



**AUTOMATIC THAI NEWS SUMMARIZATION
USING PRINCIPLE OF SUMMARIZATION
AND INTERPRETATION TEMPLATE**

PIYABUD PLOADAKSORN

**MASTER OF SCIENCE
IN STRATEGIC MANAGEMENT INFORMATION SYSTEM**

MAE FAH LUANG UNIVERSITY

2009

©COPYRIGHT BY MAE FAH LUANG UNIVERSITY

**AUTOMATIC THAI NEWS SUMMARIZATION
USING PRINCIPLE OF SUMMARIZATION
AND INTERPRETATION TEMPLATE**

PIYABUD PLOADAKSORN

**THIS THESIS IS A PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN STRATEGIC MANAGEMENT INFORMATION SYSTEM**

MAE FAH LUANG UNIVERSITY

2009

©COPYRIGHT BY MAE FAH LUANG UNIVERSITY

**AUTOMATIC THAI NEWS SUMMARIZATION
USING PRINCIPLE OF SUMMARIZATION
AND INTERPRETATION TEMPLATE**

PIYABUD PLOADAKSORN

THIS THESIS HAS BEEN APPROVED
TO BE A PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE
IN STRATEGIC MANAGEMENT INFORMATION SYSTEM

2009

EXAMINING COMMITTEE

.....*CLV*.....CHAIRPERSON

(Prof. Dr. Chidchanok Lursinsap)

.....*Kosin Chamngthai*.....MEMBER

(Assoc. Prof. Dr. Kosin Chamnongthai)

.....*Yooyativong, T*.....MEMBER

(Gp. Capt. Dr. Thongchai Yooyativong)

.....*P. Temdee*.....MEMBER

(Dr. Punnarumol Temdee)

© COPYRIGHT BY MAE FAH LUANG UNIVERSITY

ACKNOWLEDGEMENT

I would like to thank many people for their supports encouragement and guidance during my years as a graduate student in Mae Fah Luang University.

First and foremost, I would like to express sincere thanks to my advisor, Dr. Punnarumol Temdee for her help and patience. She has pushed me through the inevitable research setbacks, and encouraged me. I also extend my thanks and gratitude to Associated Professor Dr. Kosin Chamnongthai my co-advisor.

I also thank my other committee members, Prof. Dr. Chidchanok Lursinsap and Dr. Thongchai Yooyativong for their valuable time to discuss and make their comments on my research.

My special thanks are also to many other people at IT school, all friends in SMIS program, Mae Fah Luang University, and even other friends.

Finally most of all, thanks for my parents who have always been supportive, understanding, and encouraging at all times through my entire life period.

Piyabud Ploadaksorn

Thesis Title	Automatic Thai News Summarization Using Principle of Summarization and Interpretation Template
Author	Piyabud Ploadaksorn
Degree	Master of Science (Strategic Management Information System)
Supervisory Committee	Dr. Punnarumol Temdee Assoc. Prof. Dr. Kosin Chamnongthai

ABSTRACT

Online news is a well known application of information delivery in the age of digital society. There is usually time consumption to explore all sources of the online news to consume the information which is usually written with different writing styles to report the same event. Especially, in the case of Thai online news, synonyms, acronyms, superfluous words, and jargon words generally cause a redundancy problem for automatic text summarization. To deal with this problem, this thesis proposes a method for summarization of the Thai online news using a principle of summarization and an interpretation template. According to the principle of the summarization, the interpretation template is generated and filled with the common content extracted from different sources of the online news. To evaluate the quality of the summary, the evaluation task is set up to assess the satisfaction of the readers by five evaluators. The evaluation results show that the proposed method obtains 80%, 73%, 73% and 73 % of grammar, conciseness, clarity and coherence scores.

Keywords: Text summarization / Multi-document summarization / Summarization generation

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENT	(3)
ABSTRACT	(4)
LIST OF TABLES	(7)
LIST OF FIGURES	(8)
CHAPTER	
1 INTRODUCTION	1
1.1 Principle and Motivation	1
1.2 Objectives	2
1.3 Scopes	2
2 LITERATURE REVIEWS	3
2.1 Literature Reviews	3
2.2 Proposed Method	5
2.3 Principle of Summarization	6
2.4 Natural Language Processing	6
3 METHDOLOGY	12
3.1 Analysis and Design Phase	12
3.2 Implementation Phase	25

TABLE OF CONTENTS (continued)

	Page
CHAPTER	
4 EXPERIMENT AND RESULT	31
4.1 Experiment	31
4.2 Evaluation and Result	35
5 DISCUSSION	39
5.1 Analysis of Evaluation Results	39
5.2 Comparison of System and Human Generated Summaries Quality	40
6 CONCLUSION	42
6.1 Conclusion	42
6.2 Future Work and Suggestion	43
REFERENCES	44
APPENDIXES	49
APPENDIX A Evaluation Guideline	50
APPENDIX B Summary Result	51
APPENDIX C Synonym Database	67
CURRICULUM VITAE	73

LIST OF TABLES

Table	Page
3.1 Word Appearance Status in Sentence D1S1 and D2S1	15
3.2 Label Roles of the Grammar	17
3.3 The Detail of Synonym Table	26
3.4 Example of Synonym Table	26
4.1 News Headline	32
4.2 The Details of Evaluation Task	37
4.3 Average Grammar, Coherence, Clarity and Conciseness score of Overall Human Generated Summary	37
4.4 Average Grammar, Coherence, Clarity and Conciseness Score of System Generated Summary	38
4.5 Comparison of Grammar, Coherence, Clarity and Conciseness Score of System Summary and Human Generated Summary	38

LIST OF FIGURES

Figure	Page
2.1 Top-down Parsing	8
2.2 Bottom-up Parsing	9
3.1 Framework of Analysis and Implementation Phase	11
3.2 Overview of the System	12
3.3 Two input articles from two different online sources	13
3.4 Role Parsing of Sentence D1S1	18
3.5 Role Parsing of Sentence D2S1	19
3.6 The Process of Changing Synonym to Common Word of Sentence D1S1	21
3.7 The process of changing synonym to common word of sentence D2S1	22
3.8 The Interpretation Template	24
3.9 The Example of If-Else Statement for Tagging Where Information	27
3.10 The Input Page of Prototype System	28
3.11 The Output Page of Prototype System	29

CHAPTER 1

INTRODUCTION

1.1 Principle and Motivation

Text summarization is a process or a technique to be use in a computer to generate a summary from text input. This summary may serve many purposes, for example, to present the compressed descriptions of search results from search engines; to display summarized information from the Internet on specific areas; and to summarize the news in SMS or WAP format for mobile devices, etc. In addition, the rapid change of technology drives the Internet to grow up as well as the volume of information over the Internet to dramatically increase. Many information and services is turning on the Internet, for example online library, online classroom, online shop, etc. Online news articles are one of the types of information available on the Internet, because many publications are turning into public news services on the Web. This causes information redundancy because the same article is reported in many sources repeatedly. The text summarization is a possible solution to address the problem of overloaded information and information redundancy.

News summarization is one of well known applications of the text summarization. Basically, readers can access online news quickly through the Internet and the World Wide Web technology. Although the readers can access the online news conveniently, it usually requires much time for the readers to browse many sources on the Internet to consume the same news just reported with different writing styles from many sources. Particularly, for the case of Thai online news, synonyms, acronyms, superfluous words, and jargon words are generally associated with the different writing styles. These words normally cause information redundancy and incoherent problems which the traditional technique or the method of the text summarization using the statistical based method cannot handle this problem of information redundancy and

incoherent. Additionally, the opinions from reporters are generally expressed explicitly, which usually causes complexity for knowledge based summarization methods to distinguish between a fact and an opinion.

To serve the requirement of the readers which is to consume non-redundant summarized news in a short time and to obtain a compact summary having only the fact (without opinions) from all sources, this thesis aims to develop an automatic news summarization method for Thai online news. The method proposed in this thesis employs the principle of summarization to extract the main content from the different sources and to generate the summary by completing the developed interpretation template.

1.2 Objectives

The objective of this thesis is to propose a new method for summarizing Thai online news articles from difference sources in order to save time for reading news from several sources and provide an only relevant fact of the news. This thesis also aims to develop a system prototype for an experimental study and to test the proposed method.

1.3 Scopes

The thesis approaches a method of an automatic Thai news summarization. This Thai news is from different sources of online news. The output is a compact summary that limits one sentence. The prototype is a web application for testing the proposed method.

CHAPTER 2

LITERATURE REVIEW

This chapter explains the literature review and the related works for an automatic text summarization system. The proposed method and computational theories used in this thesis are also explained in this chapter.

2.1 Literature Review

2.1.1 Automatic Text Summarization

Automatic text summarization refers to the process of identifying important information and generating a condensed summary from the original document. Text summarization has been proposed for many genres of documents such as news articles (Chen, Kuo, Huang, Lin & Wung, 2003), research papers (Ou, Khoo & Goh, 2005), legal documents (Farzindar & Lapalme, 2004) and medical documents (Afantenos, Karkaletsis & Stamatopoulos, 2005). Types of summarization can be categorized by considering the types of input documents, which can be categorized into two types of methods including single document ((Ausdang & Chuleerat, 2003), (Pensiri & Sukree, 2008), (Thana, Asanee & Caelen, 2007)) and multiple documents (Radev, Winkel, & Topper, 2002). Generally, a single document summarization system aims to produce a summary from one original document while a multi-document summarization system aims to generate a single concise summary of a group of documents covering the same topic. The summary should preserve a main idea of the information contained in the original documents. This evidence presents that the multi-document summarization has more challenges than the single document one in several aspects such as redundancy, cohesion, and coherence. Consequently, traditional single-document summarization approaches do not always work well for a multiple document.

The methods or the techniques used in the previous research are broadly categorized into two types including a statistical based method and a knowledge based method. The statistical based method emphasizes to use statistics to find representations of significant features such as words, sentences or paragraphs in the original document for being selected as a summary. Some techniques employing the statistical based method are singular value decomposition (Ausdang & Chuleerat, 2003) and a centroid-based method (Radev et al., 2002). These methods aim to extract relevant information from the original documents and to arrange the relevant information in the summary as found in the original order. Ausdang & Chuleerat (2003) proposed Thai text summarization using Singular Value Decomposition. This system is an extracted paragraph by using the Singular Value Decomposition which is a linear algebra theory to calculate and to find the paragraph vector. The vector having the highest score represents the important paragraph, and the summary is extracted from that paragraph. These systems generally give incoherent output because the extracted sentences are ordered without consideration of the continuity of information. Radev et al. (2002) introduced a centroid-based multi-document summarization system or MEAD. This system is a multi-document summarizer that has been approached on statistical based methods. The system uses a centroid-based technique which is a clustering technique to measure the centroid of the document's cluster so that scores of all sentences can be created in a cluster by considering features such as distances between the sentences and the cluster centroid, distances between the sentences and the title, etc. After ranking the sentences in the cluster, summaries can be produced by extracting the highly relevant sentences according to a percentage of a compression rate.

On the other hand, knowledge based methods emphasize the use of techniques and knowledge based techniques, such as natural language processing ((Ou et al., 2005), (Farzindar & Lapalme, 2004)) Rhetorical Structure Theory ((Ou et al., 2005), (Farzindar & Lapalme, 2004), (Radev, Otterbacher, Winkle, & Blair, 2005), (Radev et al., 2002)), cross-document-relation (Radev et al., 2002), for processing or synthesizing information which can then generate an abstract summary in the form of natural language output. Rhetorical structure theory (RST) is a theory of the organization of natural text that provides a framework and a useful method to analyze text structure in terms of relations between parts of the text such as discourse structure (Mann, 1999). This method can address

the incoherent and the word ambiguity problems that occur in the statistical based methods but the requirement usually implies complex methods and an additional step is often required to perform the summarization.

For the previous research on the automatic text summarization for Thai language, most of the research aims to summarize the single document by using statistical based analysis methods ((Ausdang & Chuleerat, 2003), (Pensiri & Sukree, 2008), (Thana et al., 2007)). This kind of the summarization system usually presents the summary in a non-natural language form. A lack of content consistency which causes contrasting information is a result from a shift of information in location of the context and the extraction without regeneration into normal language. Additionally, there is generally a use of informal conversational language, superfluous words, and jargon words produced in a variety of writing styles in Thai texts, which also usually causes redundancy problems. Therefore, it is a challenging problem to obtain a good summarization method which can eliminate the redundancy and generate compact readable summaries.

To cope with the problems occurring in the previous Thai text summarization system as mentioned before, this thesis aims to propose the method for summarizing Thai text. This method approaches the knowledge based method by considering human's knowledge being used to generate the summary.

2.2 Proposed Method

This thesis proposes a new method of automatic text summarization to handle the overloaded information and information redundancy problems. The method proposed here is particularly to summarize the same Thai online news article from different sources and then generate a compact summary. This proposed method employs the knowledge based method to cope with the informal conversational language, superfluous words, and jargon words produced in a variety of writing styles which usually cause redundancy problems in Thai text. An analysis of the method is based on a principle of summarization. Additionally, this proposed method aims to generate the compact summary according to the principle of summarization

concerning a story main idea referring to “Who do what? With whom? Where? How? Or Why?” Therefore, the Interpretation template proposed in this thesis basically answers the question through the interpretation template.

2.3 Principle of Summarization

In general, human creates summaries by reading and understanding a main idea of a story. The main idea is “Who do what? With whom? Where? How? Or Why?” The next step is to rewrite on their own with their languages. The principles of summarization (Ratree, 2008) basically explain how to make a summary. These principles can be summarized as follows:

1. Revise the story in the own language of a summarizer.
2. Transform words that are difficult to understand to more common words that are easier to understand.
3. Transform first-person and second-person pronouns to third-person pronouns or mention a name instead of a third-person pronoun.
4. Transform a conversation to narration so that there are no quotation marks.
5. There is unnecessarily a need to order the summaries as in the original version of the texts, but the readers must be able to easily read and capture the main idea.

2.4 Natural Language Processing

Natural Language Processing (Liddy, 2001) is a subject between computer science and linguistics that is concerned to use computational techniques for analyzing and representing human natural language. Whereas the linguistics focuses on formal and structural models of language, the computer science focuses on the development of data representation and competence processing of these structures.

Basically, the natural language processing focuses on two distinctions: natural language processing and natural language generation. Firstly, natural language processing systems aim to force the computer to understand human language by converting the human language into more

formal representations such as parsing trees or first order logic that are easier for computer programs to manipulate. On the other hand, natural language generating systems aim to convert information from the computer into human like language.

2.4.1 Level of Natural Language Processing

In order to understand natural language texts or spoken languages that people use, it is important for the computer to be able to process for understanding meanings of those languages. Liddy (2001) divided levels of a language analysis into seven interdependent levels referred as a synchronic model of language. Details are shown as follows:

1. Phonology; this level deals with pronunciation or interpretation of speech sounds within and across the words.
2. Morphology; this level deals with the componential nature of words which are composed of morphemes, the smallest parts of words that carry meanings or suffixes and prefixes.
3. Lexical; this level deals with lexical meanings of words and parts of speech analyses.
4. Syntactic; this level focuses on analyzing the words in a sentences to uncover the grammar and the structures of the sentences, therefore this level requires the grammar and the parser.
5. Semantic; this level deals with meanings of words and sentences, determines a possible meaning of a sentence by focusing on the interactions among word level meanings in the sentence.
6. Discourse; this level deals with the structures of different texts using document structures by focusing on the properties of the whole text that convey meanings and making a connections between component sentences.
7. Pragmatic; this level deals with the knowledge from the outside which is concerned in purposeful use of language in situations and utilizes context over and above the contents of text understanding.

2.4.2 Parsing

Parsing is a process of syntactic analysis of input sentences to determine their grammatical structures. It is an indispensable component in a natural language processing system and plays an important role for systems such as machine translation, automatic summarization, question answering, etc. Moreover, the parsing is a complicated task since it must cover a broad range of linguistic phenomena. Especially, the parsing of Thai language is a problematic task in Thai language processing.

To parse a sentence, it is essential to determine a way which that sentence can be generated from the start symbol of the grammar. There are two essential ways of parsing including top-down parsing and bottom-up parsing. Firstly, the top-down parsing begins with finding left-most derivations of an input-stream and applying the grammar rules to forward them to a top-down expansion. The example output of this parse tree is shown in Figure 2.1. Secondly, the bottom-up parsing is opposite to the top-down parsing. It begins with a sentence to be parsed and applies the grammar rules reverse derivations from the terminal, the example is shown in Figure 2.2.

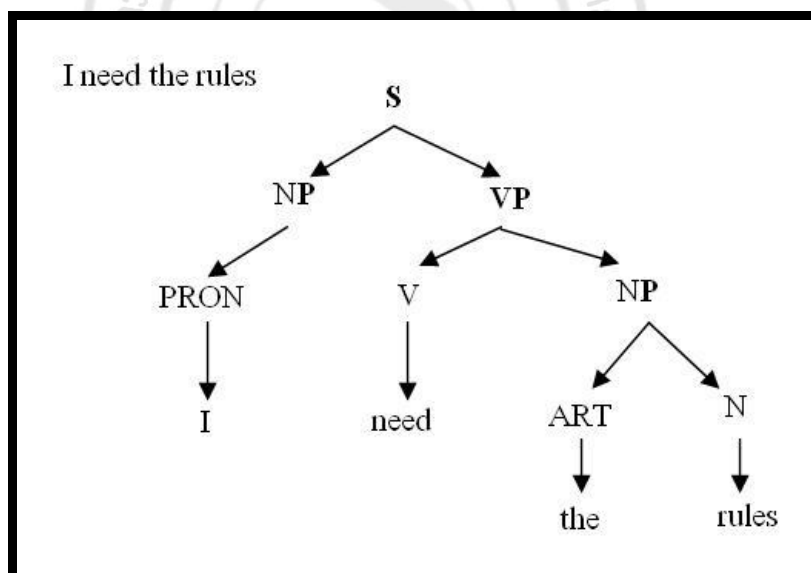


Figure 2.1 Top-down Parsing

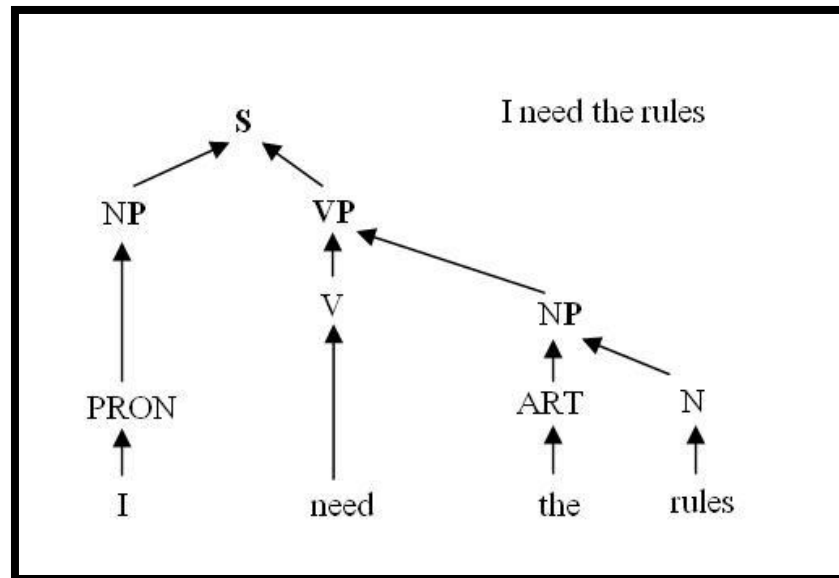


Figure 2.2 Bottom-up Parsing

2.4.3 Natural Language Processing Applications

There are many applications that apply natural language processing. The most well-known applications that employ the natural language processing include (Liddy, 2001) the following:

1. Information Retrieval

Information retrieval is science of searching documents that focus on a user information need. An information retrieval task is concerned with retrieving relevant documents to a query (Amit, 2001). The application of the natural language processing to the information retrieval is such the intelligent information retrieval applying the natural language processing to attempt to automatically assign controlled vocabularies, semantic search expansion query and etc.

2. Question-Answering

Question answering refers to a task of automatically answering a natural language question (Hovy, Gerber, Hermjakob, Junk, & Lin, 2000). To apply the natural language processing to the question-answering system requires complex natural language processing techniques to understand the text to perform correct answers.

3. Machine Translation

Machine translation is a task of a computer to translate one natural language text to another language (Dorr, Hovy & Levin, 2004). The machine translations apply the natural language processing to a grammatical and syntactic analysis for improving the system.

4. Automatic Summarization

Automatic summarization refers to a process of reducing a text document or multiple documents into a short summary but still maintaining a main idea of the original text. (Goldberg, 2007). To apply the natural language processing in the automatic summarization is to understand the texts and generate the natural language texts.



CHAPTER 3

METHODOLOGY

The thesis methodology consists of two sections which are an analysis and design phase and an implementation phase as demonstrated in Figure 3.1.

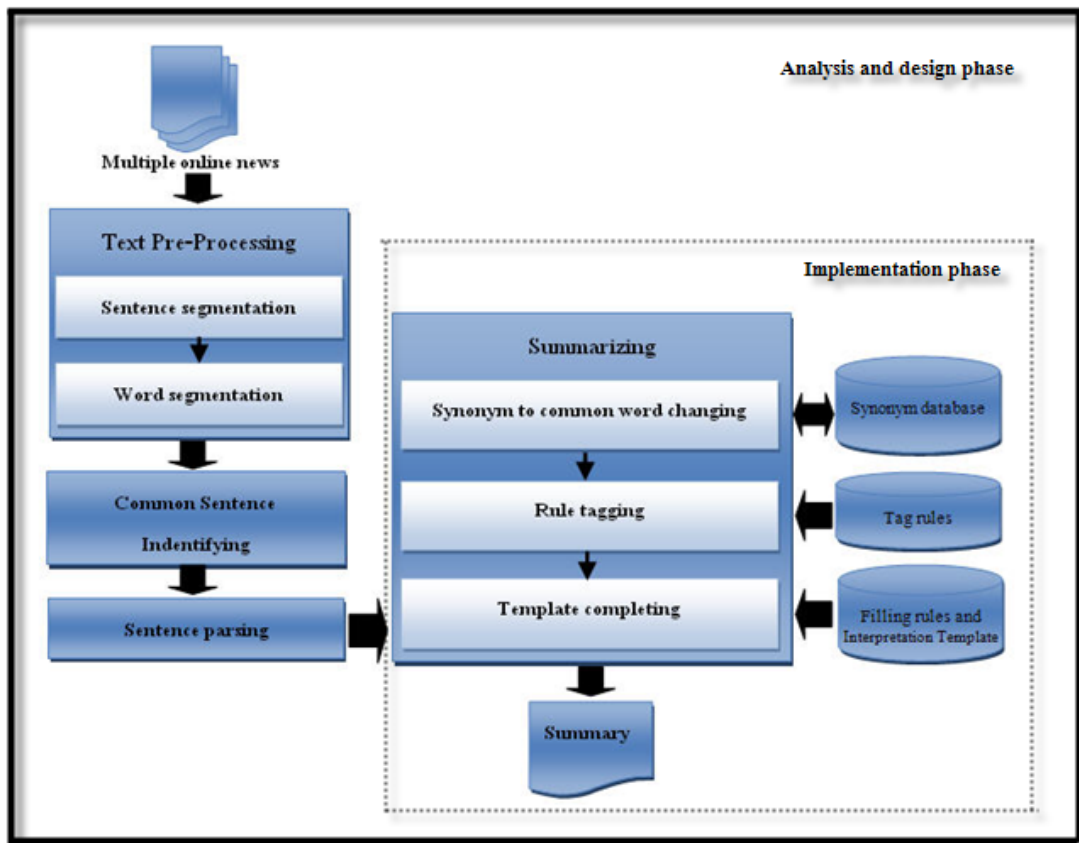


Figure 3.1 Framework of Analysis and Implementation Phase

Firstly, the design phase explains the proposed system of this thesis which covered the system overview and details of each process in the proposed system. The second phase is the implementation phase. This section shows the implementation of the system's prototype for testing the proposed system.

3.1 Analysis and Design Phase

This section describes the overview of the proposed system and the details of each process in the proposed system, and simultaneously the examples of each process are provided.

3.1.1 System Overview

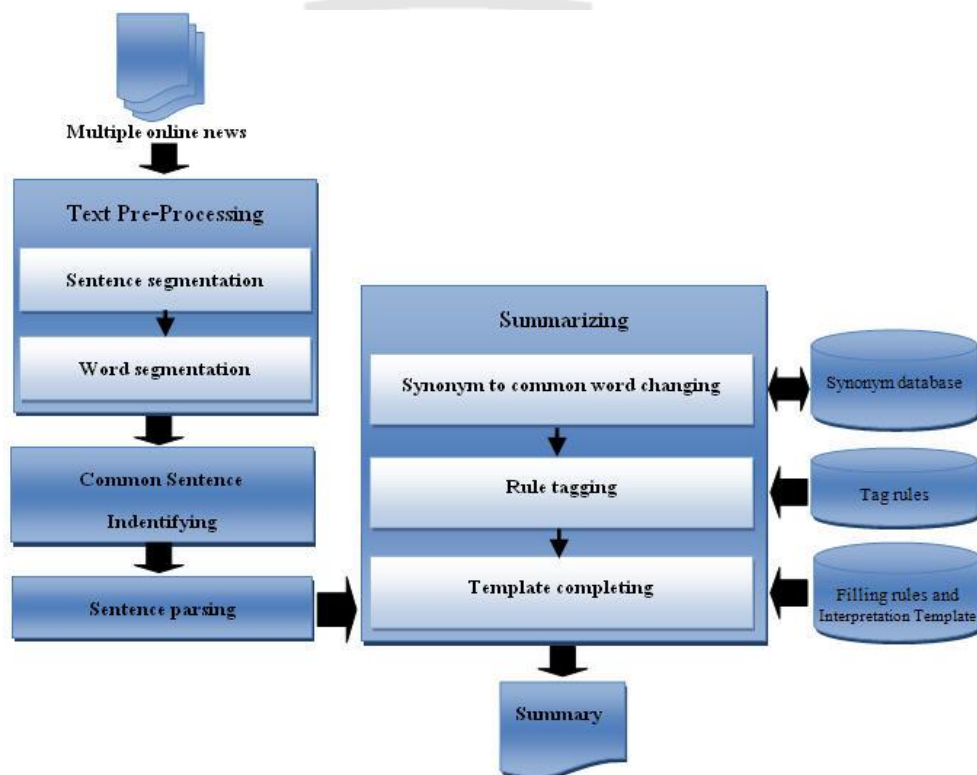


Figure 3.2 Overview of the System

The detail described here is the proposed summarization method for multiple documents, which summarizes the same new articles from different sources of online news. The overview of the system is shown in Figure 3.2. There are four processes proposed here including (1) Text pre-processing, (2) Common sentence identifying, (3) Sentence parsing and (4) Summarizing.

Firstly, the same news articles from the different online sources are input to text pre-processing for sentence and word segmentation. In this process, all texts in all documents are broken into sequences of words. The second process is common sentence identifying which is used to find the overlapping content among input sentences by computing the similarity of each sentence. The most similar sentences are chosen as relevant sentences. The next process is the sentence parsing. The obtained relevant sentences in the previous process are parsed and categorized into sets of labeled semantic roles. The last process is the summarizing process. This process consists of three sub-processes including synonym conversion, word labeling and template completion. The labeled semantic roles are selected by using the developed tag rules to complete the interpretation template to obtain a summary as an output. The interpretation template used here is predefined regarding the principle of summarization.

The details of each process are described in the next section by using examples of documents shown in figure 3.3 as input documents.

<p>Document 1 ตร.สน.ทองหล่อได้เข้าจับตัวโจรขโมยรถคาห้องพักแถวสุขุมวิท Polices from Thonglor Police Station arrested the car robber at the accommodation in Sukhumwit.</p>
<p>Document 2 วันนี้ 10 พ.ค.ตำรวจทองหล่อได้เข้าจับกุมแก๊งค์ลักรถที่ห้องพักย่านสุขุมวิท หลังจากสอบสวนพบว่าโจรกลุ่มนี้ได้ขโมยรถมาแล้วกว่า 10 คัน Today Oct 10th, Thonglor polices arrested the gang of car robber at the accommodation in Sukhumwit. After the investigation, this gang has stolen more than 10 cars.</p>

Figure 3.3 Two Input Articles from Two Different Online Sources

3.1.2 Text Pre-processing

Text pre-processing is an initial process to prepare the input data before passing into the next process. For the system proposed here, the pre-processing stage is divided into two sub processes which are sentences segmentation and words segmentation. Firstly, the process starts with segmenting the input documents into sets of sentences and then segmenting the words in those sentences. In this thesis, this word and sentence segmentation is done manually. The outputs of those examples are shown below where D indicates documents and S indicates sentences.

D1S1: |ตร.|สน.|ทองหล่อ|ได้|เข้า|จับตัว|โจร|ขโมย|รถ|คา|ห้องพัก|แถว|สุขุมวิท|

D1S1: Polices from Thonglor Police Station arrested the car robber at the accommodation in Sukhumwit.

In this example, D1S1 is the sentence 1 from the document 1.

D2S1: |วันนี้|10|ต.ค.|ตำรวจ|ทองหล่อ|ได้|พา|กำลัง|เข้า|จับกุม|แก๊งค์|ลัก|รถ|ที่|ห้องพัก|ย่าน|สุขุมวิท|

D2S1: Today Oct 10th, Thonglor polices arrested the gang of car robbers at the accommodation in Sukhumwit.

In this example, D2S1 the is sentence 1 from the document 2

D2S2: |หลังจาก|สอบสวน|พบว่า|โจร|กลุ่ม|นี้|ได้|ขโมย|รถ|มา|แล้ว|กว่า|10| คัน|

D2S2: After the investigation, this gang has stolen more than 10 cars.

In this example, D2S2 is the sentence 2 from the document 2.

3.1.3 Common Sentence Identifying

From the above sentences, the common information among groups of documents is defined by identifying the relevant sentences having overlapping information or the main idea. The cosine similarity (Garcia, 2006) between a pair of sentences is used to identify the relevant sentences. The cosine similarity is defined in equation (3.1).

$$Sim(S_1, S_2) = \frac{\sum_{i=1}^k a_i \times b_j}{\sqrt{\sum_{i=1}^k a_i^2} \times \sqrt{\sum_{j=1}^k b_j^2}} \quad (3.1)$$

Where, S_1 and S_2 are input sentences.

a_1, a_2, \dots, a_k are words contained in sentence S_1

b_1, b_2, \dots, b_k are words contained in sentence S_2 .

In this paper, the sentences having less than 5 words are ignored. From the equation (3.1), $\sum_{i=1}^k a_i \times b_j$ represents the appearances of the words in both sentences (S_1 and S_2) while a_i and b_j represent the simultaneous appearance of all words in S_1 and S_2 respectively. To address an issue of differing sentence lengths, greater differences have an impact on the similarity value. The approach used here determines the threshold of the difference in the sentence length as $|a_i - b_j| > 10$.

The cosine similarity process starts with identifying if the word appears in each pair of compared sentences. Table 3.1 shows the appearance status of each word in D1S1 and D2S1.

Table 3.1 Word Appearance Status in Sentence D1S1 and D2S1

	D1S1	D2S1
10	0	1
ต.ค.	0	1
กำลัง	0	1
แก๊งค์	0	1
ขโมย	1	0
เข้า	1	1
คา	1	0
จับกุม	0	1
โจร	1	0
ได้	1	1
ตร.	1	0
ตำรวจ	0	1

Table 3.1 (continued)

	D1S1	D2S1
แถว	1	0
ทองหล่อ	1	1
ที่	0	1
พา	0	1
ย่าน	0	1
รถ	1	1
ลัก	0	1
วันนี้	0	1
สน.	1	0
สุขุมวิท	1	1
ห้องพัก	1	1

From Table 3.1 which shows the number 0 and 1, these numbers are the appearance time of each word in each sentence.

The cosine similarity of D1S1 and D2S1 is defined as

$$Sim(D1S1, D2S1) = \frac{6}{\sqrt{13 \times 17}} = 0.404.$$

The cosine similarity of D1S1 and D2S2 is defined as

$$Sim(D1S1, D2S2) = \frac{4}{\sqrt{13 \times 15}} = 0.287.$$

From the above 3 sentence examples in section 3.1.2, all cosine similarities of each pair of the sentences are

$$Sim(D1S1, D2S1) = 0.404.$$

$$Sim(D1S1, D2S2) = 0.287.$$

The sentence pair having the highest similarity score is chosen to be the relevant sentences and the inputs for the next process. As shown in this example, the sentence 1 of the document 1 and the sentence 1 of the document 2 are selected.

3.1.4 Sentence Parsing

This process labels all words in the relevant sentences with syntactic roles. The roles used in this paper are shown in Table 3.2. The process which is done manually for this thesis is the grammar used in “Principles of the Thai Language” by Kumchai Thonglaw (1997). From the above two relevant sentences, the roles are labeled as shown in Figure 3.4.

Table 3.2 Label Roles of the Grammar

Syntactic role	Notation
Noun	N
Verb	V
Auxiliary Verb	Aux
Adverb	Adv
Adjective	Adj
Conjunction	Conj
Object	OBJ
Pronoun	Pron
Preposition	Prep
Noun Phrase	NP
Verb Phrase	VP
Preposition Phrase	PP

From the previous process of the common sentence indentifying, the sentence 1 of the document 1 and the sentence 1 of the document 2 are selected. In this process, these two sentences are parsing with the syntactic roles shown in Figure 3.4 and 3.5. Figure 3.4 shows the parsing of the sentence D1S1, which consists of groups of noun phrases and verb phrases. These groups of noun phrases are both the subject and the object of the sentence, such as the noun phrase is a subject consists of a group of nouns; ตำรวจ. (Police), สถานี. (Police station) and ทองหล่อ (Thonglor; the name of location).

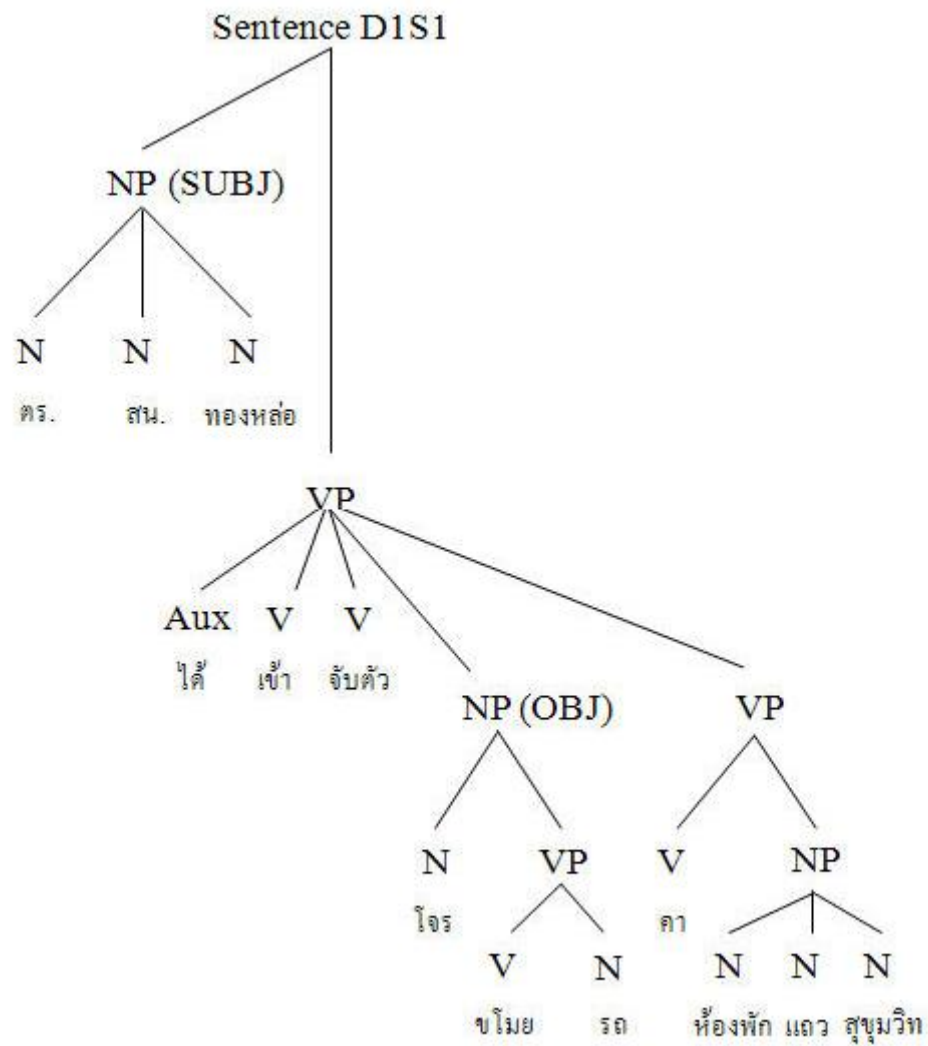


Figure 3.4 Role Parsing of Sentence D1S1

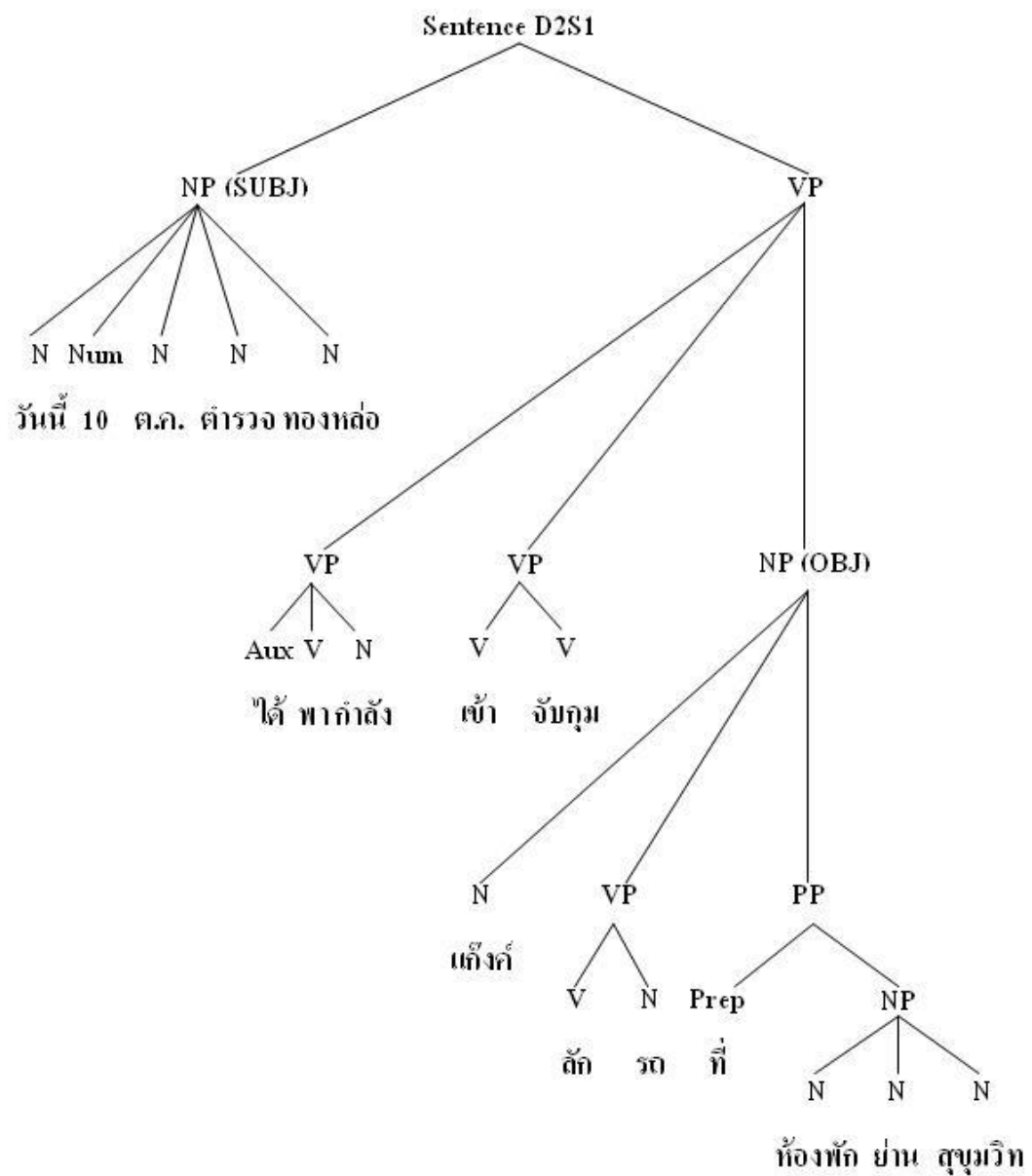


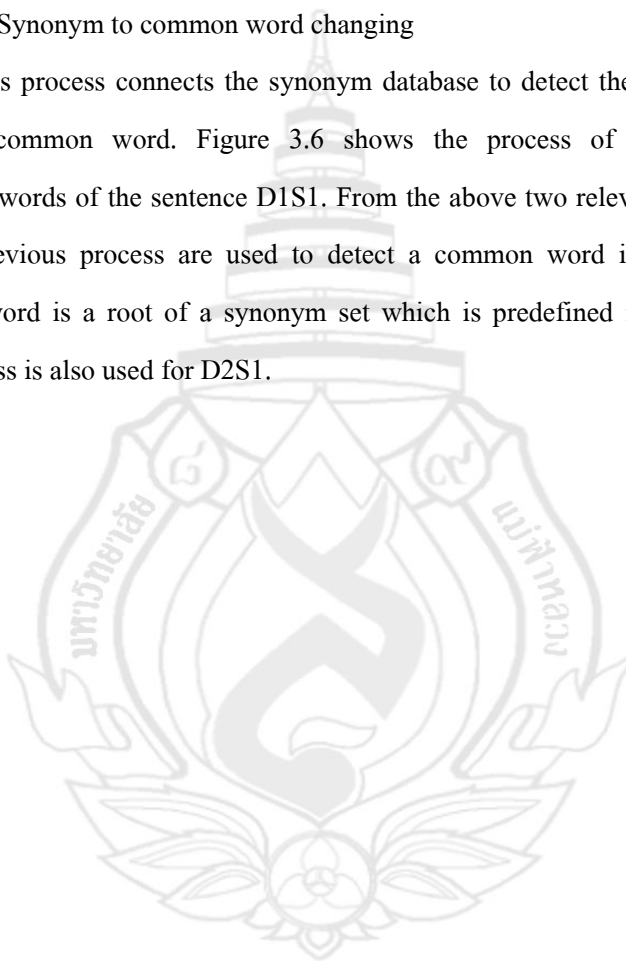
Figure 3.5 Role Parsing of Sentence D2S1

3.1.5 Summarizing

This process mainly generates a summary. There are three databases used in this process including the synonym database containing sets of synonym words, the tag rules database containing the rules used to tag the words for completing the template and the interpretation template database containing the rules for filling information to the template. From Figure 3.2, there are three sub-processes in the summarization process as follows:

1. Synonym to common word changing

This process connects the synonym database to detect the word synonym and then identifies the common word. Figure 3.6 shows the process of changing the synonyms to the common words of the sentence D1S1. From the above two relevant sentences, the labeled roles in the previous process are used to detect a common word in the synonym database. The common word is a root of a synonym set which is predefined in the synonym database. The same process is also used for D2S1.



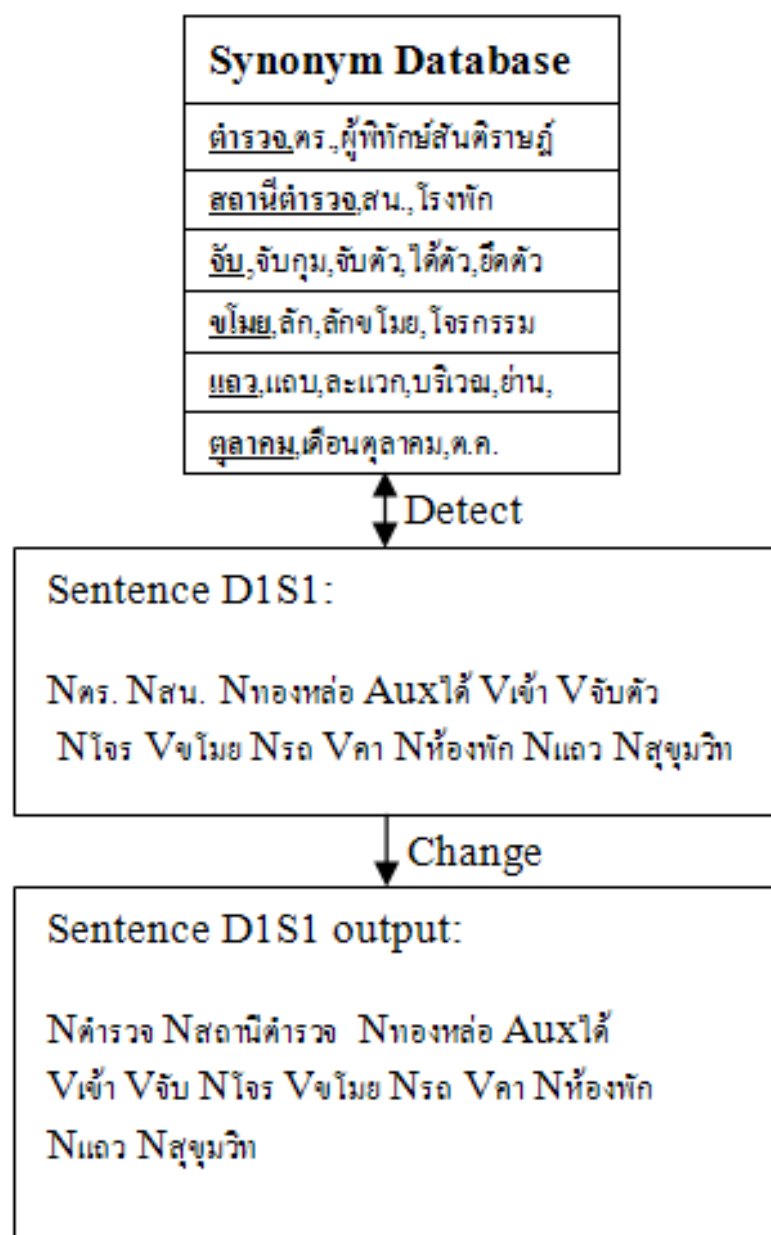


Figure 3.6 The Process of Changing Synonym to Common Word of Sentence D1S1

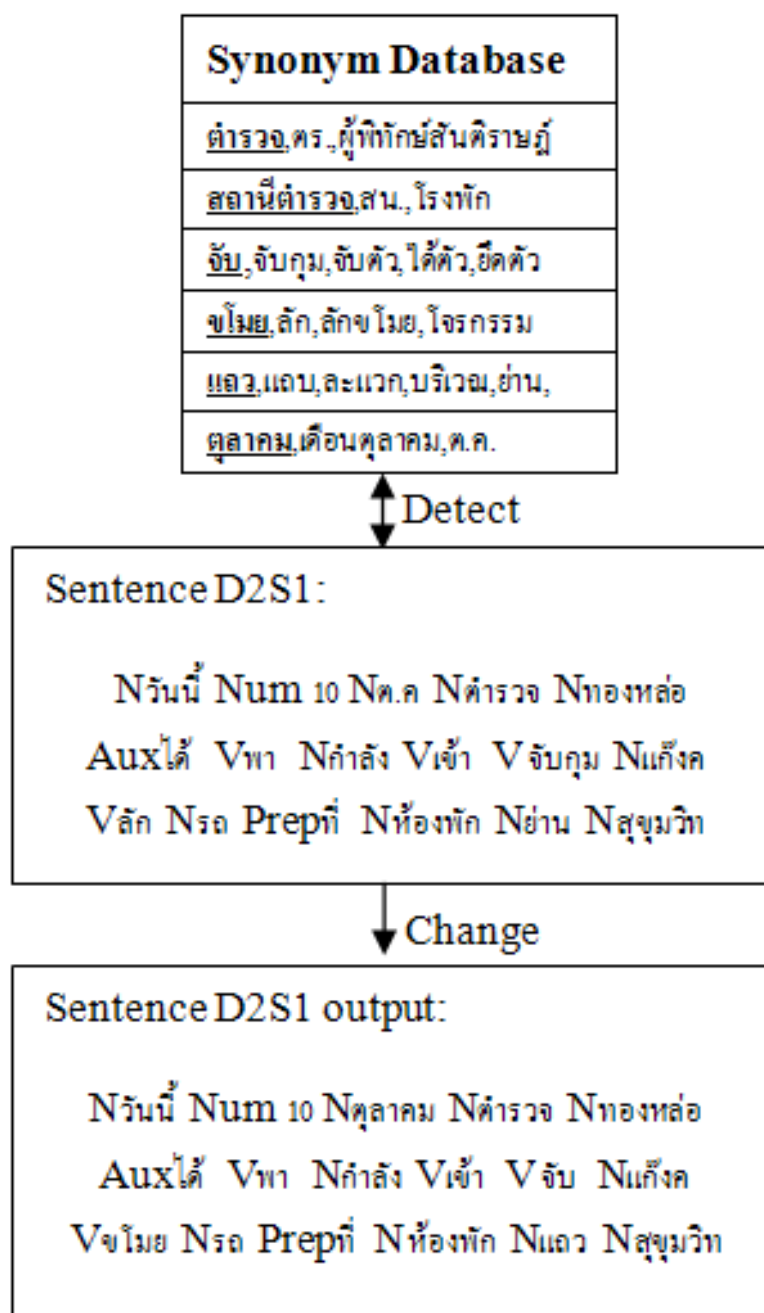


Figure 3.7 The Process of Changing Synonym to Common Word of Sentence D2S1

2. Rules tagging

After obtaining the set of the common words from the above relevant sentences, then the pre-defined tag rules are used to tag the words or phrases to be filled into the interpretation template. These rules are constructed with principles of Thai language knowledge. There are six types of tag rules used in this paper including When, Who, Whom/What, Action, Where, How/Why tags. The basic details of these tags are shown below.

1) When tag: defined as follows

- a) Numbers beginning with the clue words such as “วันที่ (date)”, “วันนี้ (today)”, “เมื่อวานนี้ (yesterday)”
- b) Numbers referring to the months such as ธันวาคม (December), มกราคม (January), ก.พ. (February).

2) Who tag: defined as follows

- a) A noun or a noun phrase that is a subject Noun or a noun phrase beginning with the cue words such as นาย (Mister.), นางสาว (Miss), or words that show a title of a person, for example ศาสตราจารย์ (Professor), ร้อยเอก (Captain), คุณหญิง (Lady), etc.

3) Whom/What tag: defined as follows

- a) Noun or noun phrase that is an object

4) Action tag: defined as follows

- a) Verb or verb phrase of a sentence
- b) Consider only a main verb

5) Where tag: defined as follows

- a) Noun or noun phrase beginning with a preposition word used to identify a place, such as ที่ (at), ใน (in), ใกล้ (near), etc.

6) How/Why tag: defined as follows

- a) Phrase or sentence beginning with a preposition or a conjunction word such as ซึ่ง (which), เพื่อ (for), โดย (by), เนื่องจาก (because), etc.

As mentioned above, there are some tag rules considering the subject and the object of the sentence other than the words or the phrases for tagging. In this thesis, the classifications of the subject and the object of the sentence have done manually together with the process of sentence parsing.

The examples below demonstrate the outputs of the rule tagging:

Sentence D1S1, this sentence has only 3 tags including

{WHO} ตำรวจสถานีตำรวจทองหล่อ (Polices from Thonglor police station)
 {ACT} เข้าจับ (arrest)
 {WHOM/ WHAT} โจรโจรมอเตอร์ (Car robber)

Sentence D2S1 has 6 tags including

{DATE} 10ตุลาคม (10 December)
 {WHO} ตำรวจทองหล่อ (Tonglor's police)
 {ACT} พากำลัง (lead to)
 {ACT} เข้าจับ (arrest)
 {WHOM/ WHAT} แก๊งค์มอเตอร์ (Gang of car robbers)
 {WHERE} ที่ห้องพักย่านสุขุมวิท (at the accommodation in Sukhumwit)

3. Template completion

The interpretation template is developed here using the principle of summarization, news component and reader satisfaction (Lynch, Nesbitt & Duke, 2002). Basically, in order to fill the interpretation template, it is necessary to answer the questions “Who, does What with/to Whom/What Where, Why/How and When” The interpretation template is shown in Figure 3.8.

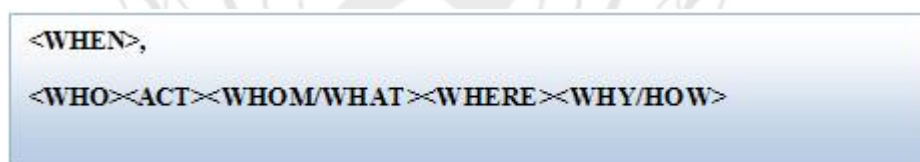


Figure 3.8 The Interpretation Template

The process that completes the template requires rules to select the tagged information from the previous process to fill into the template. The rules for filling the template used in this thesis are shown below.

1) For the same tags from the different sentences having the same meaning, the longer word is selected to fill the template. For example,

D1S1:{WHO} ตำรวจสถานีตำรวจทองหล่อ (Polices from Thonglor police station) and

D2S1: {WHO} ตำรวจทองหล่อ (Thonglor's police).

Because the {WHO} tag of D1S1 is longer than those D2S1, the {WHO} ตำรวจสถานีตำรวจทองหล่อ (Polices from Thonglor police station) of D1S1 is selected to fill the template.

2) For the same tags from the same sentences, the tags that particularly overlap with other sentences are considered particularly. For example,

D1S1: {ACT} เข้าจับ (arrest)

D2S1: {ACT} พากำลั้ง (lead to) and {ACT} เข้าจับ (arrest)

Because {ACT} เข้าจับ (arrest) of D1S1 overlaps with {ACT} เข้าจับ (arrest), the {ACT} เข้าจับ (arrest) is selected to fill the template.

3) If the information of WHO, ACT and WHOM/WHAT tags is blank, the system will return to the common sentence identifying process to select other sentence pairs.

4) When WHEN, WHERE and HOW/WHY are missing, the system will be allowed to generate a summary without those tags.

From the relevant tags information from the process of rule tagging in section 3.1.5.2., after selecting the information to fill the template in this process, the result of this process is now shown as follows:

วันที่ 10 ตุลาคม, ตำรวจสถานีตำรวจทองหล่อเข้าจับโจร โขมยรชที่ห้องพักย่านสุขุมวิท

On October 10, the polices in Thonglor police station arrested the car robber at an accommodation in Sukhumwit.

As mentioned before, the prior processes including the text pre-processing, the common sentence indentifying and the sentence parsing are done manually to prevent the errors from the parsing process.

3.2 Implementation Phase

The implementation phase shows how to implement the system's prototype. This thesis attempts to develop the prototype of system based on the web application which concentrates on the only summarizing process in the proposed system. The implement of the prototype consists of the synonym database and the interface, and their details are explained as follows.

3.2.1 Synonym Database

This thesis constructs a synonym database for testing the process of a synonym to common word change which contains words and synonym words. The words and their synonyms are collected from many sources of Thai online news. This database has one table, which is a synonym table consisting of three attributes including word, status and set. Details of the synonym table are shown in Table 3.3. An example of the synonym database used in this thesis is shown in Table 3.4. The word refers to words and also their synonyms. The status means the status of each word. There are two kinds of the status used in this thesis including the common status and the unidentified status. The set indicates the same type of the synonyms. This synonym database used in this thesis is shown in Appendix C.

Table 3.3 Details of Synonym Table

Attribute Name	Type	Length	Primary Key
words	char	50	Yes
status	char	20	-
Set	integer	10	-

Table 3.4 Example of Synonym Table

Words	Status	Set
อย.	unidentified	1021
องค์การอาหารและยา	common	1021
สำนักงานคณะกรรมการอาหารและยา	unidentified	1021

3.2.2 The Implementation of the Rule Tagging and Filling

In this thesis, the rules that are used to tag and select the information to fill the template are implemented with PHP language by using If-Else statement. The examples of the rule implemented with the If-Else statement are shown in Figure 3.9.

IF word's property = preposition and beginning word = ที่ (at), ใน (in), ใกล้ (near)
THEN label "WHERE"

Figure 3.9 Example of If-Else Statement for Tagging Where Information

3.2.3 User Interface

The user interface is a part of human (user) and computer interaction. In this thesis, the user interface (GUI) consists of two main interfaces including the input page as shown in Figure 3.10 and the output page shown in Figure 3.11. These two interfaces are implemented with HTML and PHP language.

1. Input page

The input page allows the user to interact with the system. The user inputs two sentences that are already done by the process of sentence parsing. The system will be executed a summary of the two input sentences by pressing a 'Sum' button.

Sum demo

PrincipleSUM(DEMO)

Sentence 1

Conj_เมื่อ|N_วันที่|Num_19|N_ธ.ค.|Prep_ที่|N_กระทรวงสาธารณสุข|Subj:np_พ.พิพัฒน์ ยิ่งเสรี เลขาธิการอาหารและยา|N_อย.|V_เปิดเผย|Aux_ว่า|Vp_ได้รับแจ้งว่า|Obj:vp_มีการลักลอบจำหน่ายยาแก้หวัดปลอมให้กับร้านขายยาและสถานพยาบาลบางแห่ง

Sentence 2

Conj_เมื่อ|N_วันที่|Num_19|N_ธันวาคม|Conj_ที่|V_ผ่าน|Adv_มา|Subj:np_พ.พิพัฒน์ ยิ่งเสรีเลขาธิการคณะกรรมการอาหารและยา|V_กล่าวว่า|Np_สำนักงานคณะกรรมการอาหารและยา|N_อย.|Vp_ได้รับแจ้งว่า|Obj:vp_มีการลักลอบจำหน่ายยาแก้หวัดปลอมให้แก่ร้านขายยาและสถานพยาบาลบางแห่ง

Sum Clear

Figure 3.10 Input Page of Prototype System

2. Output page

The output page shows an output summary of input sentences. This output summary is a compact sentence that is generated according to the interpret template as mentioned in the summarizing process in Section 3.1.5.



Figure 3.11 Output Page of Prototype System

From Figure 3.11 that shows the output summary, the user can get back to the first page by clicking “back to Input page”.

3.1.6 System Requirements

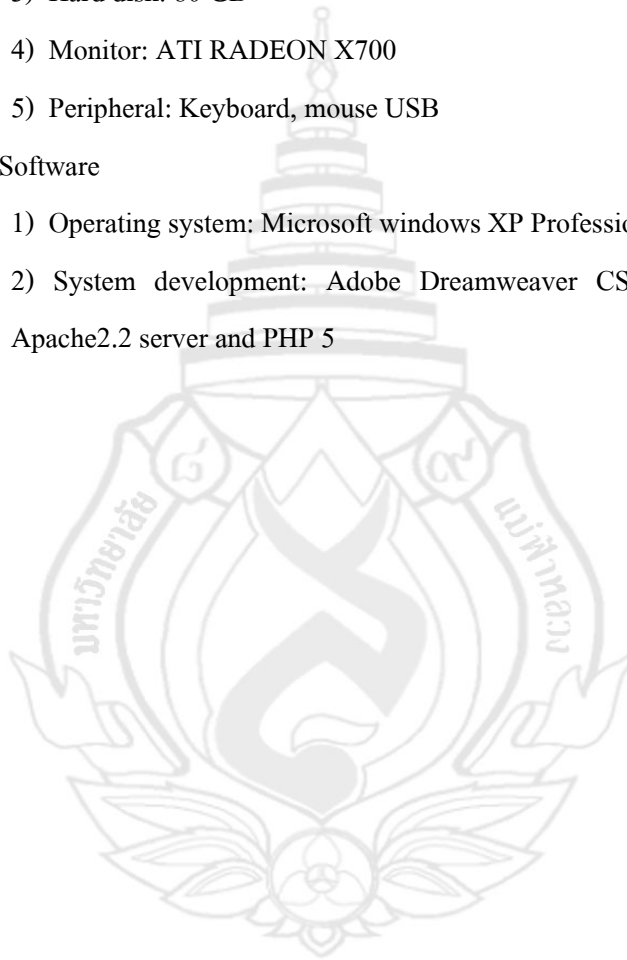
This thesis processes the minimum requirements for hardware and software as follows:

1. Hardware

- 1) CPU: Intel (R) Pentium(R) 1.73 GHz.
- 2) Ram: 1.00 GB
- 3) Hard disk: 80 GB
- 4) Monitor: ATI RADEON X700
- 5) Peripheral: Keyboard, mouse USB

2. Software

- 1) Operating system: Microsoft windows XP Professional.
- 2) System development: Adobe Dreamweaver CS4, Navicat for MySQL, Apache2.2 server and PHP 5



CHAPTER 4

EXPERIMENT AND EVALUATION

This chapter consists of two sections. Section 4.1 describes an experimental study consisting of the data set used to test the proposed method. Secondly, section 4.2 describes the details of evaluation.

4.1 Experiment

Online news articles having the same topic are used for testing the proposed method. This online news is taken from different news sources including Matichon (Matichon, 2008), Komchudluek (Komchadluek, 2008), Manager (Manager online, 2008), and Dailynews (Dailynews online newspaper, 2008). This news is general, political and entertainment news.

The proposed method was tested with 66 documents of 30 news headlines subsumed into three groups regarding political, entertainment and general news. Each type contains 10 news headlines. The headlines of the test articles are shown in Table 4.1. The task is to create a one sentence summary. The summary results of each headline are shown in Appendix B.

Table 4.1 News Headlines

Headline ID	Headline	Number of documents
P001	ม.ร.ว.สุขุมพันธุ์ บริพัตร ได้รับหนังสือรับรองเป็นผู้ว่าการ กรุงเทพ M.R. Sukhumpan Baripatr received a certificate guaranteeing Bangkok Governor.	3
P002	กลุ่มรักเชียงใหม่ 51 บุกบ้าน นายอรพล ตันติมนตรี สมาชิกสภาผู้แทนราษฎร จังหวัดเชียงใหม่ Rak Chiang Mai people enter into Mr. Tantimontree a member of the House of Representatives' accommodation.	2
P003	อัยการปิดไม่ทราบข่าว “ทักษิณ” อยู่เอเชีย Thai prosecutor refuse the case of Thaksin in Asia country now.	2
P004	กต.ชี้แลกตัว 'ทักษิณ' กับนักโทษยูเออีไม่ได้ Thailand's Ministry of Foreign Affairs had declared unable to trade fugitive ex-prime minister Thaksin with UAE.	2
P005	ผบ.สส.กล่าวถึงกรณีนายทหารไปเยี่ยมทักษิณ มีสิทธิทำได้ Army commander in chief said that military officers visited Taksin.	2
P006	กบข. แต่งดำงูกทำเนียบ Thai Civil Servant's Pension and Retirement Benefit Fund officers in black dress attack to the Government House.	2
P007	นายกฯยันไทยค้าน “เขาพระวิหาร” เป็นมรดกโลก Prime Minister Abhisit Vejjajiva confront Thai still defend over Praviharn to be World Heritage.	2

Table 4.1 (continued)

Headline ID	Headline	Number of documents
P008	“สุเทพ” เชื่อปัญหาชายแดนไทย-กัมพูชาคลี่คลาย Thai Deputy Prime Minister Suthep Thaugsuban confident Cambodia visit could reduce border tension.	2
P009	รมว.กท. ลงพื้นที่ภูเก็ต The minister of Thailand’s Ministry of Defence to inspect at Phuket.	2
P010	“สุเทพ” ไม่นักใจผลกกด. ตัดสินถือหุ่นปปช. Suthep Thaugsuban not serious with the judge of election commission.	2
G001	ฐานรากพระธาตุดอยสุเทพร้าว Phrathat Doi Suthep’s base was crack.	3
G002	ไฟไหม้ห้างบิ๊กซีอยุธยา-คาดไฟฟ้าลัดวงจร Fire at Big C superstore Ayutthaya- guess faulty electrical wiring is a cause.	3
G003	สธ.ยืนยันไทยตายเพราะหวัด 2009 แล้ว 2 คน Thailand’s Ministry of Public Health confirmed 2 deaths from the H1N1 flu virus.	2
G004	เม็กซิโกระบุการระบาดหวัดใหญ่พันธุ์ใหม่ทุเลาลงแล้ว Mexican confirmed the decrease of A (H1N1) influenza virus infections.	2
G005	ยอดติดเชื้อหวัด2009ทั่วโลกทะลุ 12,000 คน แล้ว The H1N1 swine flu has infected more than 12000 people in around the world.	2

Table 4.1 (continued)

Headline ID	Headline	Number of documents
G006	แผ่นดินไหว 5.7 ริคเตอร์ ที่เม็กซิโก Mexico Hit With 5.7 Magnitude Earthquake.	2
G007	อดีตประธานาธิบดีเกาหลีใต้ ตกเขาดับ Former South Korean President has died after falling into a ravine.	3
G008	คุณหญิงพจมานเปลี่ยนนามสกุลเป็นดามาพงศ์ Khunying Pojaman has changed her lastname to Damapong.	2
G009	องค์การอาหารและยาเตือนยาแก้หวัดปลอมระบาด Food and Drug Administration warning to the spread of imitation drug.	3
G010	แถลงลาออกทั้งน้ำตา "นักการเมืองสาวมาเลย์" หลังรูปหวีหลุด ว่อนเน็ต-มือถือ ลั่นไม่อับอาย Malaysian politician Elizabeth Wong tendered her resignation after pictures of her in the nude were circulated	2
E001	ตูน บอดี้สแลม ได้เข้าพิธีอุปสมบท ที่ จ.ขอนแก่น Mr.Toon of Bodyslam go into priesthood at Khonkhan.	2
E002	วิกตอเรียตามรอยเบ็คแฮมถ่ายแบบชุดชั้นในอาร์มานี่ Following in her husband, David Beckham's footsteps, Victoria Beckham will model of Armani underwear.	2
E003	แบม-โบ๊ท เข้าพิธีสมรสกันแล้ว Bam Janista Lewchalermwong get wedding ceremony with businessman Boat Buttrarat Charoonsamit.	2

Table 4.1 (continued)

Headline ID	Headline	Number of documents
E004	ราชาเพลงป๊อป ‘ไมเคิล แจ็กสัน’ เสียชีวิต Pop icon Michael Jackson dead.	3
E005	มาริสา แอนนิต้า ซึ่งรถชนเจ็บเล็กน้อย Top Thai model and actress Marisa Anita in a car accident.	2
E006	นุ้ย ตัด เขา เขา ไม่ขาด ยังมีคุยแต่โนรีเทิร์น Nui and ChaoChao restore good relations.	2
E007	เคน ยิ้มร่า หน้อยท้องลูกคนที่ 2 Noi Bussakorn is pregnant again with her second child.	2
E008	แซมมี่ ปิดข่าวสวิงกิ้ง ยันโสดสนิทเพราะแม่หวง Sammie Bunthita refuse the news ‘swinging’.	2
E009	จำคุก 6 เดือน “จู จี ฮุน” พระเอกดังเกาหลี The court sentence out to Ju Ji Hoon ,6 months in prison.	2
E010	โอ-วรุธ ยอมรับเมียท้อง 8 เดือน Waruth Worathum accepted that he having a wife.	2

4.2 Evaluation and Result

The goal of automatic text summarization evaluation is to measure the qualities of an algorithm or a system, which is to provide a satisfied compact summary. Generally, the research in automatic text summarization system has two main methods of evaluation including intrinsic evaluation and extrinsic evaluation.

The intrinsic evaluation is used to evaluate the system performance with a standard result, which is pre-defined by the evaluators ((Hassel, 2004), (Fukusima & Okumura, 2000)). The extrinsic evaluation is used to evaluate the system in more complex setting. For this

evaluation, the system is qualified in terms of its utility with respect to the overall task of the complex system or the user ((Hassel, 2004), (Fukushima & Okumura, 2000)).

The evaluation method used in this thesis is the extrinsic evaluation method because there is no standard dataset used to evaluate a system of multiple documents for Thai language. This evaluation assesses the quality of the summary by using an external criterion of the human judge. This evaluation method performed based on the satisfaction of the readers (Lynch et al., 2002) on four different criteria including grammar, coherence, clarity and conciseness. Descriptions of all criteria are as follows:

1. Grammar; regarding the correction of grammatical structure in terms of having completed in sentence structure (subject, verb, and object).
2. Coherence; regarding the content of summary expressed and organized in an effective or meaningful way.
3. Clarity; regarding the summary that can be representing the understandable concepts or main ideas.
4. Conciseness; regarding the appropriate length of summary.

The summaries from 30 headlines are summarized manually by the human and automatically by the system. In this thesis, all summaries are evaluated by five evaluators regarding the evaluation guidelines in Appendix A. These groups of students from different backgrounds including law, information technology and management are the summarizers for this thesis. Table 4.2 shows details of the evaluation task. The results of all human generating the summaries are shown in Appendix B.

Table 4.2 Details of Evaluation Task

	Amount	Unit
News headlines	30	Headlines
Human summarizer	3	People
Evaluator	5	People
Summary produced by human	90	Sentences
Summary produced by system	30	Sentences

To assess the system, five evaluators judged the quality of both human generated and system generated summaries with three different levels for each criterion. The average and total scores of each human generated summary for every criterion are shown in Table 4.3. Table 4.4 shows the average total scores of the system generated summary. Table 4.5 compares the average scores of all criteria between the system and human generated summaries.

Table 4.3 Average Grammar, Coherence, Clarity and Conciseness scores of Overall Human Generated Summary.

	Grammar	Coherence	Clarity	Conciseness	Total
Human 1	2.3	2.2	2.2	2.2	8.9
Human 2	2.6	2.3	2.4	2.3	9.6
Human 3	2.3	2.3	2.3	2.2	9.1
Average	2.4	2.2	2.3	2.2	9.1

Human 1 is the student from school of Law.

Human 2 is the student from school of Information Technology.

Human 1 is the student from school of Liberal Art.

Table 4.3 shows that the human generated summaries have the highest scores in the grammar criteria on average.

Table 4.4 Average Grammar, Coherence, Clarity and Conciseness Scores of System Generated Summary.

Grammar	Coherence	Clarity	Conciseness
2.4	2.2	2.2	2.2

Table 4.5 Comparison of Grammar, Coherence, Clarity and Conciseness Scores of System Generated Summary and Human Generated Summary.

	Grammar	Coherence	Clarity	Conciseness	Total
Human	2.4 (80%)	2.2 (73%)	2.3 (76%)	2.2 (73%)	9.1 (75%)
System	2.4 (80%)	2.2 (73%)	2.2 (73%)	2.2 (73%)	9 (75%)

Table 4.5 shows that the system generated summary only has less scores of the clarity than the human generated summary.

CHAPTER 5

DISCUSSION

This chapter discusses the evaluation results of both system generated and human generated summaries. Beside details of the evaluations, the discussions of all results are also explained here.

5.1 Analysis of Evaluation Results

To analyze the evaluation results of the system generated summary, each criterion is discussed as follows.

5.1.1 Grammar Score

As mentioned before, the grammar score relies on the correction of grammatical structure in term of having complete sentence structure (subject, verb, and object). The system generated summary obtained the highest score of grammar criteria. The highest score shows that the system has the competency to generate high grammatical quality summaries. This result evidences that the interpretation template and the filling rules are workable.

In order to improve this score, the system needs more rules to cover against the ambiguity of a complex having a long input sentence because one sentence may consist of two or more clauses.

5.1.2 Conciseness Score

The conciseness score demonstrates the appropriate length of the summary to ensure that the summary has the appropriate length. This score is rather high satisfaction. This evidence is direct to the objective of the thesis that is to obtain a compact summary.

The system generated summaries have high conciseness score because the length is controlled by two strategies including identification of a common sentence which only the common content is considered among sentences and by the interpretation template that is predefined only relevant information to generate a summary.

5.1.3 Clarity Score

The clarity score represents a main idea of the summary in order that readers can easily understand the concept of the news. The system generated summary also gets rather high clarity score. It is shown that the main ideas of the sentences are clearly clarified. Therefore, to answer the basic questions in the interpretation template mentioned before can ensure the clarity of the summary.

5.1.4 Coherence Score

The coherence score represents that the content in the summary is expressed in an effective or meaningful way. It can achieve through syntactical features such as sentence logical structure. The coherence score of the system generated summary is also rather high which demonstrates the effective rules and interpretation template.

In order to improve the coherence quality, more prior knowledge is needed to construct rules that construct more elegant interpretation templates.

In addition to improve the overall performance of the system, more prior linguistic knowledge is required to construct rules for creating or filling the interpretation template. However the larger synonym database is required. Particularly, more evaluation criteria are also required to assess the system's performance more effectively in term of readers' satisfaction.

5.2 Comparison of System and Human Generated Summaries Quality

The evaluation results demonstrate that the scores of the human and the system generated summaries in all criteria are rather equal, although few differences appear in the clarity scores, which are 2.3 (76%) and 2.2 (73%) of the human generated summary and the system generated summary respectively. The system generated summary obtains the lower clarity score than this

of the human generated summary because some information has been missing in the system generated summary. The future work could be arranged to improve this score by changing the rules for filling the interpretation template more generally. For example, instead of only who acts what to whom, the rules should require more information such as when, where and how too. Additionally, another reason is because of the selection of the common information. As mentioned before in the proposed system, the method considers only one pair of the sentences overlapping among the group of the documents, which leads the possibility of missing some other main ideas. More common sentences would possibly improve the clarity score.



CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Conclusion

In the digital society, the Internet plays an important role for people's daily lives with regard to the consumption of the information over the Internet. There are a various kinds of information and services available in the Internet. Particularly, for online news, there is usually time consumption to explore all sources of the online news to consume the information which is usually reported in the same news event with different writing style.

This thesis has addressed the problem of the automatic summarization based on the statistical method especially for the summarization of the online news. The problem is that synonyms, acronyms, superfluous words, and jargon words generally cause the redundancy and incoherence problem for the automatic text summarization.

This thesis proposes the system for the automatic text summarization that aims to summary Thai online news from different sources. The system generates a compact summary based on the principle of summarization. The interpretation template is proposed to generate a compact summary by using the rules constructed from linguistic knowledge.

The system was tested with 66 news articles of 30 news headlines from different online sources. The summaries were generated manually and automatically. The summaries were generated manually by three different people with different backgrounds including law, Information Technology and management. The evaluation task is done by five evaluators with regard to four criteria of the summary including grammar, coherence, clarity and conciseness. The evaluation result has been shown that the system obtains 80 %, 73%, 73% and 73% of grammar, coherence, clarity and conciseness respectively. The result of both human generated

and system generated summaries are similar. Therefore, it can be concluded that the proposed system works similarly to the human. However, the system can improve higher quality by adding more prior knowledge about rules, enlarging the synonym database and also having more varieties of adjustable interpretation template.

6.2 Future Work and Suggestion

As knowledge based approach method, this thesis is quite complicated to implement and complete every process. Therefore, various aspects are still required to improve the efficiency of the future work. These requirements are explained as follows:

6.2.1 Improving the prior knowledge base: In order improve the satisfaction of the readers and produce more natural summaries, the rules should be improved.

6.2.2 Providing a variety of the interpretation template in term of writing style and length. This thesis considers only the main ideas among the same news headlines which provide only one summary length and writing style.

6.2.3 Implementing the full automatic system: In order to test the work of knowledge and Interpretation template, this thesis implements the prototype system that provides only the summarizing process which requires the correct input manually. Thus, the performance of the overall system still is not verified. As a result, the future work studying the full automatic system would help the system verification to be more appropriate.

6.2.4 Evaluating each process of the system: In order to verify the system, the evaluation of the performance in every process to improve any problematic aspect is required to finally reach the higher performance of the system.

REFERENCES



REFERENCES

- Afantenos,S., Karkaletsis,V., & Stamatopoulos,P. (2005). Summarization from medical documents: a survey. **Artificial intelligence in medicine**, (33), 157-177.
- ASTV manager online** (2008). Retrieved November 7, 2008, from <http://www.manager.co.th/>
- Ausdang Thangthai & Chuleerat Jaruskulchai. (2003). **Text Summarization using Singular Value Decomposition for Thai**. Paper presented at the The 7th National Computer Science and Engineering Conference, Chonburi, Thailand.
- Amit, S. (2001). Modern Information Retrieval : A Brief Overview. **Bulletin of the IEEE Computer Society Technical Committee on Data Engineering**, 24 (4), 35-42
- Bing, Q. (2005). Research on Multi-Document Summarization Based on Latent Semantic Indexing. **Journal of Harbin Institute of Technology**, 12(1), 91-94.
- Carter,M. (2008). **How to Write a Newspaper Article**. Retrieved 2008 December, 12 from http://newspaperjournalism.suite101.com/article.cfm/how_to_write_a_newspaper_article
- Chen,H., Kuo,J., Huang,S., Lin,C., & Wung,H. (2003). A Summarization System for Chinese News from Multiple Sources. **Journal of the American Society for Information Science and Technology**, 54, 1224–1236.
- Daily news** (2008). Retrieved November 7, 2008, from <http://www.dailynews.co.th/web/html/home/>.
- Dorr,B., Hovy, E., & Levin,L. (2004). Machine Translation: Interlingual Methods. In, Brown, Keith (Ed.), **Encyclopedia of Language and Linguistics** (2nd ed.). New York : Elsevier.

Euzenat, J. (2007). Semantic Precision and Recall for Ontology Alignment Evaluation.

In **Proceedings of the 20th international joint conference on artificial intelligence**
(pp. 348-353). Hyderabad, India.

Farzindar, A. & Lapalme, G. (2004). The Use of Thematic structure and Concept Identification
for Legal Text Summarization. In **Proceeding of the computational linguistics in the
North East** (pp. 67-71). Montréal, Québec, Canada.

Fukusima, T. & Okumura, M. (2000). Text Summarization Challenge: Text summarization
evaluation at NTCIR Workshop2. In **Proceedings of the second NTCIR sorkshop on
research in chinese & japanese text retrieval and text summarization**.
National Institute of Informatics, Tokyo, Japan.

Garcia,E. (2006). **Cosine Similarity and Term Weight Tutorial**.
Retrieved January 10, 2009, from <http://www.miislita.com/>.

Gillespie, M. (2007). **Basic Principles of News Writing**. Retrieved 2008 November, 16
from <http://blogs.setonhill.edu/Journalism/2007/11/chris-ulicne-test-handout-60.html>

Goldberg, A. (2007). **Automatic Summarization**. In CS838-1 Advance NLP.
Lecture Notes in Advance Natural Language Processing. Department of Computer
Science, University of Wisconsin-Madison.

Hassel, M. (2004). **Evaluation of Automatic Text Summarization - a practical
implementation**. Licentiate Thesis, Department of Numerical Analysis and Computer
Science, Royal Institute of Technology, Stockholm, Sweden.

Hovy, E., Gerber,L.,Hermjakob,U.,Junk,M., & Lin,C. (2000). Question Answering
in Webclopedia. In **Proceedings of the ninth text retrieval conference (TREC-9)**
(pp.655-664). Gaithersburg, Maryland.

Kom Chad Luek. (2008). Retrieved November 7, 2008, from <http://www.komchadluek.net/>

Kumchai Thonglor (1997). **หลักภาษาไทย**. Bangkok: Ruamsam.

Liddy, E.D. (2001). Natural Language Processing. In **Encyclopedia of Library and Information Science** (2nd ed.). New York : Marcel Decker.

Lynch,S., Nesbitt, M., & Duke,S. (2002). **Inside Satisfaction: what it means, How to increase it**. Evanston, Illinois:The Media Management Center at Northwestern University.

Mann,B. (1999). **An Introduction to Rhetorical Structure Theory (RST)**.

Retrieved January 24, 2009, from <http://www.sfu.ca/rst/>

Matichon Public Company Limited. (2008). **Matichon online**.

Retrieved November 7, 2008, from <http://www.matichon.co.th/>

Mooney, R.J. (2009). **Natural Language processing: Syntactic parsing**.

[presentation slide]. Texas: University of Texas at Austin.

Ou, S., Khoo, S.G, C., & Goh, D. (2005). Constructing a taxonomy to support multi-document summarization of dissertation abstracts. **Journal of Zhejiang University SCIENCE**, 6A (11), 1258-1267.

Pensiri Leetrakul & Sukree Sinthupinyo. (2008). Hierarchical Thai Text Summarization.

In **Proceedings of international joint conference on computer science and software engineering**.(pp. 294-301). Kanchanaburi, Thailand.

Radev, D. (2000). A common theory of information fusion from multiple text sources step one: cross-document structure. In **Proceedings of the 1st SIGdial workshop on discourse and dialogue**, (pp. 74 – 83), Hong Kong.

- Radev,D.,Winkel,A., & Topper,M. (2002). Multi Document Centroid-based Text Summarization. In **Proceedings of the ACL-02 demonstrations session** (pp. 112-113). Philadelphia, Pennsylvania.
- Radev, D., Jing,H., Stys',Magorzata., & Tam, D. (2003). Centroid-based summarization of multiple documents. **Information processing and management**,(40), 919–938.
- Radev, D., Otterbacher, J., Winkle, A. & Blair,G.S. (2005). NewsInEssence : Summarizing online news topics. **Communications of the ACM**, **48**, 95-98.
- Ratree Kumwong. (2008). **การย่อความ**. Retrieved December 20, 2008, from http://www.ras.ac.th/link15/web-dang/การย่อความ_files/frame.htm.
- Thana Sukvaree, Asanee Kawtrakul & Caelen, J. (2007). **Thai Text Coherence Structuring with Coordinating and Subordinating Relations for Text Summarization**. Lecture Notes in Artificial Intelligence (LNAI), (pp.453-466), Berlin, Heidelberg: Springer-Verlag.

APPENDIXES



APPENDIX A

EVALUATION GUIDELINE

Table A.1 A Guideline of Evaluation

Criterion
<p>1. Grammar; Does a summary follow the rule of grammar? (Complete with grammatical structure S+V+O)</p> <p>3 point- The summary thoroughly presents grammatical.</p> <p>2 point- The summary presents some grammatical.</p> <p>1 point- The summary hardly presents any grammatical.</p> <p>2. Coherence; Is the content of a summary expressed and organized in an effective way?</p> <p>3 point- Summary is good organized and coherent.</p> <p>2 point- Summary is organized to discernable outline.</p> <p>1 point- Summary is not coherent or organized in a logical manner.</p> <p>3. Clarity; Does a summary represent the main idea?</p> <p>3 point- The summary demonstrates good understanding of concepts.</p> <p>2 point- The summary demonstrates some understanding of concepts.</p> <p>1 point- The summary hardly demonstrates the understanding of concepts.</p> <p>4. Conciseness; Is the length of summary is suitable?</p> <p>3 point- Summary's length is properly.</p> <p>2 point- Summary's length is too short.</p> <p>1 point- Summary's length is too long.</p>

APPENDIX B

SUMMARY RESULTS

Table B.1 Summaries Result of Human 1.

Summary ID	Summary
H1-E001	13 พฤษภาคม 52 ตูนบอดีแสดม ได้เข้าพิธีอุปสมบทที่วัดบ้านทง จ.ขอนแก่น โดยได้รับฉายาทางธรรมว่า “วราสโย” กำหนดบวช 17 วัน ลาสิกขาบท วันที่ 8 มิ.ย.
H1-E002	วิกตอเรีย แบคแฮม เป็นนางแบบชุดชั้นในคนล่าสุดให้กับ Emporium Armani ในคอลเลกชั่นสปริง/ซัมเมอร์ โดยได้รับค่าเหนื่อย ไปกว่า 600 ล้านบาท
H1-E003	28 มิ.ย. ห้องรอยัลบอลรูม โรงแรมแมนดาริน โอเรียนเต็ล แบบโบ๊ทได้ฉลองสมรส โดยมีแขกทั้งนักการเมืองนักธุรกิจและนักแสดง โดยมีพณฯท่าน บรรหาร ศิลปอาชา เป็นประธานในพิธี
H1-E004	ไมเคิล แจ็กสัน อดีตราชาเพลงป๊อปวัย 50 ปีเสียชีวิต ที่โรงพยาบาลศูนย์การแพทย์ UCLA ในลอสแอนเจลิส เมื่อวันพฤหัสบดีที่ 25 มิ.ย. จากอาการหัวใจวาย
H1-E005	29 มิ.ย. 52 นักแสดงและนางแบบ สาว – มาริสา แอนนิต้า ได้ประสบอุบัติเหตุ รถยนต์พลิกคว่ำบนถนนมอเตอร์เวย์ ซึ่งมุ่งหน้าไปพัทยา โชคดีคาดเข็มขัดนิรภัย เลยไม่เป็นไรมาก
H1-E006	นุ้ย กลับมาคืนดีกับเขาแล้ว เผยยังไม่คบเป็นแฟน คู่กันเป็นพี่เป็นน้องก่อน ไม่อยากกลับมาเป็นแบบเดิมอีก
H1-E007	หน้อย บุษกร ท้องลูกคนที่ 2 ได้ 6 สัปดาห์แล้ว เกณฐ์ได้ลูกสาว เผยไม่ตื่นเต้นเท่าท้องแรกเพราะมีประสบการณ์แล้ว

Table B.1 (continued)

Summary	Summary
ID	
H1-E008	แซมมีปัดข้าวสรวงถึงอ้วน รังสิต และเติ้ล ธนพล แคเพื่อนร่วมงาน ชันโสศนิต เพราะอยากเรียนอย่างเดียวและคุณแม่ก็ห่วงด้วย
H1-E009	วันที่ 23 มิ.ย. จู จิสุน ได้เดินทางไปขึ้นศาลเพื่อรับฟังการพิพากษาคดี โดยศาล มีคำสั่งจำคุก 6 เดือน รอลงอาญา 1 ปี บำเพ็ญประโยชน์รับใช้สังคมอีก 120 ชม. และจ่ายค่าปรับอีก 360,000 วอน
H1-E010	โอ วรุฒ เพย มีเมียแล้วแถมท้องรอกตลอด เดือน ก.ค. นี้ ตั้งใจจดทะเบียนสมรส หลังคลอดแล้ว เพยได้ลูกชาย ตั้งชื่อ “เอร์บัส”
H1-P001	เมื่อวันที่ 15 ม.ค. 52 ม.ร.ว. สุขุมพันธุ์ บริพัตร ผู้ว่าราชการ กทม. เข้ารับหนังสือ รับรองการเป็นผู้ว่าราชการกรุงเทพมหานคร ที่สำนักคณะกรรมการการเลือกตั้ง ประจำ กทม. ณ อาคารเมทัลเลียนทาวเวอร์ ย่านชิดหลักธรรมมาภิบาล
H1-P002	เมื่อวันที่ 20 ธันวาคม น.ส. กัญญากัณ มณีจักร แกนนำกลุ่มรักเชียงใหม่ 51 พร้อม กลุ่มเสื้อแดงได้เดินทางมาชุมนุมประท้วงหน้า บจก. บัณฑิตเรียฟอสเฟต ตราจักรทอง เพื่อประท้วงนายบรรพต ตันติมนตรี สส.พรรคเพื่อแผ่นดิน ที่สนับสนุนให้นายอภิสิทธิ์ เวชชาชีวะ เป็นนายกรัฐมนตรี
H1-P003	28 มิ.ย. 52 นายศิริศักดิ์ ดิยะพรร อธิบดีอัยการ ตปท. กล่าวว่า อัยการยังไม่เคย ได้รับการยืนยันจาก กท. การ ตปท. และ จนท.ตร. ซึ่งมีหน้าที่ตามหาที่อยู่ของ พตท. ทักษิณ และ กท. การ ตปท.สามารถยืนยันได้เฉพาะที่อยู่ใน ปท.สหรัฐอเมริกาบริเบมิเรสต์ และนิการากัวเท่านั้น
H1-P004	วันที่ 22 พ.ค. 52 นายชฤต จรุงวัฒน์ อธิบดีกรมสารนิเทศและโฆษกกระทรวงการ ต่างประเทศ กล่าวถึงความเป็นไปได้ในการแลกตัวผู้ต้องหาระหว่างไทยกับ สหรัฐอาหรับเอมิเรสต์ ระหว่างผู้ต้องหาอายุเออีที่ตำรวจไทยควบคุมตัวไว้นั้นเป็น ความผิดฐานถือโก่ง ซึ่งต่างกับกรณีของ พตท. ทักษิณ ที่เป็นคดีทุจริต

Table B.1 (continued)

Summary ID	Summary
H1-P005	21 ก.พ. 52 พล.อ. ทรงกิตติ จักกาบาตร์ ผบ.สส. ในฐานะนักเรียนเตรียมทหารรุ่น 10 กล่าวถึงกรณี ส.ส. พรรคเพื่อไทย และนายทหารระดับสูง เดินทางไปเยี่ยม พตท. ทักษิณ ที่ฮ่องกง ว่าหากทำตามระเบียบของทางราชการก็สามารถทำได้
H1-P006	วันที่ 29 มิ.ย. 52 เครือข่ายสมาชิกกองทุนบำเหน็จบำนาญข้าราชการแห่งประเทศไทย พร้อมด้วยสมาชิกกว่า 500 คน สวมชุดดำ เคลื่อนขบวนไปชุมนุมหน้า ทำเนียบรัฐบาล ยื่นหนังสือถึงนายกรัฐมนตรี เพื่อยื่นข้อเรียกร้อง 6 ข้อ หลัง คณะกรรมการบริการของ กบข. ได้บริหารงานผิดพลาด ส่งผลให้ประสบ ภาวะขาดทุน
H1-P007	29 มิ.ย. ที่ทำเนียบรัฐบาล นายก รมต. เปิดเผยว่า ได้ยืนยันตรงกันกับนาย ก รมต. กัมพูชา ว่าปัญหาเรื่องเขาพระวิหารจะต้องไม่เป็นปัญหาที่ส่งผลกระทบต่อ ความร่วมมือในปัจจุบันและอนาคต
H1-P008	27 มิ.ย. 52 นายสุเทพ เทือกสุบรรณ รองนายกรัฐมนตรี กล่าวถึงการเดินทางเยือน กัมพูชา โดยยืนยันว่าจะได้พบกับนายกัมพูชา จะเป็นเจ้าภาพเลี้ยงอาหาร และจะ ได้เจรจากับทางกัมพูชา และหลังจากเจรจาสถานการณ์ความสัมพันธ์ของทั้ง 2 ประเทศจะดีขึ้น
H1-P009	29 มิ.ย. 52 เจ้าหน้าที่ฝ่ายความมั่นคงไปตรวจความพร้อมในการประชุมอาเซียน ชัมมิตที่ภูเก็ต โดยมีการไปตรวจดูสถานที่จัดการประชุมและเตรียมความพร้อม มาตรการรักษาความปลอดภัย ทั้งยืนยันว่าไม่กังวลต่อการเคลื่อนไหวของกลุ่ม เสือแดง
H1-P010	30 มิ.ย. 52 นายสุเทพ เทือกสุบรรณ รองนายกรัฐมนตรี. ฝ่ายมั่นคง กล่าวว่า ไม่หนักใจ ในกรณีเรื่องคำวินิจฉัย กกต. กรณี สส. ถิฮุ่น แล้วถ้าศาลตัดสินอย่างไรก็ดำเนินไป ตามขั้นตอน

Table B.1 (continued)

Summary ID	Summary
H1-G001	เมื่อวันที่ 28 มิ.ย. 52 ผอ.ศิลปากร ชม. เปิดเผยว่า ตรวจพบรอยร้าวของ พระธาตุคอกยสุเทพ โดยเกรงว่าหากเกิดฝนตกหนัก หรือถ้ามีแผ่นดินไหวเกิน 5 ริคเตอร์อาจทำให้โครงสร้างทรุดทำให้สิ่งองค์พระธาตุพังทลายลงมาได้
H1-G002	เมื่อวันที่ 28 มิถุนายน เกิดเหตุเพลิงไหม้ที่บึงชี จังหวัดอยุธยา สาเหตุคาดว่าเกิด จากไฟฟ้าลัดวงจร ที่เกิดจากเครื่องปั่น ไปที่ติดตั้งติดกับห้องฉายหนังที่ 3 ซัด
H1-G003	วันที่ 27 มิ.ย. 52 กระทรวงสาธารณสุขเปิดเผยว่าพบคนไทย เสียชีวิตจากไข้หวัด 2009 2 ราย เป็นหญิงหนึ่งคน และชาย 1 คน จากผู้ป่วยติดเชื้อทั้งหมด 1209 ราย
H1-G004	วันที่ 4 พ.ค. รมต. กระทรวงสาธารณสุขของเม็กซิโก ระบุสถานการณ์ไข้หวัดใหญ่ สายพันธุ์ใหม่ 2009 ได้ลดความรุนแรงลงแล้ว แต่ WHO เตือนยังไม่ควรไว้วางใจ เพราะเชื้ออาจกลับมาระบาดได้อีก
H1-G005	ล่าสุดยอดผู้ป่วยจากไข้หวัดใหญ่สายพันธุ์ใหม่ 2009 ทั่วโลกพุ่งสูงถึง 12,000 คน แล้ว นำโดยสหรัฐฯ ขณะที่รัฐเซียพบผู้ติดเชื้อรายแรกเป็นชายที่เพิ่งเดินทาง กลับมาจากสหรัฐฯ
H1-G006	เกิดแผ่นดินไหวในเขตเมืองหลวงของเม็กซิโก เมื่อเวลา 14.24 วันที่ 22 พฤษภาคม วัดแรงสั่นสะเทือนได้ 5.7 ริคเตอร์ นับเป็นครั้งที่สองที่เกิด แผ่นดินไหวในเวลาไม่ถึง 1 เดือนในเขตกรุงเม็กซิโกซิตี
H1-G007	เมื่อวันที่ 23 พฤษภาคม อดีตประธานาธิบดีโนห์ มู เฮียน เสียชีวิตแล้วโดยการตาย อาจเป็นอุบัติเหตุหรือการฆ่าตัวตาย เนื่องจากกำลังถูกสอบสวนในคดีรับสินบน ขณะดำรงตำแหน่ง
H1-G008	วันจันทร์ที่ 8 ธันวาคม เวลา 8.00 น. คุณหญิงพจมาน ชินวัตร เปลี่ยนนามสกุลเป็น “ดามาพงศ์” พร้อมทำบัตรประชาชนใหม่ ที่สำนักงานเขตดุสิต หลังทำเรื่องหย่า กับ พตท.ทักษิณ ชินวัตร

Table B.1 (continued)

Summary	Summary
ID	
H1-G009	ยาแก้หวัดปลมระบบด แอบขายอ้างมีส่วนประกอบสตูโดเฟร็ดรีน-ไทรโพลิดีน เพื่อใช้ในทางที่ผิดจนก่อให้เกิดปัญหาต่อภาพลักษณ์ของประเทศไทย และยังส่งผลกระทบต่อความปลอดภัยต่อผู้บริโภคด้วย ซึ่งสารดังกล่าวนำไปสกัดเอาสารสตูโดเฟร็ดรีน สารตั้งต้นยาบ้า ผู้ผลิต-ผู้ขายมีโทษปรับ –ติดคุก
H1-G010	นางสาวอลิชาเบธ หว่อง สมาชิกสภาผู้แทนราษฎรพรรคเคอาดิลัน ประกาศขอยุติการทำหน้าที่ทางการเมือง หลังจากถ่ายภาพเปลือยของเรอนอนอยู่ในห้องนอนถูกมัดมือนำไปเผยแพร่ตามสื่อต่างๆ



Table B.2 Summaries Result of Human 2.

Summary ID	Summary
H2-E001	ตูนบอดี้สแลม บวชที่วัดเกาะทอง จังหวัดขอนแก่น โดยมีเพื่อนศิลปินมาร่วมงานจำนวนมาก
H2-E002	วิคตอเรีย แบคแฮม เป็นนางแบบโฆษณาชุดชั้นในให้อาร์มานี ในคอลเล็กชั่นสริงซัมเมอร์
H2-E003	แบม จมิสตา ลีวเฉลิมวงศ์ เข้าพิธีแต่งงานแล้วกับ โบ๊ท บุตรรัศย์ จรูญสมิทธิ์ ที่โรงแรมแมนดารินโอเรียนเต็ล โดยมีแขกมากมายทั้งนักแสดง นักการเมือง และนักธุรกิจ
H2-E004	ไมเคิล แจ็กสัน ราชาเพลงป๊อป เสียชีวิตแล้ว ด้วยวัย 50 ปี โดยแพทย์สันนิษฐานสาเหตุการเสียชีวิตเบื้องต้นว่า เกิดจากหัวใจวาย
H2-E005	ดารานางแบบ สามีริสา แอนนิต้า ประสบอุบัติเหตุรถยนต์พลิกคว่ำบนถนนมอเตอร์เวย์ แถว อ.สตึก จ.ชลบุรี ขณะไปถ่ายละคร
H2-E006	นุ้ย-เกศรินทร์ เอกธวัชกุล กลับมาคืนดีกับดีเจ “เซาเซา” แล้ว แต่ยัง ไม่กล้าคบแบบแฟน
H2-E007	เคน ชีรเดช วงศ์พัฑฒ์เปิดเผยว่า หน้อย บุษกร กำลังตั้งครรภ์ลูกคนที่สอง
H2-E008	แซมมี่ ปัทมทิศา ภูวิจารณ์ เคา์เวลล์ ออกมาปฏิเสธข่าว กวาง 2 หนุม "อ้วน" รังสิต ศิรนานนท์ และ "เต๋อ" ธนพล นิมิตสุข ขึ้นคอนโด
H2-E009	จู จิฮุน พระเอกดังจากละครเจ้าหญิงวุ่นวายกับเจ้าชายเย็นชา เดินทางไปฟังการพิจารณาคดี ข้อหาเสพติด และยาเสพติด โดยศาลมีคำสั่งจำคุก จิฮุนเป็นเวลา 6 เดือน
H2-E010	โอ-วรุฒ วรธรรม ออกมายอมรับว่าตนมีภรรยาแล้ว และกำลังจะมีลูก
H2-P001	ม.ร.ว.สุขุมพันธุ์ บริพัตร ผู้ว่าราชการกรุงเทพมหานคร เข้ารับหนังสือรับรองตำแหน่ง ผู้ว่าฯ กทม.ที่สำนักงานคณะกรรมการการเลือกตั้งประจำกรุงเทพมหานคร อาคารเมทลเลียน ทาวเวอร์

Table B.2 (continued)

Summary ID	Summary
H2-P002	เสื่อแดง นำโดยนางสาวกัญญากัก มณีจักร เดินทางประท้วงกันที่ บจก. ปุ๋ยอินทรีย์ฟอสเฟต ตราฉัตรทอง ซึ่งเป็นบ้านพักของนายรพล ดันติมนตรี ส.ส.พรรคเพื่อแผ่นดิน
H2-P003	นายศิริศักดิ์ ดิยะพรรณ อธิบดีอัยการฝ่ายต่างประเทศ ออกมาปฏิเสธว่าไม่เคยได้รับการยืนยันว่าตอนนี้ พ.ต.ท.ทักษิณ ชินวัตร อดีตนายกรัฐมนตรี ผู้ต้องคำพิพากษาศาลฎีกาแผนกคดีอาญาของผู้ดำรงตำแหน่งทางการเมืองคดีทุจริตซื้อขายที่ดินรัชดาภิเษก กลับมาอยู่ในประเทศแถวเอเชีย
H2-P004	นายธฤต จรุงวัฒน์ อธิบดีกรมสารนิเทศและโฆษกกระทรวงการต่างประเทศ กล่าวถึงเรื่องความเป็นไปได้ในการแลกตัวผู้ต้องหา กับประเทศสหรัฐอเมริกาสำหรับอเมริกา
H2-P005	พล.อ.ทรงกิตติ จักกาบาตร์ ผู้บัญชาการทหารสูงสุด กล่าวถึง กรณี ส.ส.พรรคเพื่อไทย (พท.) และนายทหารระดับสูง รุ่นเดียวกัน เดินทางไปเยี่ยม พ.ต.ท.ทักษิณ ที่ฮ่องกง ว่าหากทำตามระเบียบของทางราชการ สามารถไปได้
H2-P006	นายวิศว์ อัครสันติสุขกุล ประธานเครือข่ายสมาชิกกองทุนบำเหน็จบำนาญข้าราชการแห่งประเทศไทย นำข้าราชการที่เป็นสมาชิก กบข. ใส่ชุดดำมาประท้วงที่หน้าทำเนียบรัฐบาล
H2-P007	นายอภิสิทธิ์ เวชชาชีวะ นายกรัฐมนตรี เผยว่า รัฐบาลยังยืนยันจุดยืนในการคัดค้านเขาพระวิหารขึ้นเป็นมรดกโลก
H2-P008	นายสุเทพ เทือกสุบรรณ รองนายกรัฐมนตรี กล่าวถึง การเดินทางไปกัมพูชาว่าจะได้พบกับ สมเด็จฮุนเซน นายกรัฐมนตรีกัมพูชา และ จะเป็นเจ้าภาพเลี้ยงอาหารกลางวัน นอกจากนี้ยังเชื่อว่าทางกัมพูชาจะเข้าใจ และทำให้สถานการณ์ตามแนวชายแดนของทั้งสองประเทศ ลดความตึงเครียดลงได้
H2-P009	พล.ต.อ.ประวิตร วงษ์สุวรรณ รัฐมนตรีว่าการกระทรวงกลาโหม ลงพื้นที่เพื่อตรวจสอบความพร้อมของการจัดการประชุมรัฐมนตรีอาเซียนที่ จ.ภูเก็ต

Table B.2 (continued)

Summary ID	Summary
H2-P010	นายสุเทพ เทือกสุบรรณ รองนายกรัฐมนตรีฝ่ายความมั่นคง กล่าวถึงการ วินิจฉัยกรณีถือครองหุ้น ว่า ตนก็ไม่แน่ใจ วินิจฉัยว่าอย่างไรก็ว่าไปตามนั้น และไปสู้คดีที่ศาล
H2-G001	นายสหัสวัฒน์ แนนหนา ผอ.สำนักศิลปากรที่ 8 เชียงใหม่ ออกมาเปิดเผยว่าพบรอย ร้าวที่ฐานรากชั้นที่ 2 ของพระธาตุดอยสุเทพเพิ่ม คาดว่าหากเกิดฝนตกหนัก หรือแผ่นดินไหวเกิน 5 ริกเตอร์ อาจทำให้โครงสร้างทรุดจากดินเหลวทำให้ ดิ่งองค์พระธาตุพังทลายลงมาได้
H2-G002	เกิดเหตุเพลิงไหม้ขึ้นที่ห้างบิ๊กซีอยุธยา โดย พ.ต.ท.สมาน ศรีอรุณ สารวัตรเวร สภ.บางปะอิน จ.พระนครศรีอยุธยา รับแจ้งเหตุ ก่อนประสานให้รถดับเพลิง 5 คัน ไปยังที่เกิดเหตุ
H2-G003	กระทรวงสาธารณสุขเปิดเผย ปัจจุบันมีผู้ป่วยติดเชื้อหวัดใหญ่ 2009 ทั้งหมด 1,209 ราย ในจำนวนนี้หายเป็นปกติแล้ว 1,190 ราย มีผู้รักษาตัวอยู่ในโรงพยาบาล 16 ราย และ พบคนไทยเสียชีวิตจากไข้หวัดใหญ่สายพันธุ์ใหม่ 2009 เป็นของไทย แล้ว 2 คน เป็นผู้หญิงใน กทม.วัย 40 ปี 1 ราย และผู้ชายวัย 42 ปี ที่ จังหวัดชลบุรี
H2-G004	รัฐมนตรีสาธารณสุขเม็กซิโก ออกมาเผยว่า สถานการณ์การแพร่ระบาดของโรค ไข้หวัดใหญ่สายพันธุ์ใหม่ 2009 ในเม็กซิโกลดความรุนแรงลงแล้ว
H2-G005	เชื้อไวรัส H1N1 ชนิดเอ สาเหตุของไข้หวัดใหญ่สายพันธุ์ใหม่ 2009 ยัง แพร่กระจายไปทั่วโลกใน 46 ประเทศ โดยล่าสุดตัวเลขพุ่งเป็น 12,000 คนแล้ว
H2-G006	เกิดแผ่นดินไหว 5.7 ริกเตอร์ ใจกลางประเทศเม็กซิโก รวมถึงกรุงเม็กซิโกซิตี
H2-G007	อดีตประธานาธิบดีโนห์ มู เฮียน เสียชีวิตแล้ว หลังจากตกลงไปในหุบเขาขณะปีน เขาแห่งหนึ่ง

Table B.2 (continued)

Summary	Summary
ID	
H2-G008	คุณหญิงพจมาน ชินวัตร อดีตภริยา พ.ต.ท.ทักษิณ ชินวัตร เดินทางมายังสำนักงานเขตดุสิต เพื่อเปลี่ยนนามสกุล เป็น ดามาพงศ์ หลังจากทำเรื่องหย่ากับ พ.ต.ท.ทักษิณ เรียบร้อยแล้ว
H2-G009	นพ.พิพัฒน์ ยิ่งเสรี เลขาธิการคณะกรรมการอาหารและยา ออกมาเปิดเผยว่า ขณะนี้มีการลักลอบขายยาปลอมให้กับสถานพยาบาลบางแห่งเกิดขึ้น
H2-G010	น.ส.อลิษาเบธ หว่อง นักการเมืองหญิงมาเลย์ ประกาศลาออกหลังมีรูปภาพเปลือย แพร่ตามสื่อต่างๆ

Table B.3 Summaries Result of Human 3.

Summary	Summary
ID	
H3-E001	ตูน บอดี้สแลมบวชที่ จ.ขอนแก่น ได้ชื่อทางธรรมว่า ปาสโย มีกำหนดบวช 17 วัน
H3-E002	วิกตอเรีย ได้ถ่ายภาพโฆษณาในชุดชั้นในของ Emporium Armani หลังจาก เดวิด เบ็คแฮม ได้ถ่ายโฆษณาทางเกงในก่อนแล้ว
H3-E003	แบบ จณิดา ฉลองสมรสกับโบ๊ท ที่ห้องรอยังบอลรูม โรงแรมแมนดาริน เมื่อวันที่ 28 มิ.ย. ที่ผ่านมา
H3-E004	ไมเคิล แจ็กสัน เสียชีวิตอย่างกะทันหัน หลังจากเกิดอาการหัวใจวายเฉียบพลัน
H3-E005	มาริสา แอนนิต้า ประสบอุบัติเหตุชนกำแพงข้างถนน ได้รับบาดเจ็บเล็กน้อย บนถนนมอเตอร์เวย์ ช่วงก่อนออกจากบ้านบึง
H3-E006	นุ้ย เกรริน หลังเลิกกับ เขาเขา ก็กลับไปปริเทรินกัน แต่ยังไม่กล้าคบแบบแฟนเหมือนเดิม แต่ก็ยังคุยกันอยู่ในฐานะเพื่อน
H3-E007	เคน หลังจากได้ลูกสาวคนแรกแล้วก็ลุ้นได้ลูกคนที่สองทันที
H3-E008	หลังจากข่าวของแซมมี่ ออกมาว่าสวิงกับ 2 หนุ่ม อ้วน รังสิต และเตี๊ยะ ธนพล เจ้าตัวเขาเพราะว่ายังไม่เคยมีแฟนและไปไหนกับแม่ตลอดเวลา
H3-E009	จำคุกจูจิสัน ข้อหาเสพยาอี ถูกศาลตัดสินจำคุกหลายปี
H3-E010	โอ วรุธ ขอมรับว่ามีภรรยาแล้ว หลังจากคบกันมานาน 2 ปี แล้วภรรยาตั้งครรภ์ได้ 8 เดือน โดยมีกำหนดคลอดบุตรในเดือน ก.ค. นี้
H3-P003	นายสิริศักดิ์ ดิยะพรรณ อธิบดีอัยการต่างประเทศ กล่าวว่าไม่ทราบข่าวกระแสที่ พ.ต.ท. ทักษิณ ชินวัตร อดีตนายกรัฐมนตรี้ พักอาศัยในเอเชีย
H3-P004	นายธฤตจรุงวัฒน์ อธิบดีกรมสารนิเทศและโฆษกกระทรวงการต่างประเทศ กล่าวถึงความเป็นไปได้ในการแลกตัวผู้ต้องหากับยูเออี
H3-P005	ผบ.สส. ไม่ขัดขวางพวกนายทหารสูงสุด เดินทางไปพบ พ.ต.ท. ทักษิณ ที่ฮ่องกง

Table B.3 (continued)

Summary	Summary
ID	
H3-P006	สมาชิก กบข. แต่งชุดดำบุกทำเนียบ เรียกร้อง 6 ข้อ ยกเลิกกองทุน
H3-P007	นายอภิสิทธิ์ เวชชาชีวะ นายกรัฐมนตรี ย้ายจุดยืนไทยคัดค้านเขาพระวิหารขึ้นเป็นมรดกโลก
H3-P008	นายสุเทพ เทือกสุบรรณ รองนายกรัฐมนตรีฝ่ายความมั่นคง เชื่อว่าปัญหาชายแดนไทย-กัมพูชาจะคลี่คลาย
H3-P009	พล.อ. ประวิตร วงษ์สุวรรณ รัฐมนตรีว่าการกระทรวงกลาโหม พร้อมคณะเดินทางลงพื้นที่ จ.ภูเก็ต เพื่อตรวจดูความพร้อมการจัดประชุมอาเซียน
H3-P010	นายสุเทพ เทือกสุบรรณ รองนายกรัฐมนตรีฝ่ายความมั่นคง ในฐานะเลขาธิการพรรคประชาธิปัตย์ ให้สัมภาษณ์ถึง กกต. วินิจฉัย 28 ส.ส. พรรค ภูมิใจไทย ว่าจะฟ้องหรือไม่ ว่าทุกอย่างดำเนินไปตามกระบวนการทางกฎหมาย
H3-G001	พบว่าเกิดรอยร้าวบนพระธาตุคอกสุเทพ และจะทำการซ่อมแซม หากปล่อยไว้เกรงว่าจะเกิดอันตราย จึงจะทำการบูรณะขึ้น
H3-G002	เกิดเพลิงไหม้ขึ้นที่ห้างบิ๊กซี จ.พระนครศรีอยุธยา คาดว่าเกิดจากไฟฟ้าลัดวงจร ทำให้ผู้คนหนีกันอย่างอลหม่าน
H3-G003	คนไทยติดเชื้อไข้หวัดใหญ่ 2009 มีผู้เสียชีวิตแล้ว 2 ราย
H3-G004	ประเทศเม็กซิโกการระบาดของไข้หวัด 2009 ลดลง แต่ก็ยังมีผู้ติดเชื้อเพิ่มขึ้น จึงได้ทำความสะอาดที่ต่างในกรุงเม็กซิโก
H3-G005	ยอดติดเชื้อหวัด 2009 ทั่วโลกพุ่งสูงขึ้นจนทะลุถึง 12,000 คนแล้ว
H3-G006	เกิดแผ่นดินไหว ในประเทศเม็กซิโก มีความแรง 5.7 ริกเตอร์ ทำให้อาคารต้นสะเทือน ผู้คนแตกตื่นหนีตายกันจ้าละหวั่น
H3-G007	อดีตประธานาธิบดีประเทศเกาหลีใต้ กระโดดฆ่าตัวตายตกลงหน้าผา ก่อนตายทิ้งจดหมายลาตายไว้

Table B.3 (continued)

Summary	Summary
ID	
H3-G008	คุณหญิงพจมานทได้ทำการเปลี่ยนนามสกุลจาก ชินวัตร เป็น ดามาพงศ์ หลังจาก อย่างกับ พ.ต.ท. ทักษิณ ที่สำนักงานเขตดุสิต
H3-G009	อย. ออกเคทอนยาแก้หวัดปลอมที่ซื้อมากินเอง หลังจากมีการจับกุมผู้ลักลอบซื้อ ขายยาที่มีส่วนผสมของยาที่มีส่วนประกอบ สู่โคเอฟีดรีน-ไตรโพลิตินอยู่
H3-G010	นักการเมืองสาวมาเลย์ลาออกจากการเป็น สส. ของพรรค หลังจากกรู๊ปหวิวหลุด ว่อนในอินเทอร์เน็ตและมีมือถือ

Table B.4 Summaries Result of System.

Summary	Summary
ID	
TS-E001	วันที่ 23 พฤษภาคม นายอาทิตย์ราห์ คงมาลัย หรือ ตูน บอดี้สแลม เข้าพิธีอุปสมบทที่วัดเกาะทอง อ.เมืองขอนแก่น จ.ขอนแก่น
TS-E002	วิคตอเรีย เดินตามรอยสามิดารานักเตะ เดวิด เบคแฮม สดกีฬาโพสท่าถ่ายแบบโฆษณา ชุดชั้นในสตรีคอลเล็กชั่นสปริงซัมเมอร์ให้ไฮเบอร์เนตส์สุดหรู เอ็มโพริโออามานี
TS-E003	วันที่ 28 มิถุนายน แบลม จนิศดา ล้วนเฉลิมวงศ์ คง เจ้าบ่าว โป๊ท บุตรรัตน์ จรุงสมิทธิ จัดพิธีฉลองมงคลสมรสพระราชทาน ที่ห้องรอยัลบอลรูม โรงแรมแมนดาริน โอเรียนเต็ล
TS-E004	วันที่ 25 มิถุนายน ไมเคิล แจ็กสัน นักร้องชื่อดังของวงการดนตรีโลก เสียชีวิตที่ลอสแอนเจลิส
TS-E005	วันที่ 29 มิถุนายน นักแสดงและนางแบบสาว สา-มาริสา แอนนิต้า ประสบอุบัติเหตุรถยนต์พลิกคว่ำ ที่ อ.สตึก จ.ชลบุรี เพื่อไปถ่ายละครเสาร์ 5
TS-E006	สาว น้อย เกศริน เอกธวัชกุล รีเทิร์นกลับมาสวีทหวาน เพราะคงกันไปไหนมาไหนด้วยกันหลายต่อหลายครั้ง
TS-E007	วันที่ 24 มิถุนายน ศรีภรรยาหน้อย บุญกร ตั้งท้องทายาทคนที่สองได้ 6 สัปดาห์แล้ว
TS-E008	ดาราสาว แซมมี่ ปัทมिता ภูวิจารณ์ เคา์เวลล์ ตกเป็นข่าวฉาวไม่หยุดไม่หย่อน แอบควง 2 หนุ่ม อ้วน รังสิต ศิรนานนท์ และ เต๋อชนพล นิมพัชสุข
TS-E009	วันที่ 23 มิถุนายน นักแสดงชายเกาหลีชื่อดัง จู จีฮุน ได้เดินทางไปขึ้นศาลเพื่อรับฟังการพิจารณาคดี หลังยอมรับว่ากระทำผิดจริงในข้อหาเสพยา อีและยาเค
TS-E010	โอเพย์มีภรรยาแล้วท้องแก่คลอดเดือนหน้า ยัมกริมหมอบอกได้ลูกชายตั้งชื่อล่วงหน้าว่า แอร์บัส
TS-P001	วันที่ 15 มกราคม ม.ร.ว.สุขุมพันธุ์ บริพัตร ผู้ว่าฯ กทม. เข้ารับ หนังสือรับรองตำแหน่งผู้ว่าฯ กทม.คนที่ 16

Table B.4 (continued)

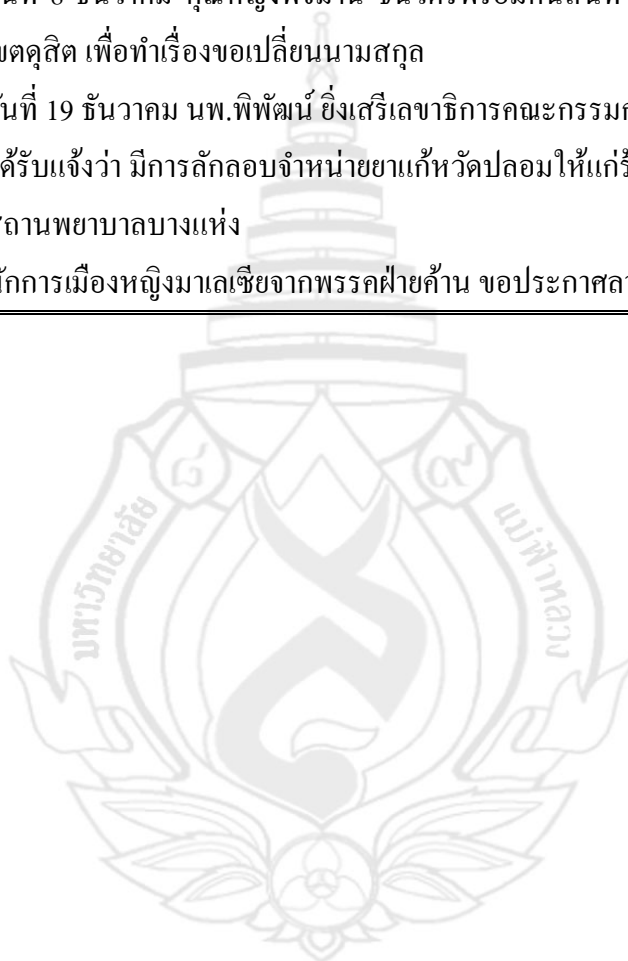
Summary	Summary
ID	
TS-P002	วันที่ 20 น.ส.กัญญาภักดิ์ มณีจันทร์ แกนนำกลุ่มรักเชียงใหม่ 51 หรือกลุ่มเสื้อแดง เชียงใหม่ พร้อมด้วย ชาวบ้าน เดินทางมาชุมนุมประท้วง ที่ บจก.ปุยอินทรีย์ ฟอสเฟตตราจักรทอง
TS-P003	วันที่ 28 มิถุนายน นายศิริศักดิ์ ดิยะพรรณ อธิบดีอัยการฝ่ายต่างประเทศ กล่าวถึง กระแสข่าว พ.ต.ท.ทักษิณ ชินวัตร กลับมาอยู่ในประเทศแถบเอเชียว่า อัยการยังไม่เคยได้รับการยืนยันจากกระทรวงการต่างประเทศ และเจ้าหน้าที่ตำรวจ
TS-P004	วันที่ 22 พฤษภาคม นายธฤต จรุงวัฒน์ อธิบดีกรมสารนิเทศและ โฆษกกระทรวง การประเทศ กล่าวถึง เรื่องความเป็นไปได้ในการแลกตัวผู้ต้องหาระหว่างไทยกับ สหรัฐอาหรับเอมิเรตส์ ที่กระทรวงการต่างประเทศ โดยหลักการแล้วประเทศที่ไม่มีการทำสนธิสัญญาการส่ง ตัวผู้ร้ายข้ามแดนจะใช้หลักต่างตอบแทน
TS-P005	วันที่ 21 กุมภาพันธ์ พล.อ.ทรงกิตติ จักกาบาตร์ผู้บัญชาการทหารสูงสุด กล่าวถึง กรณี ส.ส.พรรคเพื่อไทยและนายทหารระดับสูงรุ่นเดียวกันเดินทางไปเยี่ยม พ.ต.ท.ทักษิณ ที่ฮ่องกง
TS-P006	วันที่ 29 มิถุนายน นายวิศร อัครสันตติสกุล ประธานเครือข่ายสมาชิกกองทุน บำเหน็จบำนาญข้าราชการแห่งประเทศไทยนำข้าราชการ ที่เป็นสมาชิกกบข. มายัง สะพานชมัยมรุเชฐ ถนนพิษณุโลก ด้านหน้าทำเนียบรัฐบาล เพื่อเรียกร้องยื่นข้อ เรียกร้อง 6 ข้อ ให้นายอภิสิทธิ์ เวชชาชีวะ นายกรัฐมนตรี
TS-P007	วันที่ 29 มิถุนายน นายอภิสิทธิ์ เวชชาชีวะ นายกรัฐมนตรี เผยว่าประเทศไทยยังคง สงวนสิทธิ์การคัดค้านเขาพระวิหารขึ้นเป็นมรดก โลก
TS-P008	วันที่ 27 มิถุนายน นายสุเทพ เทือกสุบรรณ รองนายกรัฐมนตรี ฝ่ายความมั่นคง เปิดเผยก่อนเดินทางเยือนกัมพูชา โดยยืนยันว่าจะได้พบกับนายกรัฐมนตรีของ กัมพูชาอย่างแน่นอน

Table B.4 (continued)

Summary	Summary
ID	
TS-P009	วันที่ 29 มิถุนายน พล.ต.อ.ประวิตร วงษ์สุวรรณ รัฐมนตรีว่าการกระทรวงกลาโหม ได้นำ ผบ.เหล่าทัพทุกเหล่าทัพและเจ้าหน้าที่ฝ่ายความมั่นคง ไปตรวจความพร้อมในการประชุมอาเซียนซัมมิต ที่ภูเก็ต เพื่อตรวจความพร้อมของการจัดการประชุม รัฐมนตรีอาเซียน
TS-P010	วันที่ 30 มิถุนายน นายสุเทพ เทือกสุบรรณ รองนายกรัฐมนตรีฝ่ายความมั่นคง ในฐานะเลขาธิการพรรคประชาธิปัตย์ กล่าวถึงกรณีที่คณะกรรมการการเลือกตั้ง จะวินิจฉัยชี้ขาด 28 ส.ส.ของพรรคประชาธิปัตย์ในการถือครองหุ้น ที่ทำเนียบรัฐบาล
TS-G001	วันที่ นายสทวัฒน์ เน้นหนา ผู้อำนวยการสำนักศิลปากรที่ 8 เชียงใหม่ เปิดเผยว่า นายธีระ สลักเพชร รัฐมนตรีว่าการกระทรวง วัฒนธรรมเดินทางมายัง วัดพระธาตุดอยสุเทพ จ.เชียงใหม่ เพื่อตรวจดูความแข็งแรงของพระธาตุดอยสุเทพ
TS-G002	วันที่ 28 มิถุนายน พ.ต.ท.สมาน ศรีอรุณ พนักงานสอบสวน สภ.บางปะอิน จ.พระนครศรีอยุธยา รับแจ้งเหตุ เพลิงลุกไหม้ ที่ห้างสรรพสินค้าบิ๊กซี ซูเปอร์เซ็นเตอร์ สาขาจ.พระนครศรีอยุธยา
TS-G003	สำนักงานสาธารณสุขได้รับรายงานแจ้งจากโรงพยาบาลรามคำแหงว่า มีผู้ป่วยเพศ หญิงวัย 40 ปี ซึ่งเข้ารับการรักษาตัวโดยแพทย์วินิจฉัยว่าเป็นไข้หวัด 2009 เสียชีวิตลง
TS-G004	วันที่ 4 พฤษภาคม นายโฮเซ อังเกล คอร์โดวา รัฐมนตรีสาธารณสุขเม็กซิโก ระบุว่า สถานการณ์การแพร่ระบาด เชื้อไวรัสไข้หวัดใหญ่สายพันธุ์ใหม่ ในประเทศ เม็กซิโกลดความรุนแรงลงแล้ว
TS-G005	วันที่ 23 พฤษภาคม สำนักข่าวต่างประเทศรายงานว่าเชื้อไวรัส H1N1 ชนิดเอ สาเหตุของไข้หวัดใหญ่สายพันธุ์ใหม่ 2009 ยังแพร่กระจายไปทั่วโลกใน 46 ประเทศ
TS-G006	วันที่ 23 พฤษภาคม เกิดแผ่นดินไหววัดแรงสั่นสะเทือนได้ 5.7 ริกเตอร์ ที่กรุง เม็กซิโกซิตี

Table B.4 (continued)

Summary	Summary
ID	
TS-G007	วันที่ 23 พฤษภาคม อดีตประธานาธิบดี โรห์ มู-ฮย็อนของเกาหลีใต้เสียชีวิต
TS-G008	วันที่ 8 ธันวาคม คุณหญิงพจมาน ชินวัตรพร้อมคนสนิท เดินทางมา ยังสำนักงานเขตคูสิต เพื่อทำเรื่องขอเปลี่ยนนามสกุล
TS-G009	วันที่ 19 ธันวาคม นพ.พิพัฒน์ ยิ่งเสรีเลขาธิการคณะกรรมการอาหารและยา ได้รับแจ้งว่า มีการลักลอบจำหน่ายยาแก้หวัดปลอมให้แก่ร้านขายยาและสถานพยาบาลบางแห่ง
TS-G010	นักการเมืองหญิงมาเลเซียจากพรรคฝ่ายค้าน ขอประกาศลาออกจากตำแหน่ง



APPENDIX C

SYNONYM DATABASE

Table C.1 Synonym Database



Word	Status	Set
1	common	1001
หนึ่ง	unidentified	1001
กล่าว	unidentified	1002
กล่าวว่	common	1002
บอก	unidentified	1002
เล่า	unidentified	1002
ม.ค.	unidentified	1003
มกราคม	common	1003
สส	unidentified	1004
สมาชิกสภาผู้แทนราษฎร	common	1004
ส.ส.	unidentified	1004
กิน	unidentified	1005
รับประทาน	common	1005
ช่างภาพ	common	1006
ตากล้อง	unidentified	1006
สามี	common	1007
ผ้า	unidentified	1007
สลดผ้า	unidentified	1008
เปลือยกาย	common	1008

Table C.1 (continue)

Word	Status	Set
ชื่อตั้ง	common	1009
มีชื่อเสียง	unidentified	1009
ก.พ.	unidentified	1010
กุมภา	unidentified	1010
กุมภาพันธุ์	common	1010
มี.ค.	unidentified	1011
มีนา	unidentified	1011
มีนาคม	common	1011
เม.ย.	unidentified	1012
เมษา	unidentified	1012
เมษายน	common	1012
พ.ค.	unidentified	1013
พฤษภา	unidentified	1013
พฤษภาคม	common	1013
มิ.ย.	unidentified	1014
มิถุนา	unidentified	1014
มิถุนายน	common	1014
ก.ค.	unidentified	1015
กรกฎา	unidentified	1015
กรกฎาคม	common	1015
ส.ค.	unidentified	1016
สิงหา	unidentified	1016
สิงหาคม	common	1016
ก.ย.	unidentified	1017
กันยา	unidentified	1017
กันยายน	common	1017

Table C.1 (continue)

Word	Status	Set
ต.ค.	unidentified	1018
ตุลา	unidentified	1018
ตุลาคม	common	1018
พฤศจิกายน	common	1019
ธ.ค.	unidentified	1020
ธันวา	unidentified	1020
ธันวาคม	common	1020
อย.	unidentified	1021
องค์การอาหารและยา	common	1021
สำนักงานคณะกรรมการอาหารและยา	unidentified	1021
จำหน่าย	unidentified	1022
ขาย	common	1022
ปลอม	common	1023
เทียม	unidentified	1023
เปิดเผยว่า	unidentified	1002
นายกฯ	unidentified	1024
นายกรัฐมนตรี	common	1024
ม.ร.ว.	unidentified	1025
หม่อมราชวงศ์	common	1025
กทม.	unidentified	1026
กรุงเทพ	unidentified	1026
กรุงเทพมหานคร	common	1026
กกต.	unidentified	1027
คณะกรรมการการเลือกตั้ง	common	1027
ผู้ว่าฯ	unidentified	1028
ผู้ว่าราชการ	common	1028

Table C.1 (continue)

Word	Status	Set
กลุ่มเสื้อแดง	unidentified	1029
เสื้อแดง	unidentified	1029
แนวร่วมประชาธิปไตยต่อต้านเผด็จการแห่งชาติ	common	1029
นปก.	unidentified	1029
นส.	unidentified	1031
นางสาว	common	1031
บจก.	unidentified	1032
บริษัทจำกัด	common	1032
ส.ส.จ.	unidentified	1033
สมาชิกสภาจังหวัด	common	1033
แม่่ว	unidentified	1034
ทักษิณ	common	1034
พล.อ.	unidentified	1035
พลเอก	common	1035
พ.ต.ท.	unidentified	1036
พันตำรวจโท	common	1036
มือถือ	unidentified	1037
โทรศัพท์มือถือ	common	1037
ยาแก้หวัด	unidentified	1038
ยาแก้ไข้หวัด	common	1038
ครู	unidentified	1039
อาจารย์	common	1039
หลับ	common	1040
นอน	unidentified	1040
นอนหลับ	unidentified	1040
มกรา	unidentified	1003

Table C.1 (continue)

Word	Status	Set
ประชาธิปไตย	common	1041
ม.	unidentified	1042
มหาลัย	unidentified	1042
หงส์แดง	unidentified	1044
ลิเวอร์พูล	common	1044
ผีแดง	unidentified	1045
เมนยู	unidentified	1045
แมนฯยู	unidentified	1045
แมนเชสเตอร์ยูไนเต็ด	common	1045
ปืนใหญ่	unidentified	1046
อาร์เซนอล	common	1046
จ.	unidentified	1047
จังหวัด	common	1047
มีชื่อ	unidentified	1009
โด่งดัง	unidentified	1009
ชค	unidentified	1020
พย	unidentified	1019
ตค	unidentified	1018
กย	unidentified	1017
กพ	unidentified	1010
มีค	unidentified	1011
เมย	unidentified	1012
พค	unidentified	1013
มิย	unidentified	1014
กค	unidentified	1015

Table C.1 (continue)

Word	Status	Set
สค	unidentified	1016
มค	unidentified	1003
ไป	common	1048
เดินทาง	unidentified	1048
ถึงแก่อสัญกรรม	unidentified	1049
เสียชีวิต	common	1049
ดอด	unidentified	1050
ตั้งท้อง	common	1051
ตั้งครรภ์	unidentified	1051
มีครรภ์	unidentified	1051
สำนักงานสาธารณสุข	common	1052
สธ.	unidentified	1052

CURRICULUM VITAE

NAME Miss Piyabud Ploadaksorn

DATE OF BIRTH 6 December 1984

ADDRESS 290/1, Tambon Thayang,
Amphoe Thungyai, Nakonsitammarat 80240

EDUCATIONAL BACKGROUND

2007 Bachelor of Science
Major in Information Technology
Prince of Songkla University University

