้นฐานในการจัดกลุ่มคำหลัก ซึ่งเกิดจากการรวมเอาค่าการวัดความถูกต้อง (Precision) และค่าความครบถ้วน (Recall) มาคำนวณ โดยค่า Recall คือ ค่าที่บ่งบอกถึงอัตราผลลัพธ์ที่ถูกต้องจากการจัดกลุ่มคำหลัก และค่า Precision คือค่าที่บ่งบอกถึงอัตราผลลัพธ์ที่ไม่ถูกต้องจากการจัดกลุ่มคำหลัก สามารถคำนวณได้จากสมการ ดังนี้

$$F - measure = \frac{2RP}{R + P}$$

$$P = \frac{A}{A + B}$$

$$R = \frac{A}{A + C}$$

เมื่อ P คือ ค่าความถูกต้อง (Precision), R คือ ค่าความครบถ้วน (Recall), A คือ จำนวนคำหลักที่สามารถจัดกลุ่มได้ถูกต้อง B คือจำนวนคำหลักที่จัดกลุ่มไม่ถูกต้อง และ C คือจำนวนคำหลักที่ต้องการ แต่ไม่ถูกจัดกลุ่ม จากการทดลองตัวอย่างในงานวิจัยนี้ พบว่าค่า Precision เท่ากับ 91.61% ค่า Recall เท่ากับ 81.92% และค่า F-measure เท่ากับ 84.90%

## 5. สรุปผลและข้อเสนอแนะ

งานวิจัยนี้พัฒนาระบบการสรุปเอกสารเชิงความหมายโดยใช้ออนโทโลยี เพื่อลดเวลาในการพิจารณาคัดเลือกข้อมูลข่าวสารอิเล็กทรอนิกส์ โดยอาศัยเทคโนโลยีทางด้านเชิงความหมาย (Semantic) ในการสกัดคำหลัก การจัดกลุ่มคำหลัก

ด้วยการสร้าง Lexical Chain โดยอาศัยฐานข้อมูลคำศัพท์เวิร์ดเน็ต ซึ่งผลลัพธ์ของการจัดกลุ่มคำหลักที่วัดประสิทธิภาพด้วยค่า Precision เท่ากับ 91.61% ค่า Recall เท่ากับ 81.92% และค่า F-measure เท่ากับ 84.90% ถือว่าผลลัพธ์ที่ได้ค่อนข้างมีประสิทธิภาพดี แต่งานวิจัยนี้มีข้อจำกัดที่สามารถใช้กับบทความภาษาอังกฤษเท่านั้น เนื่องจากอ้างอิงจากฐานข้อมูลคำศัพท์เวิร์ดเน็ต ซึ่งหากมีการพัฒนาฐานข้อมูลเชิงความหมายสำหรับภาษาอื่นๆ และพัฒนางานวิจัยให้สามารถใช้ร่วมกับบทความภาษาอื่นๆ ได้ จะเกิดประโยชน์ยิ่งขึ้น

## 6. เอกสารอ้างอิง

- [1] U. Hahn and I. Mani, "The Challenges of Automatic Summarization," *IEEE Computer Society Press*, vol. 33, pp. 29-36, 2000.
- [2] E. G. M. Petrakis, G. Varelas, A. Hliaoutakis, and P. Raftopoulou, *X-Similarity: Computing Semantic Similarity between Concepts from Different Ontologies.*, Department of Electronic and Computer Engineering Technical University of Crete (TUC).
- [3] J. Morato, M. Á. Marzal, J. Lloréns, and J. Moreiro, "WordNet Applications," *GWC 2004*, pp. 270-278.
- [4] C. S. L. P. University, *WordNet a lexical database for the English language : About WordNet*, Cognitive Science Laboratory Princeton University 221 Nassau St. Princeton, NJ 08542 2006.
- [5] E. Hovy, C. Y. Lin, and D. Marcu, "Automated Text Summarization (SUMMARIST)." [Cited; Available from: <http://www.isi.edu/natural-language/projects/SUMMARIST.html>.]
- [6] C.-Y. Yang. and J. C. Hung., "Word Sense Determination using WordNet and Sense Co-occurrence.," *Proceedings of the 20th International Conference on Advanced Information Networking and Applications.*, vol. Volume 1 (AINA'06), pp. 779-784, 2006.
- [7] P. D. Turney, "Lexical Cohesion - Applications." [Cited; Available from: [http://www.apperceptual.com/ml\\_text\\_cohesion\\_ap.html](http://www.apperceptual.com/ml_text_cohesion_ap.html)
- [8] J. Morris and G. Hirstt, "Lexical Cohesion Computed by Thesaural Relations as an Indicator of the Structure of Text," *Association for Computational Linguistics*, vol. 17, pp. 21-48, 1991
- [9] ศรีสุดา.พิละมาตย์ and ส. อาจอินทร์., "การจัดกลุ่มเอกสารบนเว็บเพื่อการสืบค้นเชิงความหมาย (Web Document Clustering for Semantic Search)," *NCSEC*, pp. 19-21, 2007.
- [10] Guizhen Yang, Saikat Mukherjee, and I. V. Ramakrishnan, "On precision and Recall of Multi-Attribute Data Extraction from Semistructured Sources [Electronic version]," 2003.