Unify Framework for Crime Data Summarization using RSS Feed Service*

Tichakorn NETSUWAN and Kraisak KESORN*

Department of Computer Science and Information Technology, Faculty of Science, Naresuan University, Phitsanulok 65000, Thailand

(*Corresponding author's e-mail: kraisakk@nu.ac.th)

Received: 2 September 2016, Revised: 1 April 2017, Accepted: 1 May 2017

Abstract

This research presents online crime news analysis using text mining, Natural Language Processing framework (General Framework for Text Mining: GATE), and data warehouse (DW) technologies. The proposed framework aims at extracting key features of crime data available on newspaper website and classifies them into crime categories which are later transformed into a star schema for speedy retrieving and online analytical processing (OLAP). This system can present data in multidimensional structure to perform data analytics to support police officers for determining the security policies to protect locals and tourists who live in the risk areas. The main novelty of this framework is the demonstration of using information available through RSS feed service to generate reports to support decision making. The experimental results show that the extracted data from the Internet can effectively represent the actual crime data occurred in the study areas (low error rate) and allow data analysts to get an insight of the information represented through OLAP.

Keywords: Crime news, data analytics, OLAP, text mining, data warehouse

Introduction

Crime can reflect the social problems and should be prevented or decreased for the good of living of people and tourists in the areas. Typically, crime information is daily presented on the website and read by millions of audience worldwide. Unfortunately, this information is rarely used to benefit for people e.g. safety and security aspects. According to the survey in the first 6 months of 2016 by *Numbeo* website [1], crime index in Thailand is 52.16 and it is 4th among South-East Asian countries. While Malaysia has the highest crime rate at 65.56 followed by Vietnam at 53.45 and Combodia at 52.72. In Thailand, crime statistics illustrates that Pattaya obtains the highest crime at 58.55 followed by Phuket (57.38), Bangkok (47.70), and Chiang Mai (34.94). It is noticeable that they are major cities for tourists of the country.

Typically, crime information is manually collected only when there is someone report to the officers and this data will usually be stored in database which allows police officers to retrieve, analyze, and manually generate crime reports which is very labor intensive. The main limitations of this method are 1) data is not up to date as manually collection process is time consuming and 2) scalability problem could occur because numbers of tables and amount of data are increased in the future and, consequently, querying and reporting speed are affected due to integrity checking of Database Management System (DBMS). To solve these limitations, we proposed to exploit crime news available on the Internet to

*Presented at 1st International Conference on Information Technology: October 27th - 28th, 2016 support decision making of police officers by constructing the crime news extraction and analytics system which includes several functions as follows.

1) Extract crime data from online news website and later use for data analytic purpose.

2) Store crime data using star schema in DW which efficiently resolve the scalability problem and effectively represent data in multidimensional forms, so called online analytical processing (OLAP).

3) Generate interactive reports of this data via website which users can perform drill-down or roll-up operations. The proposed system can aid police officers to form the policies in order to prevent crime that may occur in the risk areas.

Related works

News usually influences our lifestyles and, thus, several researchers exploited news to automate analytics in various aspects. Shojaee *et al*. [2] proposed a study of classification learning algorithms to predict crime status using crime dataset from University of California, Irvine (UCI) Machine Learning Repository for data mining. However, crime dataset from UCI is not updated and this results in practically useless. Yang *et al*. [3] studied learning approaches for detecting and tracking news events of interests using Reuters and CNN news stories. However, the presented method only be useful in some circumstances and need users involve in the loop of event identification. Seo *et al.* [4] presented the financial news analysis for intelligent portfolio management which is a text classification agent that takes advantage of information retrieval techniques to complement quantitative financial information. These news articles were gathered from various electronic news providers e.g. CNN Financial Network, Forbes, Reuters, NewsFactors, Motley Fool, CNet, ZDNet, Morningstar.com, Associate Press (AP), AP Financial, and Business wire. Nonetheless, it is unclear that how they store data for later use and support for scalability issue. Online news allows users to easily access news anytime and anywhere via mobile devices and personal computers. Really Simple Syndication (RSS) technology enables publishers to syndicate data automatically. A standard XML file format ensures compatibility with many different machines/programs. RSS feeds also benefit to users who want to receive timely updates from favorite websites or to aggregate data from many sites. Wanglee *et al*. [5] developed an automatic news aggregator system which users can read all news. However, crime news information has never been used to support decision making for the police officers in the literatures. Sudhahar *et al.* [6] presented a system for large scale quantitative narrative analysis (QNA) of news corpora. The task is to identify the key actors (criminals and victims) in news and the actions they performed. The system demonstrated that men were most commonly responsible for crimes against the person, while women and children were most often victims of those crimes. Wang *et al.* [7] analyzed online news and classified into categories by means of adaptive clustering. Moreover, the news comments were classified into categories such as negative, positive, which are also grouped into clusters helping the experts to get the view of the common people to the news. The main purpose of this work is to help the experts find which news that the people concerned the most. However, the main drawback of this work is it does not store data for offline processing and cannot represent data in multi-dimensions. Thus, it is not efficiently support for decision making of users. Most of researches in the literatures deployed database system to store the extracted data which is usually suffer from integrity checking and resulting in poor querying speed when the number of tables and volume of data become large. To the best of our knowledge, there is no crime news extraction and analytics system existing in the literatures and none of them use data warehouse as an architecture for data storage. Hence, we present this idea as a main contribution in this paper.

Materials and methods

To develop online crime news extraction and analysis framework, 4 major components are introduced as illustrated in **Figure 1** and can be described as follows.

(1) *Collection*: crime data in this work focused on English news only and was automatically collected from http://www.pattayapeople.com exploiting RSS feed service. Finally, the collection in this research contains 1,596 crime news.

(2) *Extraction and analysis*: The collected data is analyzed and extracted key information using GATE framework [8]. GATE is preferred because it is open source framework and well known among researchers in this area. This step includes 3 sub-processes: 1) Name entity extraction to detect names, locations, organizations, address, date etc. 2) Crime terms detection to find key terms related to crimes. This research uses Oxford Learner's Dictionaries [9] and MyVocabulary [10] to detect crime keywords. Example of crime terms are shown in **Table 1**. 3) All crime terms are processed in order to extend to other related data. **Figure 2** demonstrates the example of other indirectly relevant data that can be extended from the crime terms e.g. gender, criminal, and victims. To find other relevant information, several related information is used e.g. system date, name prefix, and structure grammatical formalisms. **Table 2** shows examples of structure grammatical formalism rules for criminal and victim identification. At this stage of experiment, the rules are fixed regarding to grammatical formalism. If a user wants to add more rules, he can manually do this.

(3) *Classification*: news classification using text mining to classify crime news into 5 categories. **Table 3** depicts crime categories [11] used in this work. This task aims at showing the highest number of crime types occurred in Pattaya. To classify crime news, all extracted keywords must be counted the frequency and put them into a matrix (**Table 4**).

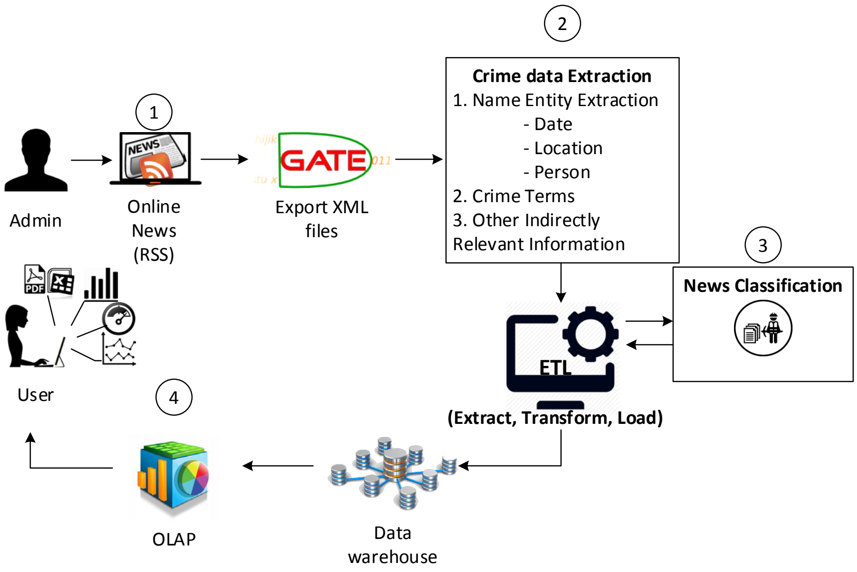


Figure 1 Online crime news extraction and analysis framework.

Finally, the important between keywords and news are computed using *TF-IDF* (Eqs. (1) and (2)) [12] and Artificial Neural Network (*ANN*) is applied for document classification task.

$$IDF_k = \log(n / DF_k) \tag{1}$$

$$TF - IDF_{jk} = TF \times IDF_k \tag{2}$$

where *n* is number of news in the collection, *TF* refers to the frequency of term *t* in a document and *DF* represents a number of documents containing term *t*.

(4) *Representation*: the extracted crime data is represented in a multidimensional form using OLAP which is a process to integrate data based upon star schema. Microsoft SQL Server 2008 R2 is used as a

tool in this work for DW construction. It helps to perform data analytics in multidimensional tables and pre-summarized across dimensions to drastically improve query speed over relational databases. In addition, designing data at multiple levels of aggregations allows user to perform drill-down or roll-up operations. Drill-down presents data at a level of increased detail, while roll-up is the reverse operation of drill-down by decreasing detail of data. **Figure 3** shows an example of star schema which comprises a fact table and dimension tables.

Example sentence

<Sentence> Dongtan sub police station was called out by the Sawang Boriboon Rescue Volunteers in the early hours of <Date> Monday , 13th October </Date> to investigate an attempted suicide by a Swiss male expat .<Person> Mr . Hans Maurer </Person>, aged around 60 was found unconscious in a parked Chevrolet in a secluded area of Tesco Lotus car park off <Location> Thepprasit Road </Location>. Sentence>

Crime data extraction	Other indirectly relevant information
Date : Monday , 13th October	Year : 2014
Person : Mr. Hans Maurer	Gender : Male
	Identify Person : Victim
Location : Thepprasit Road	Region : South Pattaya

Figure 2 Example other indirectly relevant information extended from crime data.

Performance evaluations and discussion

To evaluate the presented system, the experiments have been conducted using a customized dataset contain 1,596 crime news. Since Pattaya has highest numbers of crimes, we scope crime news to this area only. Several measures are used to validate the performance of the proposed system e.g. Lift chart, ROC curve, Precision, Recall, and F-measure. The experiments are divided into several sections in order to investigate all aspects of the system performance such as classification performance, data extraction efficiency, and error rate of the extracted data compared to the actual data. Finally, an example of OLAP report is demonstrated. The evaluation results are described as the follows sections.

Table 1 Example crime terms.

Crime terms						
killing	murder	slash	jump	stab	shoot	fire
smash	rape	steal	hit	fight	punch	skid
threaten	rob	snatch	extort	ransack	pickpocket	theft
impound	cheat	pretend	bribe	bribing	attack	blow
kidnapping	abduct	electrocuted	arm	terrorist	prostitutes	YABA
kick	assault	violate	thieving	drug	gamble	addict

Table 2 Criminal identification rules based on sentence structures.

Criminal identification rules	**Example sentence**
Rules 1 found keywords "arrest", "caught", "suspect", "detained"	Sriracha police announced the **arrest** of **Mr. Watid** on Friday.
Rules 2 <**Person**> + active voice Or Passive voice + <**Person**>	Mr. Vladimir Ragoziw who **was hit** **by** **Mr. Paing** riding along this busy highway.
Victim identification rules	**Example sentence**
Rules 1 found a keyword "victim"	The **victim** was 26 year old Russian tourist **Mr. Chore Dnichonko**, identified from a rental motorcycle paper found on the body.
Rules 2 Passive voice <**Person**> + passive voice	Pattaya police received a call from **Mr. Rewat**, a motorcycle taxi rider, to report that he **had** just **been threatened and robbed** by a male customer.

Table 3 Crime categories.

Crime category	**Crime feature**
1. Serious offense	Murder, Kidnapping, Arson
2. Bodily harm case	Manslaughter, Death by negligence, Attempted to murder, Assault, Rape
3. Crimes against property	Theft, Snatching, Blackmail, Extortion, Receiving stolen property, Vandalism
4. Interesting case	Motorcycle theft, Car theft, Bus robbery, Taxi robbery, Cheater and fraud, Misappropriation
5. Corrupt state case	Weapon act, Gambling act, Narcotics act, Prostitution act, Materials act

Table 4 News-Terms frequency matrix.

News	**TF (Term Frequency)**								
	slash	**stab**	**attack**	**shoot**	**rob**	**steal**	**jump**	**kidnapping**	**kick**
News_1	1	0	0	0	0	0	0	0	1
News_2	0	2	1	0	0	0	0	0	0
News_3	0	0	1	5	0	0	0	0	0
…	…	…	…	…	…	…	…	…	…
News_{1596}	0	0	1	0	0	0	0	1	0
DF (Document frequency)	**20**	**30**	**35**	**39**	**31**	**172**	**8**	**28**	**5**

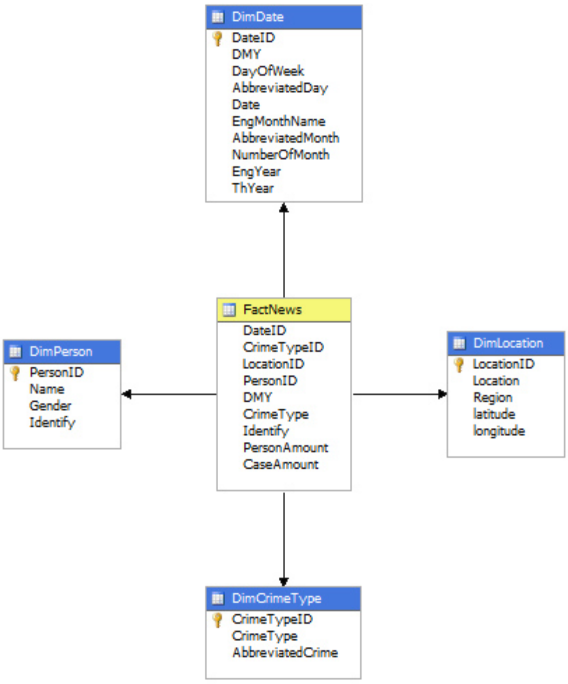


Figure 3 Star schema of crime news in data warehouse.

News classification performance

Two state of the art techniques of data mining are deployed and compared with the presented technique in order to evaluate the classification power. If the extracted data has high quality, it will enhance the classification performance. We also want to study which data mining is the best classification model for crime news information. In this experiment, Artificial Neural Network (*ANN*), Decision Tree (DT), and Naïve Bayes (NB) are compared together. The lift chart in **Figure 4** illustrates the performance of classification of 3 models where the x-axis represents the percentage of data population and the y-axis is the percentage of correctly data classification. The ideal line represents the model with 100 % correct classification and illustrated by diagonal line (dark bold line) and it is used to measure the classification performance. The Lift chart shows *ANN* superiors than other 2 models as its performance is closer to the ideal model than others. The average correct classification is about 84.16 % while NB and DT obtain 81.86 and 75.12 %, respectively. To confirm this result, Relative Operating Characteristic (ROC) curve is used to re-validate the classification power of those 3 models. ROC curve is a line chart that shows the true positive (TP) rate versus its false positive (FP) rate of a classifier. The best possible prediction technique would obtain a point in the upper left corner of the ROC chart, representing 100 % true positive and 0 % false positive. An ideal model would give a point along a diagonal line from the left bottom to the top right corners which divides the chart. Areas above the diagonal represent good classification results while areas below the diagonal line refer poor results [13,14]. **Figure 5** demonstrates that *ANN* obtains the highest performance (area under the curve = 0.86) followed by NB and DT (area under the curve = 0.83 and 0.64, respectively). This result is consistent with the result in **Figure 4** and, thus, this can confirm that *ANN* the best model for this task. However, there are some parameters of *ANN* needed to be adjusted to increase the prediction accuracy of *ANN*. Next section will describe about *ANN* parameters tuning.

Artificial Neural Network parameters tuning

ANN is deeper investigated for parameters adjustment in order to obtain the highest performance of the model. There are some parameters of *ANN* that are needed to tune to optimize the classification power. This section aims at investigating the effect of those 2 parameters (Hidden_node_ratio and hold_out_seed) on the performance of *ANN* prediction model by varying one parameter at a time [15]. The important parameters are adjusted and shown in **Table 5**. Various pairs of (β,δ) values were tried, and the one with the best accuracy was selected. These parameter adjustments result in better classification performance by obtaining 87.40, 88.10 and 87.44 % of precision, recall, and F-measure, respectively as illustrated in **Figure 6**.

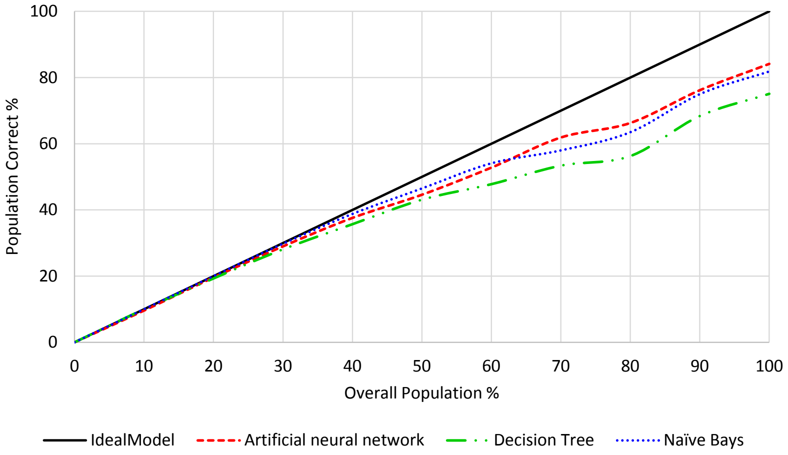


Figure 4 Lift chart compared the classification performances of 3 models.

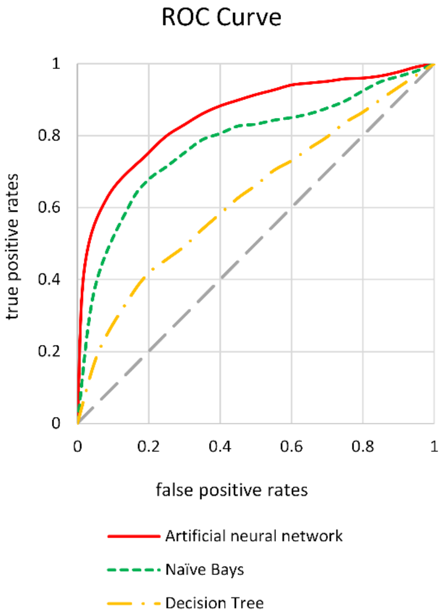


Figure 5 ROC curve compared 3 classification models.

Table 5 Artificial Neural Network parameters.

Parameter	Default	Range/Adjusted value
Hidden_node_ratio (β)	4	[0, 4]
Holdout_seed (δ)	0	[0, 4]
Maximum_input_attributes	255	107
Maximum_output_attributes	255	1
Simple_size	1,000	1,569

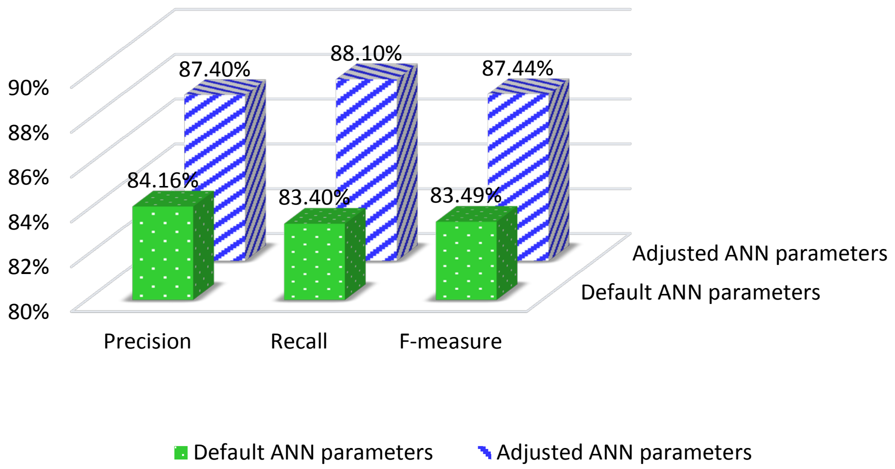


Figure 6 Classification performance after *ANN* parameters adjustment.

Data extraction efficiency

This experiment aims at evaluating GATE framework exploited in this research for Natural Language Processing (NLP) task. We study the performance of GATE to identify person, location, and date measured by precision, recall and F-measure. The result is shown in **Figure 7**. The lowest accuracy obtained from GATE is person identification about 77.85 % of precision. We analyzed this result and found that GATE poorly identifies a person with Thai name. This is because GATE is mainly designed for English name entity processing and, consequently, it is not able to detect Thai name as well as lacking of the prefix name. This can lower the name entity recognition (NER) power of GATE. This is also similar to location recognition because they are also Thai names. In contrast, date format in crime news is written using standard format. As a result, GATE effectively recognizes date data and, as such, it obtains highest performance at 90.29, 94.57 and 86.48 % of precision, recall, and F-measure, respectively.

Grammatical formalism rules evaluation

Since we extended the result of GATE framework in order to identify other relevant data e.g. criminal, victim, and gender (**Table 2**). Therefore, we need to evaluate the performance of the presented grammatical formalism rules. **Figure 8** demonstrates the performance of grammatical formalism rules to identify criminal and victim. Victim are identified more correctly than criminal at 83.22 % of precision because it is easier to identify victims based on passive voice whereas criminal is written using various grammatical formalisms. Hence, criminal identification receives poorer performance compared to victim.

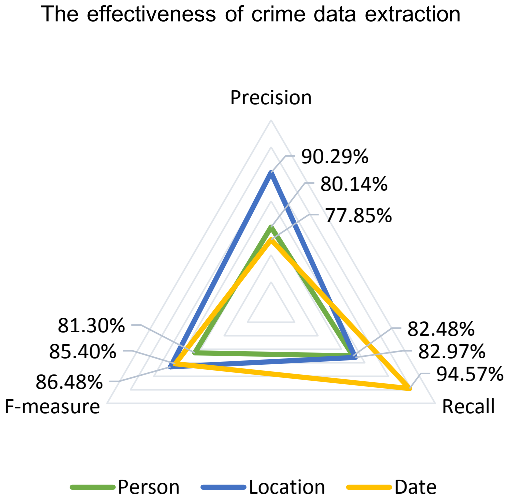


Figure 7 Effectiveness of crime data extraction using GATE framework.

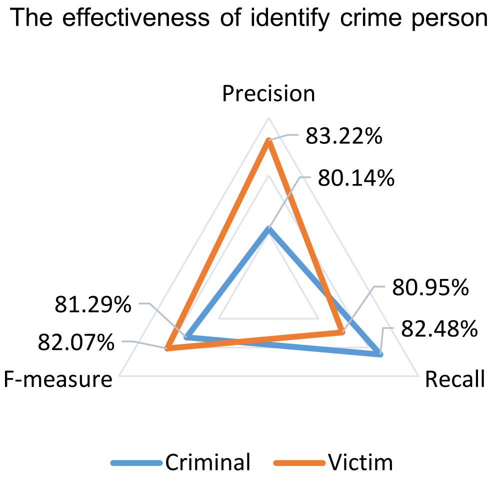


Figure 8 Criminal and victim identification using structure grammatical formalism rules.

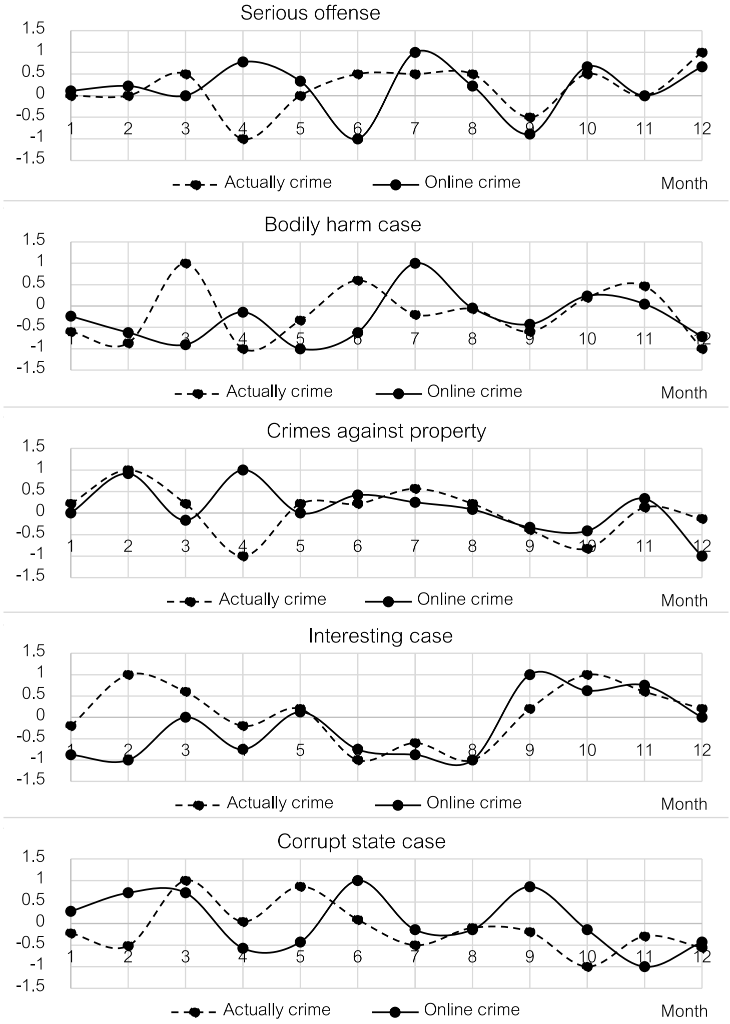


Figure 9 Normalized data comparison between extracted crime data from the Internet and actual data collected by polices.

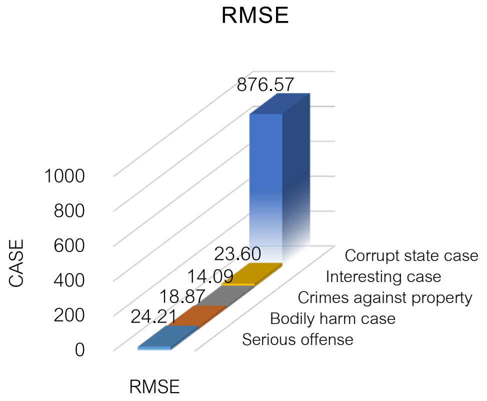


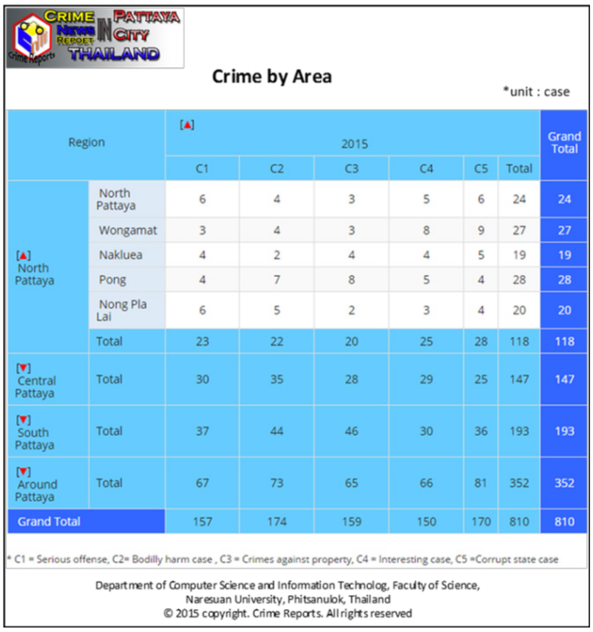
Figure 10 Root Mean Squared Error (RMSE) value of each crime category.

Extracted data and actual crime data comparison

To investigate whether crime news available on the Internet through RSS feed can represent the numbers of actual crime data collected by police officers, those 2 sources of data need to be compared and compute the error rate between them. In this work, Root Mean Squared Error (RMSE) is deployed to investigate the quality of the extracted crime data. **Figure 9** depicts the comparison number of crime data of those 2 sources in 2015. It is clearly seen that some categories consistency with actual data such as serious offensive body harm case, crime against property, and interesting case (fraud, deceit, cheat etc.) whereas corrupt state case seems to be diverted with the actual data. After deep investigation in this category, we found that the studied website contains a few number of this news. The major reason because this kind of news is not interested by readers. Then, they ignore to put it into their website. As a result, the number of this crime reported online is far different from the actual cases leading to having highest RMSE up to 867.57 while other types of crimes obtain error rate less than 25 as shown in **Figure 10**.

Example of OLAP report

OLAP performs based on the multidimensional data model or star schema. It allows data analysts to get an insight of the information through fast, consistent, and interactive access to information. **Figure 11** shows the example OLAP report of crime in each area as well as number of criminals and victims in Pattaya in 2015 which users can perform drill-down and roll-up operations. This report can support the decision of police officers to determine the policy to decrease the number of crimes in the risk areas for example, arrange more polices to check more often at the high crime frequency location.



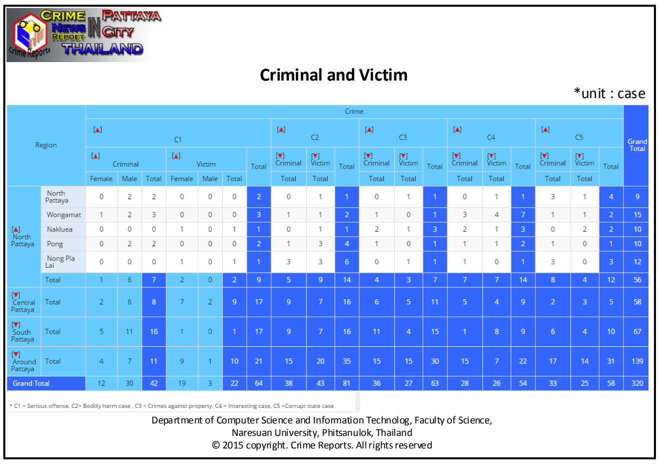
Crime by Area

*unit : case

Region		[▲] 2015						Grand Total
		C1	C2	C3	C4	C5	Total	
[▲] North Pattaya	North Pattaya	6	4	3	5	6	24	24
	Wongamat	3	4	3	8	9	27	27
	Nakluea	4	2	4	4	5	19	19
	Pong	4	7	8	5	4	28	28
	Nong Pla Lai	6	5	2	3	4	20	20
	Total	23	22	20	25	28	118	118
[▼] Central Pattaya	Total	30	35	28	29	25	147	147
[▼] South Pattaya	Total	37	44	46	30	36	193	193
[▼] Around Pattaya	Total	67	73	65	66	81	352	352
Grand Total		157	174	159	150	170	810	810

* C1 = Serious offense, C2= Bodilly harm case , C3 = Crimes against property, C4 = Interesting case, C5 =Corrupt state case

Department of Computer Science and Information Technolog, Faculty of Science,
Naresuan University, Phitsanulok, Thailand
© 2015 copyright. Crime Reports. All rights reserved



Criminal and Victim

*unit : case

Region		Crime																			Grand Total
		[▲] C1							[▲] C2			[▲] C3			[▲] C4			[▲] C5			
		[▲] Criminal			[▲] Victim			Total	[▼] Criminal	[▼] Victim	Total	[▼] Criminal	[▼] Victim	Total	[▼] Criminal	[▼] Victim	Total	[▼] Criminal	[▼] Victim	Total	
		Female	Male	Total	Female	Male	Total		Total	Total		Total	Total		Total	Total		Total	Total		
[▲] North Pattaya	North Pattaya	0	2	2	0	0	0	2	0	1	1	0	1	1	0	1	1	3	1	4	9
	Wongamat	1	2	3	0	0	0	3	1	1	2	1	0	1	3	4	7	1	1	2	15
	Nakluea	0	0	0	1	0	1	1	0	1	1	2	1	3	2	1	3	0	2	2	10
	Pong	0	2	2	0	0	0	2	1	3	4	1	0	1	1	1	2	1	0	1	10
	Nong Pla Lai	0	0	0	1	0	1	1	3	3	6	0	1	1	1	0	1	3	0	3	12
	Total	1	6	7	2	0	2	9	5	9	14	4	3	7	7	7	14	8	4	12	56
[▼] Central Pattaya	Total	2	6	8	7	2	9	17	9	7	16	6	5	11	5	4	9	2	3	5	58
[▼] South Pattaya	Total	5	11	16	1	0	1	17	9	7	16	11	4	15	1	8	9	6	4	10	67
[▼] Around Pattaya	Total	4	7	11	9	1	10	21	15	20	35	15	15	30	15	7	22	17	14	31	139
Grand Total		12	30	42	19	3	22	64	38	43	81	36	27	63	28	26	54	33	25	58	320

* C1 = Serious offense, C2= Bodilly harm case , C3 = Crimes against property, C4 = Interesting case, C5 =Corrupt state case

Department of Computer Science and Information Technolog, Faculty of Science,
Naresuan University, Phitsanulok, Thailand
© 2015 copyright. Crime Reports. All rights reserved

Figure 11 Generated report crime grouped by areas of Pattaya in 2015, Thailand.

Conclusions

This research proposed a prototype system that exploits online crime news to support decision making of police officers to determine security and safety policy for locals and visitors. This unify framework automatically extract crime data from RSS feeds of a newspaper website and restructure them into star schema and stored in data warehouse. This allows users can effectively perform data analysis through drill-down and roll-up operations. Several experiments have been conducted to validate the performance of the presented framework and obtained satisfied results. Different from state-of-the arts, we focus on crime domain and exploited GATE framework to process and analyze news which have never been conducted in the previous literatures. Although the presented system work effectively on this task, it cannot replace the existing system of the police because the presented system can only estimate the numbers of crimes but it cannot replace the official data collected by polices. However, it can be used as a guideline for polices to see overview of crimes and their trends which benefit to resolve crime problems in the risk areas and improve the quality of living of natives and tourists. Another limitation of this work is the grammatical formalism rules are fixed. Thus, these rules might not cover all structures of sentences. As a consequence, the system cannot detect key features in some cases. Our future work could modify these rules to be more dynamic and apply GIS to the framework to plot crime density into map and this allows officers to easier monitor crime situation in the desire areas.

Acknowledgements

This work has been supported by Naresuan University Research Funding. Contract number R2559C073.

References

[1] “South-Eastern Asia: Crime Index by Country 2016 Midyear, Available at: http://www.numbeo.com/crime/rankings_by_country.jsp?title=2016-mid&region=035, accessed September 2016.

[2] S Shojaee, A Mustapha, F Sidi, and A J Marzanah. A study on classification learning algorithms to predict crime status. *Int. J. Digit. Content Tech. Its Appl.* 2013; **7**, 361-9.

[3] Y Yang, JG Carbonell, RD Brown, T Pierce, BT Archibald and X Liu. Learning approaches for detecting and tracking news events. *IEEE Intell. Syst.* 1999; **14**, 32-43.

[4] YW Seo, J Giampapa and K Sycara. *Financial News Analysis for Intelligent Portfolio Management.* Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, Technical Report CMU-RI-TR-04-04, 2004.

[5] J Wanglee, C Thaina, S Yodkaew and L Preechaveerakul. Automatic news aggregator system based on users’ preference. *In*: Proceedings of the Conference of Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology Association of Thailand. Bangkok, Thailand, 2009, p. 155-60.

[6] S Sudhahar, R Franzosi and N Cristianini. Automating quantitative narrative analysis of news data. *In*: Proceedings of the 2nd Workshop on Applications of Pattern Analysis. Castro Urdiales, UK, 2011, p. 63-71.

[7] W Wang, X Cui and A Wang. News analysis based on meta-synthesis approach. *In*: Proceedings of the 32nd Annual IEEE International Computer Software and Applications Conference. Turku, Finland, 2008, p. 923-8.

[8] H Cunningham, D Maynard, K Bontcheva and V Tablan. GATE: An architecture for development of robust HTL applications. *In*: Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics. Philadelphia, 2002, p. 168-75.

[9] Committing Crime Topic from the Oxford Advanced Learner’s Dictionary, Available at: http://www.oxfordlearnersdictionaries.com/topic/committing_crime, accessed September 2016.

[10] Crime Vocabulary, Crime Word List, Available at: https://myvocabulary.com/word-list/crime-vocabulary, accessed September 2016.

[11] P Krongyuth, K Pattanagul, Y Tongrasit, W Chaisiwamongkol, S Ungpansattawong, R Naimsanit and A Maneesriwongul. A multivariate statistical analysis of crime in provincial level of Thailand. *KKU Res. J.* 2013; **18**, 642-50.

[12] T Joachims. A probabilistic analysis of the rocchio algorithm with TFIDF for text categorization. *In*: Proceedings of the 14th International Conference on Machine Learning. San Francisco, USA, 1997, p. 143-51.

[13] JA Swets. Signal Detection Theory and Roc Analysis in Psychology and Diagnostics: Collected Papers, Available at: https://www.questia.com/library/91082318/signal-detection-theory-and-roc-analysis-in-psychology, accessed September 2016.

[14] J Fogarty, RS Baker and SE Hudson. Case studies in the use of ROC curve analysis for sensor-based estimates in human computer interaction. *In*: Proceedings of Graphics Interface. Victoria, British Columbia, 2005, p. 129-36.

[15] K Kesorn, P Ongruk, J Chompoosri, U Thavara, A Tawatsin and P Siriyasatien. Morbidity rate prediction of Dengue Hemorrhagic Fever (DHF) using the Support Vector Machine and the *Aedes aegypti* infection rate in similar climates and geographical areas. *Plos One* 2015; **10**, e0125049.