# Impact of Financial News Headline and Content to Market Sentiment

Tan Li Im, Phang Wai San, Chin Kim On, Rayner Alfred, and Patricia Anthony

*Abstract*—**Business and financial news are important resources that investors referred to when monitoring the stock performance. News brings us the latest information about the stock market. Studies have shown that business and financial news have a strong correlation with future stock performance. Business and financial news can be used to extract sentiments and opinions that may assist in the stock price predictions. In this paper, we present a sentiment analyser for financial news articles using lexicon-based approach. We utilized two most important elements of news, the headline and the content as our test data. We use polarity lexicon to distinguish between positive and negative polarity of each term in the corpus. We further investigate on how news headline will affect the sentiment analysis by adjusting the weights of the news headline and news content's sentiment value. Three sets of experiments were carried out using headline only, content only and headline and content as test data. In the experiment, we used non-stemming tokens and stemming tokens when considering individual word found in the news article. The preliminary results are presented and discussed in this paper.**

*Index Terms*—**Lexicon, news headline, news content, sentiment analysis.**

## I. INTRODUCTION

The availability of online newspaper, made is possible for us to easily access news articles relating to business and financial news. Business and financial news are important resources that can be used by investors to monitor stock performance as it provides the latest information about the stock market.

Studies have shown that business and financial news have a strong correlation with future stock performance. Therefore, it is useful to extract sentiments and opinions from business and financial news. Studies have also shown that business and financial news play important role and can indirectly affect the performance of a particular stock. Most of the investors make decision to either buy or sell a particular stock based on the opinions and sentiments expressed in the news. Hence, a tool that is able to provide financial analysis based on specific news articles is desirable. To be effective, this particular tool or software must be able to analyse multiple news in order to provide financial analysis that will aid investors in their buy or sell decision.

Studies have shown that sentiment analysis plays an important role in creating this required financial analysis software. Sentiment analysis can be divided into two categories: (1) lexicon-based approach and (2) machine learning based approach [1]. The lexicon-based approach requires a prefixed or predefined dictionary or lexicon and it involves computing the orientation for a document from the polarity of words or phrases in the document [2]. This can be done either manually, semi-automatically or automatically [3]. Manually created lexicon makes use of existing dictionary such the General Inquirer Dictionary which contains information about English word senses, including polarity tags such as positive, negative negation, overstatement, or understatement [4]. Semi-automatic lexicon approach involves resources such as WordNet and SentiWordnet [1]. Automatically created lexicon is produced via association [1]. The latter approach involves a supervised classification technique which involves building classifiers from labeled instances of texts or sentences [3]. Naïve Bayes, SVM and N-gram are some of the most popular sentiment classification techniques [5].

In other studies, [6] investigated the correlation between trade figures on the stock exchange and its online media coverage. They showed that the overall market trend correlated with the distribution of general sentiment within articles taken from the corresponding periods of trend. [7] demonstrated the way to extract financial related sentiment values from news sources and using it as a market predictor. In this work, it reported a prediction accuracy of approximately 70%. In [8], the author employed a lexical cohesion based text representation in detecting the sentiment and polarity in news story. [9] utilizes lexicon based approach for discovering sentiments of Twitter's tweet that contains word variations, emoticons and hash tags. However, to date no research has been conducted for extracting headline and content sentiments from news articles for financial analysis purposes. Hence, this has motivated us to conduct this research.

In this paper, we will be focusing on manually created lexicon-based approach. We find the common words used in financial news, classify each word into one of the two sentiment classes (positive or negative), and add it to the lexicon. We use two most important elements of news, the headline and the content as our test data. We use polarity lexicon to distinguish between positive and negative polarity of each term in the corpus. We further investigate how news headline will affect the final value of sentiment analysis. We

Tan Li Im, Pang Wai San, Chin Kim On, and Rayner Alfred are with the Center of Excellence in Semantic Agents, Universiti Malaysia Sabah, Jalan UMS, 88400, Kota Kinabalu, Sabah, Malaysia (e-mail: im_87@hotmail.com, imsan87@gmail.com, kimonchin@gmail.my, ralfred@ums.edu.my).

Patricia Anthony is with the Department of Applied Computing, Faculty of Environment, Society and Design, Lincoln University, Christchurch, New Zealand (e-mail: patricia.anthony@lincoln.ac.nz ).

do this by combining the sentiment values for the news headline content where we adjust their values accordingly to get the desired result. Three sets of experiments were conducted using 1) headline only, 2) content only and 3) headline plus content. In the experiments, we used non-stemming tokens and stemming tokens for the individual words found in the news article.

This paper is organized as follows: In Section II we describe our sentiment analysis using lexicon-based approach. Next, we discuss the experimental evaluation in Section III. In this section, we also analyse the performance of our sentiment analysis algorithm using stemmed text and non-stemmed text. Section IV describes and discusses the result, and in Section V we conclude and describe the future work.

## II. SENTIMENT ANALYSIS APPROACH

In this section, we discuss the algorithm used to implement the lexicon-based approach for financial news articles. This involves four processes: 1) building lexicon, 2) data preprocessing, 3) lexicon matching and polarity detection, and 4) sentiment calculation.

### A. Building Lexicon

The subjectivity lexicon proposed by [8] is used in this research as a guideline in manually creating the required lexicon. In this research, the lexicon used is created based on the norms of business and financial domain. The process of constantly adding new words to the lexicon is needed in order to produce larger domain specific lexicon. A list of new words from financial related articles is generated and each of the words found is tagged as either positive or negative polarity. The list is then added to the existing lexicon. To date, there are approximately 8,000 words that have been tagged and saved in our lexicon. Fig. 1 shows the process of building lexicon.

```
Begin
1.0    Read text from articles
2.0    While (article not empty) loop
    2.1    If (related text found) then
        2.1.1  Tag the whole text (including the noises)
        2.1.2  Add to existing lexicon
    2.2    End if
3.0    End while loop
End
```

Fig. 1. Pseudocode: building lexicon.

### B. Data Preprocessing and Sentiment Analysis

We make use of pre-build lexicon in the sentiment analysis and data preprocessing is required to remove the noises in the text. This is important as the list of words collected involved some noises such as stop words, redundant words and punctuations, which are considered as not important information to the financial analysis. In this step, each financial news article is tokenized into tokens. Then, a parsing algorithm is used to separate the news articles using whitespace. In this research work, the data is preprocessed and then divided into two categories based on non-stemming algorithm and Porter stemming algorithm [11]. The Porter stemming algorithm or popularly known as "Porter Stemmer" is a process for removing the common morphological endings

from words in English. It is considered as a basic normalisation process when setting up information retrieval systems [12]. This is experiment is conducted to compare the performance of stemming and non-stemming algorithms in this research work. Fig. 2 (a) and (b) show the pseudocode for stemming and non-stemming data preprocessing.

```
Begin
1.0    Read news articles from text file
2.0    While (text file is not empty) loop
    2.1    Remove the punctuations
    2.2    Remove all redundant words
    2.3    Tokenize the articles into tokens
    2.4    For (each token) loop
                2.4.1  Match token with stop word list
        2.4.2  If (token matched) then
            2.4.2.1    Remove the stop word
                    2.4.3  End if
    2.5    End for loop
3.0    End while loop
End
```

(a)

```
Begin
4.0    Read news articles from text file
5.0    While (text file is not empty) loop
    5.1    Remove the punctuations
    5.2    Remove all redundant words
    5.3    Tokenize the articles into tokens
    5.4    For (each token) loop
        5.4.1  Match token with stop word list
        5.4.2  If (token matched) then
            5.4.2.1    Remove the stop word
            5.4.2.2    Apply for Porter Stemmer [11]
        5.4.3  End if
                5.5    End for loop
6.0    End while loop
End
```

(b)

Fig. 2. Pseudocode: data processing for non-stemming and stemming algorithms.

### C. Lexicon Matching and Polarity Detection

This stage involves matching the words in the lexicon [13]. Our sentiment analysis approach makes use of the pre-built lexicon. There are approximately 8,000 pre polarity tagged words used. As mentioned, each word in the lexicon is tagged for its polarity of either positive or negative. If the token matches with a word in the lexicon, then the token will be tagged as positive or negative polarity according to the pre-tagged words in the lexicon. Tokens that do not match with the words in the lexicon are ignored in the system. Fig. 3 shows the pseudocode to perform polarity detection task.

```
Begin
1.0    Read tokenized articles from text file
2.0    While (token in not empty in text file) loop
    2.1    Match token with words in lexicon
    2.2    If (token match with words in lexicon) then
        2.2.1  If (word found is positive) then
            2.2.1.1    Tag token as positive
            2.2.1.2    Positive token count +1
        2.2.2  Else if (word found is negative) then
            2.2.2.1    Tag token as negative
            2.2.2.2    Negative token count +1
        2.2.3  End if
    2.3    End if
3.0    End while loop
End
```

Fig. 3. Pseudocode: polarity detection.

### D. Sentiment Calculation

Two formulas called positivity/negativity ratio (P/N ratio) are applied to determine the polarity of the entire news article.

$$P/N\_ratio = -1 * (negative\,count / positive\,count) \quad (1)$$

$$P/N\_ratio = (positive\,count / negative\,count) \quad (2)$$

where *P* is defined as Positivity ratio and *N is* the Negativity ratio. (1) is used for computing the *P/N* ratio only when the positive count is less than or equal to negative count and (2) is used for computing the *P/N* ratio only if the negative count is less than positive count. The news articles will be classified as positive or negative polarity based on above formulas. The *P/N* ratio is always in the range of [-10, 10] based on the polarity.

The formulas are used to measure the number of positive tokens versus the number of negative tokens in an article. In the experiment that considered the news headline and the news content independently, the end result is to get an average sentiment value for that news headline or news content respectively. For example, in the news content, if 100 tokens were detected with its polarity, then the average sentiment value for those 100 tokens is recorded. The same process is applied to the news headline. The algorithm for computing the sentiment value is shown in Fig. 4.

The calculation of the sentiment value for the experiment that used combination of the news headline and the news content is slightly different. Here, we first calculate the sentiment value of the news headline and the news content separately by using the sentiment calculation algorithm shown in Fig. 4. After that, we allocate the weight to the news headline and the news content's sentiment values. The final sentiment values of that particular news article are calculated based on the following formula.

$$Sentiment\,value = (wh * hsv) + (wc * csv) \quad (3)$$

where *wh* represents the weight of the headline, *wc* represents the weight of the content, *hsv* represents the news headline sentiment value, and *csv* represents the news content sentiment value.

We manipulated different combinations of the news headline and the news content sentiment value's weights from 0.9 and 0.1 to 0.1 and 0.9 and selected the weight combination that yielded the best result.

## III. EXPERIMENTS

This section describes the experiments carried out using financial news articles that we collected. Three experiments were conducted for the lexicon-based sentiment analysis algorithm. We tested our algorithm using three sets of data, which are the news headline, the news content, and news headline plus the news content. In the first and second experiments, we used the news headline and the news content as test data respectively. In these two experiments, our aim is to calculate the overall polarity sentiment of the news article

using our sentiment analysis algorithm purely by considering the news headline only and the news content only. We also evaluated the performance of the lexicon algorithm with and without stemming. As mentioned, we performed the stemming using the Porter Stemming algorithm proposed by [11]. In this experiment, the stemmer was applied to both tokens in the news articles and words in the lexicon in order to increase the probability of getting a match.

```
Begin
1.0    Read the positive and negative counts
2.0    If (positive count <= negative count) then
  2.1    Apply ratio = -1 * (negative count/positive count)
3.0    Else if (positive count > negative count) then
  3.1    Ratio = positive count/negative count
4.0    Else if (positive count < negative count) and (positive count = 0)
       then
  4.1    Assign ratio = -10
5.0    Else if (positive count>negative count) and (negative count = 0) then
  5.1    Assign ratio = 10
6.0    End if
7.0    If (ratio > 10) then
  7.1    Assign ratio = 10
8.0    Else if ratio < -10) then
                           8.1    Assign ratio = -10
9.0    End if
End
```

Fig. 4. Sentiment calculation using *P/N* ratio.

The second aim of the experiment is to investigate whether the news headline has an impact on the overall calculation of the sentiment value. We did this by calculating the sentiment values for the news headline and the news content and we applied Equation (3). In our experiment, we used different combinations of weights for the news headline and the news content's to find the best result.

### A. Dataset

In this research, the dataset were collected from the local online newspapers which are *The Star Online*, The *National News Agency of Malaysia (BERNAMA)*, *The Sun Daily*, and *The New Straits Times (NST)*. However, we only collected the financial and business related news articles. These news articles contain news related to individual stock in Kuala Lumpur Stock Exchange (KLSE) or also known as Bursa Malaysia. The topics include, but not limited to earning announcements, quarterly reports, annual reports, and product unveilings. These articles were collected in the first half of 2013. Each news article contains two elements which are the news headline and the content of the news. Approximately 200 financial news articles were collected and used in the experiments.

### B. Gold Standard

A standard measurement of news articles polarity is required to measure the performance of the proposed sentiment analyser. In this work, the articles were manually annotated as negative or positive polarity and then the dataset are categorized into positive and negative polarity. The polarity of the news article could be concerned with whether the outcome of the news is considered good or bad. Increment of profit for a company will be considered a positive polarity and a decrease or lost in profit of one company is considered

as a negative polarity.

### C. Performance Metrics

Another dimension of the measurement is the sentiment values of the news articles which focused on the task of assigning a value to indicate the strength of the positive or the negative polarity in the news article [10].

The performance of the sentiment analysis algorithm is reported using positive predictive value as shows in (4), negative predictive value as shows in (5), sensitivity as shows in (6), specificity as shows in (7) and also the overall accuracy as shown in (8) of the algorithm for correctly predicting the polarity of the news articles.

$$Positive\ predictive = \frac{\sum True\ positive}{\sum Test\ outcome\ positive} \quad (4)$$

$$Negative\ predictive = \frac{\sum True\ negative}{\sum Test\ outcome\ negative} \quad (5)$$

$$Sensitive = \frac{\sum True\ positive}{\sum Condition\ positive} \quad (6)$$

$$Specific = \frac{\sum True\ negative}{\sum Condition\ negative} \quad (7)$$

$$Accurate = \frac{No.of\ (True\ negative + True\ Negative)}{No.of\ (True\ negative + false\ positive + True\ Negative + false\ negative)} \quad (8)$$

## IV. RESULTS AND DISCUSSION

We conducted three experiments for the lexicon-based sentiment analysis; the first one is to determine the news article's sentiment by using the news headlines only, the second is to determine the sentiment based on the news content only, and the third one by combining the sentiment of the news headline and the news content. We also evaluated the performance of the lexicon-based algorithm in two ways, one without stemming and another with stemmed tokens.

Table I shows the results of the experiment that used the news headline as the test data. The positive predictive value recorded high percentage in both non-stemming and stemming algorithm which achieved 95.1% and 98.5% respectively. The algorithm performed poorly in predicting the negative polarity using news headline. The negative predictive value for non-stemming and stemming recorded 42.3% and 53.8% respectively. The negative prediction values are low in both cases as some articles may consist of negative tagging for headline but slightly positive tagging in the content. As an example, one of the article headline was ". . . XX company loss 5 million in the 3rd quarter of 2010 ." and in the article content was ". . . the profit will be increased in the last quarter as the XX company has just signed an MOU with YY company that estimated to have an increment of 40% profit compared to last year ". The true positive rate for non-stemming lexicon-based approach recorded 47.5% whereas the true negative rate recorded 96.0%. This is due to as our dataset having a lot of positive articles and very few negative polarity articles. The stemming algorithm recorded an overall accuracy of 71.4% as opposed to 65.8% recorded by the non-stemming algorithm. Based on these observations, we can conclude that using stemming is better when we have to deal with the news headline only.

TABLE I: SUMMARY OF RESULTS USING NEWS HEADLINE

| Measurement (%) | Non Stemming | Stemming |
|---|---|---|
| Positive Predictive Value | 95.1 | 98.5 |
| Negative Predictive Value | 42.3 | 53.8 |
| Sensitivity (True positive rate) | 47.5 | 54.9 |
| Specificity (True negative rate) | 96.0 | 98.7 |
| Accuracy | 65.8 | 71.4 |

TABLE II: SUMMARY OF RESULTS USING NEWS CONTENT

| Measurement (%) | Non Stemming | Stemming |
|---|---|---|
| Positive Predictive Value | 76.7 | 82.4 |
| Negative Predictive Value | 68.3 | 73.2 |
| Sensitivity (True positive rate) | 83.6 | 84.4 |
| Specificity (True negative rate) | 58.1 | 70.3 |
| Accuracy | 74.0 | 79.1 |

TABLE III: SUMMARY OF RESULTS USING NEWS HEADLINE AND NEWS CONTENT

| Measurement (%) | Weightage (Headline + Content in %) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 10+ 90 | 20+ 80 | 30+ 70 | 40+ 60 | 50+ 50 | 60+ 40 | 70+ 30 | 80+ 20 | 90+ 10 | Headline only | Content only |
| Positive Predictive Value | 85.6 | 90.7 | 93.1 | 93.9 | 96.4 | 96.4 | 96.4 | 96.4 | 96.4 | 98.5 | 82.4 |
| Negative Predictive Value | 78.8 | 80.8 | 82.5 | 82.7 | 83.3 | 83.3 | 83.3 | 83.3 | 83.3 | 53.8 | 73.2 |
| Sensitivity (True positive rate) | 87.7 | 87.7 | 88.5 | 88.5 | 88.5 | 88.5 | 88.5 | 88.5 | 88.5 | 54.9 | 84.4 |
| Specificity (True negative rate) | 75.7 | 85.1 | 89.2 | 90.5 | 94.6 | 94.6 | 94.6 | 94.6 | 94.6 | 98.7 | 70.3 |
| Accuracy | 83.2 | 86.7 | 88.8 | 89.3 | 90.8 | 90.8 | 90.8 | 90.8 | 90.8 | 71.4 | 79.1 |

Table II shows the results of using stemmed and non-stemmed algorithm for just the news content. It can be observed that the positive predictive value for non-stemming and stemming lexicon-based approach recorded 76.7% and 82.4% respectively. In addition, the negative predictive value for non-stemming lexicon-based approach gives 68.3% whereas stemming lexicon-based approach recorded 73.2%. As expected, the sentiment analyser that utilized stemming

algorithm achieved a higher overall accuracy of 79.1%.

The lexicon-based approach sentiment analysis algorithm using news content performs better in assigning positive news articles compared to negative news articles. The true positive rate for non-stemming lexicon-based approach recorded 83.6% whereas the true negative rate recorded only 58.1%, which is much lower that the true positive rate. This is because of the nature of the news articles itself. Reporters tend to write

something positive in the news even though the news is actually negative. It appears that there are more positive words than negative words in negative news articles. Thus the proposed algorithm can be considered as positive bias.

Table III illustrates the results of using combinations of news headline and news content with Porter Stemming showed better results. We manipulated the news headline and news content's weightage from 90% to 10% and from 10% to 90% respectively. It can be seen that our algorithm's accuracy improved from 71.4% (using the news headline only) and 79.1% (using the news content only) to 90.8%. We can also observe that by using a weight greater than or equal to 50% for the news headline, the result remain constant. In this experiment, we are satisfied that equal weights of 0.5 for the news content and the news headline give the best value.

## V. CONCLUSION AND FUTURE WORKS

This paper presented the used of lexicon-based approach for analyzing financial news sentiment, particularly for the positive and negative polarity. We proposed a lexicon-based approach to identify and extract sentiments from financial news articles. We also proposed to add weights to the news headline and the news content to generate a better result.

Three experiments were conducted involving news headline, news content and combinations of news headline and news content. We evaluated our lexicon-based algorithm in two ways, one without stemming and another with stemmed tokens. In the experiment using the news headline, the accuracy for non-stemmed token recorded 65.8% and stemmed token recorded 71.4%. In the experiment using the news content, non-stemmed token algorithm achieved 74% and stemmed token algorithm achieved 79.1% accuracy. The results obtained from these two experiments show that stemming the news produced better accuracy. It also reflects that the news content yielded better accuracy compared to news headline.

In the third experiment, we proposed to use both the news headline and the news content as the test data by adding weights to their sentiment values. The results show that this approach improved our previous lexicon-based approach's performance. Using a weight combination of 0.5 and 0.5 for the news headline and the news content, the algorithm achieved a high accuracy of 90.8%, which is much higher than the result obtained when using the news headline only (71.4%) or the news content (79.1%) only.

A lexicon-based approach is a simple, robust solution in sentiment analysis. Currently, our algorithm only matches one word at a time. One of the drawbacks of this algorithm is that the word before or the word after may affect the sentiment of that particular word e.g. negation. Also, we cannot easily detect the sentiment by considering just a single word. There are many instances where a sentiment is analysed by looking at multiple words or phrase in a sentence such as "He is a good liar.", "The bank reduces interest." For future work, we will examine the lexicon-based approach by analyzing phrases in the news articles rather than just looking at individual words. We would also like to investigate the possibility of using SentiWordnet lexicon to evaluate different dimensions of polarity such as polarity intensity which measures the strength of negative/positive polarity in a text.

## REFERENCES

[1] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Computational Linguistics*, in press.

[2] P. D. Turney and M. L. Littman, "Unsupervised learning of semantic orientation from a hundred-billion-word corpus," Technical Report EGB-1094, National Research Council Canada, 2002.

[3] Q. Ye, Z. Zhang, and R. Law, "Sentiment classification of online reviews to travel destinations by supervised machine learning approaches," *Expert system with application: An introduction Journal*, vol. 35, issue 3, April, 2009, pp. 6527-6535.

[4] P. J. Stone, D. C. Dunphy, M. S. Smith, and D. M. Ogilvie, *The General Inquirer: A Computer Approach to Content Analysis*, MIT Press, Cambridge, MA, 1966.

[5] A. Kennedy and D. Inkpen, "Sentiment classification of movie reviews using contextual valence shifters," *Computational Intelligence*, vol. 22, pp. 110-125, 2006.

[6] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," in *Proc. the Conference on Empirical Methods in NLP*, Philadelphia, PA. 2002, pp. 79-86.

[7] Z. Agic, N. Ljubesic, and M Tadic, "Towards sentiment analysis of financial texts in croatian," in *Language Resources and Evaluation (LREC)*, N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner and D. Tapias, Eds. European Language Resources Association.

[8] J. J. Zhai, N. Cohen, and A. Atreya, "Sentiment analysis of news articles for financial signal prediction," unpublished.

[9] The Subjectivity lexicon. [Online]. Available: http://mpqa.cs.pitt.edu/

[10] M. Losada and E. Heaphy, "The role of positivity and connectivity in the performance of business teams: a nonlinear dynamics model," *American Behavioral Scientist*, vol. 47, no. 6, pp. 740-765, 2004.

[11] The Porter Stemmer. [Online]. Available: http://www.ling.gu.se/~lager/mogul/porter-stemmer/

[12] A. Devitt and K. Ahmad, "Sentiment Polarity Identification in Financial News: A Cohesion-based Approach," in *Proc. the Association for Computational Linguistics (ACL)*, pp. 984 Č991, 2007.

[13] P. Palanisamy, V. Yadav, and H. Elchuri, "Serendio: simple and practical lexicon based approach to sentiment analysis," SemEval-2013: Semantic Evaluation Exercises International Workshop on Semantic Evaluation, Atlanta.

**Tan Li Im** is currently pursuing her master degree in computer science with the Center of Excellent in Semantic Agents under School of Engineering and Information Technology, in Universiti of Malaysia Sabah, Sabah, Malaysia. The author's research interests include agent technologies, sentiment analysis, stock market prediction and data mining.
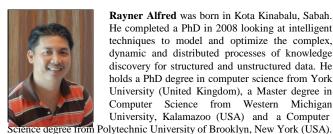
**Phang Wai San** is currently pursuing her master degree in computer science with the Center of Excellent in Semantic Agents under School of Engineering and Information Technology, in Universiti of Malaysia Sabah, Sabah, Malaysia. The author's research interests include agent technologies, neural networks, backpropagation, and stock market prediction.

**Chin Kim On** received his PhD in artificial intelligence with the Universiti of Malaysia Sabah, Sabah, Malaysia. The author's research interests included gaming AI, evolutionary computing, evolutionary robotics, neural networks, image processing, semantics based visual information retrieval, agent technologies, evolutionary data mining and biometric security system with mainly focused on fingerprint and voice recognition.

He is currently working as a senior lecturer at the Universiti Malaysia Sabah in the School of Engineering and Information Technology, Sabah,

Malaysia. He has authored and co-authored more than 60 articles in the forms of journals, book chapters and conference proceedings. He is a member of IEEE and IAENG societies.

**Rayner Alfred** was born in Kota Kinabalu, Sabah. He completed a PhD in 2008 looking at intelligent techniques to model and optimize the complex, dynamic and distributed processes of knowledge discovery for structured and unstructured data. He holds a PhD degree in computer science from York University (United Kingdom), a Master degree in Computer Science from Western Michigan University, Kalamazoo (USA) and a Computer Science degree from Polytechnic University of Brooklyn, New York (USA).

Dr. Rayner leads and defines projects around knowledge discovery and information retrieval at Universiti Malaysia Sabah. One focus of Dr. Rayner's work is to build smarter mechanism that enables knowledge discovery in relational databases. His work addresses the challenges related to big data problem: How can we create and apply smarter collaborative knowledge discovery technologies that cope with the big data problem.

.

Dr. Rayner has authored and co-authored more than 75 journals/book chapters and conference papers, editorials, and served on the program and organizing committees of numerous national and international conferences and workshops. He is a member of the Institute of Electrical and Electronic Engineers (IEEE) and Association for Computing Machinery (ACM) societies.

**Patricia Anthony** received her PhD in computer science from the University of Southampton in 2003. She is currently working as a senior lecturer at the Department of Applied Computing, Lincoln University, New Zealand. Her research interest is in semantic agents and multi-agent systems and how these agents can interact with each other within an open domain to solve problems. She is also interested in investigating how agents can communicate with each other at the semantic level using semantic technology. To date, she has published more than 80 articles in the forms of journals, book chapters and conference proceedings. She is a member of IEEE, ACM and IACSIT.