International Journal of Engineering & Technology, 7 (3.20) (2018) 864-869

SPC

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

The Identification of the Top Agro Food Keywords in The Star Online Newspaper

Mohamad Farhan Mohamad Mohsin[1], Siti Sakira Kamaruddin[2], Fadzilah Siraj[3], Hamirul Aini Hambali[4] Mohammed Ahmed. Taiye[5]

School of Computing, College of Arts & Sciences, Universiti Utara Malaysia, Kedah, Malaysia
**Corresponding Author Email: farhan@uum.edu.my*

Abstract

News websites produce thousands of articles and the number is growing. To publish up to date breaking news, each news providers apply different styles of journalism and as consequent readers may find different words and concepts used although when reporting a similar event. In this study, we examined the Malaysian online newspaper in order to seek the most regular keywords used in daily online news. To achieve that, the Star Online newspaper was chosen and the news related to Malaysian Agro food industries was selected for mining. During the mining, 173 news articles were scrapped from the website within the time frame of 2008-2017 and the keyword were mine using RAKE algorithm. From the analysis, 51 frequently used agro food related keywords has been identified and fish is the top keyword used in The Star Online. The other popular keywords found were palm oil, rice, fruits, livestock, vegetables, chicken, and crops. It is also noticed that different terms were used to describe similar concept. The indentified keywords can be used to form future agro inventory. In future work, the other Malaysian Online newspaper such as The New Strait Time and The Sun will be further explored and comparison of the top agro food keyword will be made among them.

Index Terms*: Agro, text mining, online news, news mining, RAKE algorithm*

1. Introduction

Internet has turned out to be one of the successful channels for web clients to convey and look for data, for example, read the online news.Because of its availability and affordable devices, internet creates opportunities for online users to get update with latest breaking news in a timelier manner. It has become a huge part of the society to refer to the online news and it is now overwhelmed with large volume of recent articles which are generated and updated every day. Although the online news does not provide a detail story about an event but it gives a quick synopsis to inform people what actually happened[1].

News is a report of a current event that contains information about something that has just happened or will happen soon [2]. News represents type of information that citizen needs and it is a reflection of organizational, sociological, and cultural combined with economic factors [3]. It can be shared to public via newspaper, television, radio or news in internet. The online news websites on the web are many. Such an outstanding worldwide online news sites are Yahoo! News,

Google News, Huffington Post, CNN, NY Times, Fox News and NBC News. Agreeing eBizMBA[4] positioning in July 2017, the Yahoo! News is most prominent online news sites with 175,000,000 perusers and it was trailed by Google News with 150,000,000 perusers. Google News is a PC produced news benefit that totals features from in excess of 50,000 news sources around the world, bunch comparative stories together, and shows them as indicated by every peruser's advantages. The administration covers news articles showing up inside the previous 30 days on different news sites. Altogether, Google News totals content from in excess of 25,000 distributers including Malaysian news distributers. In Malaysia, there are 29 papers suppliers that give online news form as appeared Table 1.

Table I: Online News Providers in Malaysia

#	News Providers	Descriptions / online news name	URL Address
1	Berita Harian	National Malay newspaper, BH Online	www.bharian.com.my
2	Utusan Malaysia	National Malay newspaper, Utusan Online	www.utusan.com.my
3	Bernama	Malaysian National News Agency	www.bernama.com
4	Borneo Post	Sabah, Sarawak, East Malaysia newspaper, Utusan Borneo	www.theborneopost.com
5	China Press	National Chinese newspaper	www.chinapress.com.my

6	Daily Express Sabah	East Malaysian newspaper	www.dailyexpress.com.my
7	Digital News Asia	Technology & IT industry news, DNA	www.digitalnewsasia.com
8	Free Malaysia Today	Online news portal, FMT	www.freemalaysiatoday.com
9	Guang Ming Daily	National Chinese newspaper	www.guangming.com.my
10	Harian Metro	National Malay newspaper, myMetro	www.hmetro.com.my
11	Ipoh Echo	Ipoh community news	www.ipohecho.com.my
12	Malay Mail Online	Online news portal with sister newspaper	www.themalaymailonline.com/
13	Malaysia Chronicle	Online news portal	www.malaysia-chronicle.com
14	Malaysiakini	Online news portal, malaysiakiniTV	www.malaysiakini.com
15	MyCen News	Aggregated Malaysian, regional and world news, also science, health, lifestyle & tech	www.mycen.com.my/news
16	Nanyang Siang Pau	National Chinese newspaper	www.nanyang.com
17	New Sarawak Tribune	Sarawak daily	www.newsarawaktribune.com
18	New Straits Times	National English newspaper, NST	www.nst.com.my
19	Oriental Daily	National Chinese newspaper	www.orientaldaily.com.my
20	Overseas Chinese Daily News	Sabah Chinese newspaper, OCDN	www.ocdn.com.my
21	Sin Chew	National Chinese newspaper, mySinChew	www.sinchew.com.my www.mysinchew.com
22	Tamil Nesan	Tamil news paper	www.tamilnesan.com.m
23	The Ant Daily	Online news & commentary	www.theantdaily.com
24	The Edge Markets	Business, share market & financial news	www.theedgemarkets.com
25	The Heat Online	Online news & commentary	www.theheatonline.asia
26	The Malaysian Insider	Online news portal, TMI	www.themalaysianinsider.com
27	The Rakyat Post	Online news portal	www.therakyatpost.com
28	The Star	National English newspaper, Staronline	www.thestar.com.my
29	The Sun Daily	National English newspaper, theSundaily	www.thesundaily.my

The volume of online news articles are huge since they are generated daily, thus the processing and analysis of those data become more challenging to discover hidden knowledge [5]. News sites create a huge number of articles covering wide territory which can be considered as large information. To process such an enormous volume of information, huge information systems are plausible to convey the outcomes inside restricted run times. In this unique situation, huge information online applications are connected with Internet content or records or Internet seek ordering. Subsequently, inquire about territory that are especially imperative for this situation is content mining connected to web news mining. The preparing and investigation of this substantial corpus of information is a vital test. This test should be handled by utilizing huge information methods which process substantial volume of information inside constrained run times. Additionally, since we are going into an online life information blast, methods, for example, content mining or interpersonal organization investigation should be truly thought about

At the point when an occasion happened, news suppliers dependably battle to give perusers a la mode breaking news and every one of them have diverse news-casting styles in distributing it. Therefore, perusers may discover distinctive words and ideas utilized in various news suppliers yet typically they are alluding to comparative setting. Other than that issue, the current inventories and databases partook in open information vaults are less enlightening and inclined to moderate refreshing rate it may not profit the partners to decide. Preliminary investigation was carried out to search and download database from the Government of Malaysia's Open Data Official Portal (http://data.gov.my). The investigation revealed that most of the shared data are summaries of data which were last updated 6 months ago. Based on the assumption that the news is generated everyday in big volume, there is a possibility to construct a database from the available news.

In this investigation, we analyzed the Malaysian online paper to look for the most customary catchphrases utilized in every day online news which can possibly shape a database dependent on the mined idea and watchword. Watchwords are characterized as words or short expressions that speak to the substance of a bigger content. To accomplish that, the Star Online paper was picked and the news about Malaysian agro sustenance businesses was chosen for mining. The online news identified with Malaysian Agro businesses are distributed day by day and the data from the news is ceaselessly developing and always showing signs of change quick. With the web news mining innovation, the concealed idea applicable to Argo Industry can be removed to frame another learning for sometime later. The distinguishing proof of the agro sustenance is term recurrence that has the most elevated term, where it is utilized (news article) and what sort of data it depicts in the given article. These news articles were rejected from the web with a predefined depiction of time allotment, that is from what date the data was removed and the point of dialog in the article is engaged. Along these lines, a versatile dataset that ready to be intermittently refreshed from different recourses identified with Argo, for example, every day news is required.

This paper is composed as pursues. Segment II is the related works which includes three talk; the online news article mining, The Star Online Newspaper and the Agro Food Industries. At that point, the strategy used to lead this investigation is talked about in Section III. In Section IV, the discoveries and dialog of the examination is exhibited. The last segments close this work..

2. Related Works

This section discusses the recent work on online news mining, the background of agro food industries, and The Star Online news website.

A. Online News Article Mining

News mining is under the content mining territory where it includes the way toward extricating fascinating example from content reports, for example, from online news. Since news contains most recent and current refresh on an occasion, it gives contribution from numerous gatherings to basic leadership. For instance, news mining innovation has been broadly utilized in financial, for example, to figure between day stock costs dependent on the practices that originated from occasion provided details regarding the mass media[6-8]. In accordance with this, a securities exchange examination expectation framework has been produced that can figure the share trading system condition whether it will fall or rise. In view of news things, the framework distinguishes and portrays real occasions that affect the market.

To enhance business execution, [9] used web news articles identified with sun based cells to recognize powerless flag subjects by abusing catchphrase based content mining. They trusted that the powerless signs are early pointer of basic occasions or patterns to plan new potential business thought. Other than that, [10] utilized the intensity of online news mining to find fascinating contender connections dependent on the event of the organization been refered to in online news articles. Different works identified with business and enhancement of client relationship that used news mining are [11, 12] .

B. The Star Online

The Star Online (http://www.thestar.com.my) is Malaysia's first news site which was eaten on June 23, 1995. With the plan to furnish perusers with exceptional breaking news and thorough data, the Star Online spreads current news, business, sports, network, innovation, property, work, world news, and way of life. In 2014, this site has been granted as truly outstanding in Asia by the World Association of Newspapers and News Publishers (WAN-IFRA). As to fabricate solid faithful association with its online network perusers, the Star interfaces them through Twitter and Facebook. Advancing with times, The Star Online additionally offers its substance through The Star ePaper and a portable application rather than the regular printed paper. Other than that, perusers' in a hurry who need breaking news and business refreshes conveyed to them can decide on their SMS administrations accessible by means of Maxis and DiGi

C. Agro Food Industries

Agro-nourishment enterprises alludes to the matter of creating agronomically based sustenance that covers a wide scope of exercises that using homestead, creature and ranger service based items as crude materials. The business can be ordered into two; on the land that alludes to horticulture in nature where its gather is the last item. Furthermore is on the table where they are prepared where the gather progress toward becoming sustenances [13]. There are an expansive scope of agro nourishment wares such grains and heartbeats; oilseeds and nuts; leafy foods; roots and tubers; meat and dairy items; nectar and sugar; flavors and stimulants.

In Malaysia, agro-nourishment division is one of the national key outcome regions. It is observed by the Ministry of Agriculture and Agro-Based Industry which capable to configuration, facilitate and guarantee the execution of the agrarian improvement agro-sustenance program. Under the Malaysian National Agro Food Policy (2011-2020), the legislature is quick to guarantee the accessibility, moderateness and openness of sustenance security and wellbeing and in addition the intensity and maintainability of the agro-nourishment industry.

Stock of agro-nourishment is characterized as dataset, database or gathering of data that depicts specific occasion of agro sustenance businesses, for example, add up to income, import and fare data which can be utilized for examination. It very well may be as single database table where each section of the table speaks to a specific variable, and each line relates to a given individual from the informational index being referred to.

3. Methodology

This section explains the method used to conduct the research and how the experiment was set up. To identify the top Agro food product terms in the Star online news article, this study applied the data mining approach which is divided into 3 main phases, as shown in Figure 1. The phases are news extraction, pre-processing, and news mining.

D. News Extraction

This phase involves scrapping news article from The Star Online website. These news articles are basically related to the agricultural events, policies and development made by the Malaysian Government raging from various sectors and product of agricultural implementation of the country. During scrapping, the keyword ’agro food’ was used as a search key to filter news related to agro. Only the agro articles from March 2008 – Jun 2017 (116 months) was scrapped from the website using a scrapper written in Python.

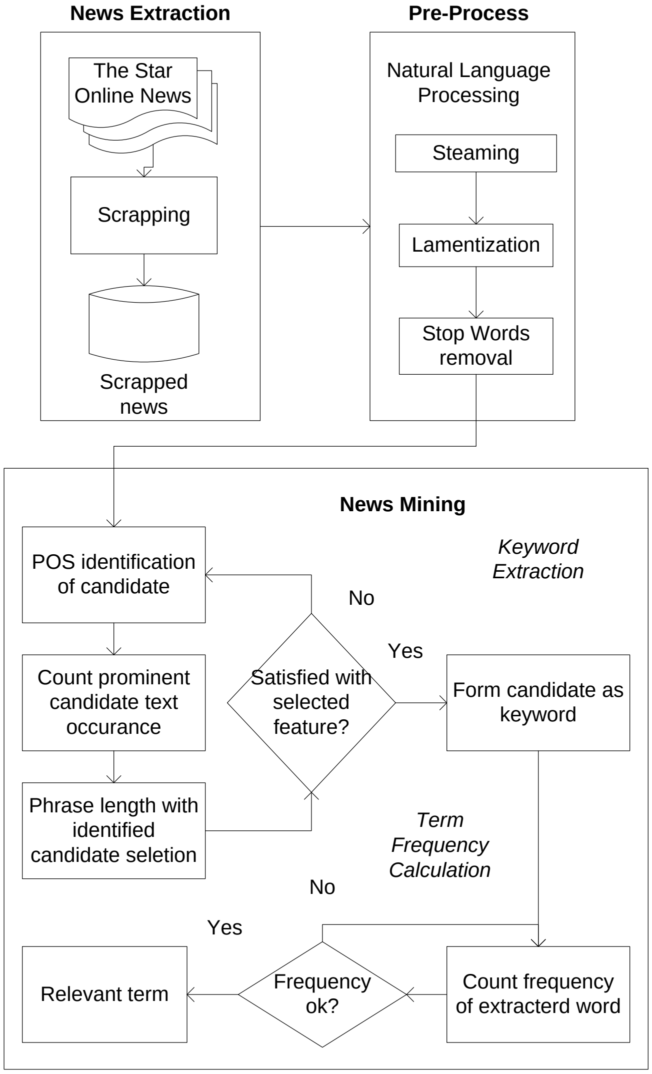

News Extraction
Pre-Process
The Star Online News
Scrapping
Scrapped news
Natural Language Processing
Steaming
Lamentization
Stop Words removal
News Mining
Keyword Extraction
POS identification of candidate
Count prominent candidate text occurance
Phrase length with identified candidate seletion
Satisfied with selected feature?
No
Yes
Form candidate as keyword
Term Frequency Calculation
No
Yes
Relevant term
Frequency ok?
Count frequency of extracterd word

Figure 1: The methodology

Fig. 2 is a diagrammatic illustration of target maps that needs to be extracted from the provided news articles. The extracted information will then be processed and finaly analysed for knowledge identification and business excellence All the news articles were stored as raw data in csv format called as the ‘StarOnline.csv. It is then presented to the next phase.

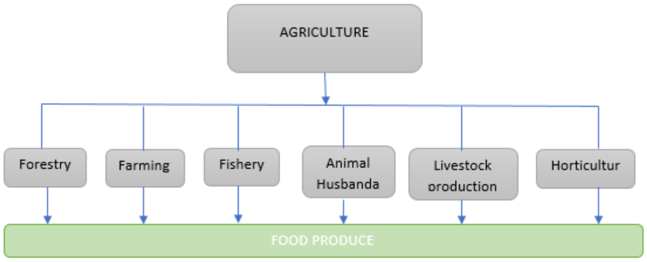


Figure 2: Extraction map

E. Pre-Process

Pre-process phase involves the process of preparing data for mining where the scrapped news was preprocessed to remove unnecessary item and symbol. In this phase, the natural language processing (NLP) was used to clean the input file. There were three steps; streaming, lemmatization, and removing stop word.

F. News Mining

The news articles were mined in this phase. The process was split into two phases that were keyword extraction and term frequency calculation.

1) Keyword extraction

Extracting keywords is one of the most important tasks when working with text. By definition, keywords describe the main topics expressed in a document. In this study, we focused on two specific tasks and their evaluation:

- Extracting the most significant words and phrases that appear in given text
- Identifying a set of topics from a predefined vocabulary that match a given text

There are three main components in keyword extraction algorithms which are candidate selection, properties collection, and scoring and selecting keywords. Candidate selection process involves extracting all possible words, phrases, terms or concepts that can potentially be keywords. After that, it is followed by properties calculation that count prominent candidate text occurrence. It means that, for each possible candidate, the algorithm will count properties that indicate that it may be a keyword for agro food. Lastly is the scoring and selecting keywords process where each candidate is given a score. All candidates can be scored by either combining the properties into a formula, or using a machine learning technique to determine probability of a candidate being a keyword. A score on the number of keywords is then used to select the final set of keywords.

2) Term Frequency Calculation

Term frequency is frequently utilized in data recovery and content mining. It is a factual measure used to assess how imperative a word is to a record in an accumulation or corpus. The significance builds relatively to the occasions a word shows up in the record however is balanced by the recurrence of the word in the corpus. Ordinarily, the tf-idf weight is formed by two terms: the first registers the standardized Term Frequency (TF), otherwise known as. The occasions a word shows up in an archive, partitioned by the aggregate number of words in that report. The second term is the Inverse Document Frequency (IDF), processed as the logarithm of the quantity of the reports in the corpus isolated by the quantity of records where the explicit term shows up.

- TF: Term Frequency, which estimates how as often as possible a term happens in an archive. Since each report is diverse long, it is conceivable that a term would seem substantially more occasions in long archives than shorter ones. In this manner, the term recurrence is regularly isolated by the archive length (otherwise known as. the aggregate number of terms in the report) as a method for standardization: TF(t) = (Number of times term t shows up in a record)/(Total number of terms in the archive).
- IDF: Inverse Document Frequency, which estimates how imperative a term is. While processing TF, all terms are considered similarly imperative. Anyway it is realized that specific terms, for example, "is", "of", and "that", may show up a great deal of times yet have little significance. Hence we have to overload the successive terms while scale up the uncommon ones, by registering the accompanying: IDF(t) = log_e(Total number of archives/Number of records with term t in it).

To mine the news, we embraced the Rapid Automatic Keyword Extraction (RAKE) [14] which incorporates extricating the catchphrase and compute the tem recurrence as appeared in Algorithm 1. RAKE is area freedom where it ready to work autonomously on reports without alluding to a corpus, and has great execution in term on accuracy, effortlessness and computational productivity. It attempts endeavors to decide enter states in an assemblage of content by investigating the recurrence [15] of word appearance and its co-occurance with different words in the content. The outcomes from the calculation were descendingly arranged.

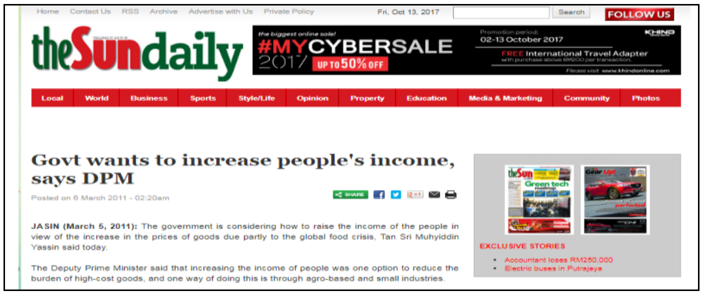
```
Input: Scrapped news from website
Output: A list of the most occurrence keywords
Start
Scrap news from websites
Pre-process the news to filter out unwanted text
    - Split the document into an array of words, breaking it at word
      delimiters (like spaces and punctuation).
Split the words into sequences of contiguous words, breaking each
sequence at a stopword. Each sequence is now a “candidate
keyword”.
Calculate the “score” of each indivudual word in the list of
candidate keywords. This is calculated using the metric
        degree(word)/frequency(word)

For each candidate keyword, add the word scores of its constituent
words to find the candidate keyword score.
Take the first one-third highest scoring candidates from the list of
candidates as the final list of extracted keywords.
Sort the extracted keywords
End
```

Algorithm 1: The Rake Psedocode [14]

4. Result & Discussion

The aim of this study to identify the top agro food related keywords that mostly used in The Star Online newspaper. During news extractions, there were 173 news articles scrapped from website within the time frame 2008-2017. Fig. 3 shows the sample of the original agro food related news article from The Star Online on March 11, 2011. Using scrapper, the articles was extracted as csv file as depicted in Fig. 4.



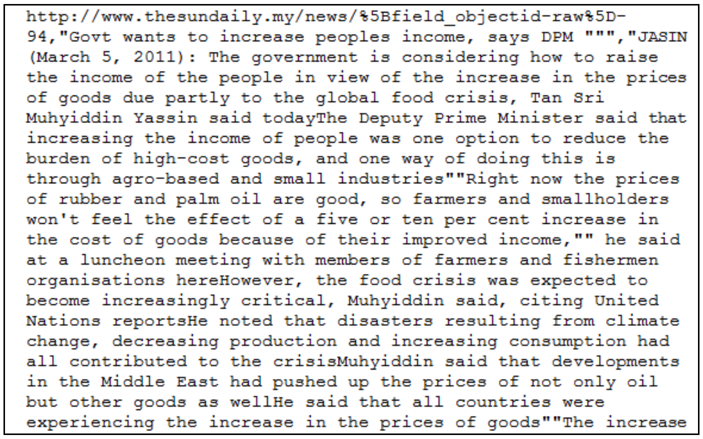
Govt wants to increase people's income, says DPM

Posted on 6 March 2011 - 02:20am

JASIN (March 5, 2011): The government is considering how to raise the income of the people in view of the increase in the prices of goods due partly to the global food crisis, Tan Sri Muhyiddin Yassin said today.

The Deputy Prime Minister said that increasing the income of people was one option to reduce the burden of high-cost goods, and one way of doing this is through agro-based and small industries.

Figure 3: The agro food related article from The Star Online on March 11, 2011



```
http://www.thesundaily.my/news/%5Bfield_objectid-raw%5D-
94,"Govt wants to increase peoples income, says DPM """,""JASIN
(March 5, 2011): The government is considering how to raise
the income of the people in view of the increase in the prices
of goods due partly to the global food crisis, Tan Sri
Muhyiddin Yassin said todayThe Deputy Prime Minister said that
increasing the income of people was one option to reduce the
burden of high-cost goods, and one way of doing this is
through agro-based and small industries""Right now the prices
of rubber and palm oil are good, so farmers and smallholders
won't feel the effect of a five or ten per cent increase in
the cost of goods because of their improved income,"" he said
at a luncheon meeting with members of farmers and fishermen
organisations hereHowever, the food crisis was expected to
become increasingly critical, Muhyiddin said, citing United
Nations reportsHe noted that disasters resulting from climate
change, decreasing production and increasing consumption had
all contributed to the crisisMuhyiddin said that developments
in the Middle East had pushed up the prices of not only oil
but other goods as wellHe said that all countries were
experiencing the increase in the prices of goods""The increase
```

Figure 3: The extracted news in csv file

The mining result is presented in Table 2 with 51 keywords. Form the table, the fish is the most frequent terms used in the The Star online with 24 counts and it was closely followed with palm oil in the second place with 22 counts [16]. Rice places as the third with 19 counts and followed with fruits, livestock, vegetable, corps, and chicken with the number of counts is slightly lower with rice. This indicate fish, palm oil, rice, fruits, livestock, vegetable, corps, and chicken are among the most cases that were reported in the Star Online articles within the studied period [17]. There are other agro keywords that appear in the articles but there is less [18].

Table 2: Top agro food related keywords that mostly appears in The Start Online newspaper

#	Agro Food Related Terms	Term Freq
1	Fish	24
2	palm oil	22
3	Rice	19
4	Fruits	18
5	Livestock	16
6	Vegetables	16
7	Chicken	14
8	Crops	12
9	Durian	8
10	Farmers	8
11	paddy	8
12	Padi	7
13	seafood	7
14	Beef	7
15	poultry	5
16	breeders	5
17	coffee	5
18	seeds	4
19	meat	4
20	farms	4
21	cattle	4
22	herbs	3
23	plants	3
24	food	3
25	prawn	1
26	animals	1
27	mussels	1
28	beef	1
29	cockle	1
30	biodiseal	1
31	fields	1
32	porks	1
33	orange	1
34	park	1
35	rossle	1
36	beverages	1
37	bird	1
38	wildlife	1
39	mutton	1
40	frozen	1
41	seedlings	1
42	onion	1
43	shellfish	1
44	lamb	1
45	orchard	1
46	crude	1
47	turtle	1
48	sea	1
49	citrus	1
50	mangosteen	1
51	elephant	1

In the list, there are different terms that describe similar concept of agro such:

- ‘paddy and ‘padi,
- ‘meat’ and ‘beef’
- ‘lamb’ and ‘mutton’
- ‘chicken’ and ‘poultry’

5. Conclusion

In this study, we have identified the highest number of occurrence of the agro food keywords in The Star Online newspaper. 173 articles related to agro food were scrapped from the web and the RAKE algorithm was used to mine the top keywords. From the analysis, 51 agro food keywords were identified and fish is the top keyword used in The Star Online. The other popular keywords found were palm oil, rice, fruits, livestock, vegetables, chicken, and crops. It is also noticed that different terms were used to describe similar concepts. The indentified keywords can be used to form an agro inventory. In future, this study will examine the other Malaysian Online newspaper such The New Strait Time and The Sun and compare the top agro food keyword among them.

Acknowledgment

The authors gratefully acknowledge the XXXX Research Grant Scheme for supporting this research project through grant XXX/1/2015/STG/UUM/02/49.

References

[1] A. Taylor, *The People's Platform: Taking Back Power and Culture in the Digital Age*: Random House Canada, 2014.

[2] T. Harcup and D. O'Neill, "What is news?," *Journalism Studies*, vol. 1, pp. 1-19, 2016.

[3] Weaver, David, R. Beam, P. V. Bonnie Brownlee, and G. C. Wilhoit, *The American Journalist in the 21st Century*. Mahwah, NJ:: Lawrence Erlbaum Associates, 2007.

[4] Shakeel PM. Neural Networks Based Prediction Of Wind Energy Using Pitch Angle Control. International Journal of Innovations in Scientific and Engineering Research (IJISER). 2014;1(1):33-7.

[5] P. Mohamed Shakeel; Tarek E. El. Tobely; Haytham Al-Feel; Gunasekaran Manogaran; S. Baskar., “Neural Network Based Brain Tumor Detection Using Wireless Infrared Imaging Sensor”, IEEE Access, 2019, Page(s): 1

[6] X. Tang, C. Yang, and J. Zhou, "Stock Price Forecasting by Combining News Mining and Time Series Analysis," in *International Joint Conferences onWeb Intelligence and Intelligent Agent Technologies (WI-IAT '09)*, Milan, Italy, 2009, pp. 1-5.

[7] S. S. Abdullah, M. S. Rahaman, and M. S. Rahman, ""Analysis of stock market using text mining and natural language processing","

presented at the International Conference on Informatics Electronics & Vision (ICIEV) 2013.

[8] Sridhar KP, Baskar S, Shakeel PM, Dhulipala VS., “Developing brain abnormality recognize system using multi-objective pattern producing neural network”, Journal of Ambient Intelligence and Humanized Computing, 2018:1-9. https://doi.org/10.1007/s12652-018-1058-y

[8] G. P. C. F. J. X. Y. W. Lam, "Stock prediction: Integrating text mining approach using real-time news," in *IEEE International Conference on Computational Intelligence for Financial Engineering*, Hong Kong, 2003, pp. 1-6.

[9] Selvakumar S, Inbarani H, Shakeel PM. A Hybrid Personalized Tag Recommendations for Social E-Learning System. International Journal of Control Theory and Applications. 2016;9(2):1187-99.

[10] Z. Ma, G. Pant, and O. R. L. Sheng, "Mining competitor relationships from online news: A network-based approach," *Electronic Commerce Research and Applications,* vol. 10, pp. 418-427, 2011/07/01/ 2011.

[11] P. Kroha, R. Baeza-Yates, and B. Krellner, "Text Mining of Business News for Forecasting," in *International Workshop onDatabase and Expert Systems Applications (DEXA '06)* Krakow, Poland, 2006, pp. 1-5.

[12] M. Tsagkias, W. Weerkamp, and M. d. Rijke, "Predicting the volume of comments on online news stories," presented at the Proceedings of the 18th ACM conference on Information and knowledge management, Hong Kong, China, 2009.

[13] D. Goodman, Ed., *Agro-Food Studies in the 'Age of Ecology': Nature, Corporeality, Bio-Politics* 1). USA: Blackwell Publishers, 1999, p.^pp. Pages.

[14] S. Rose, D. Engel, N. Cramer, and W. Cowley, *Automatic Keyword Extraction from Individual Documents*, 2010.

[15] eBizMBA. (2017, 9 October). *Top 15 Most Popular News Websites (July 2017)*. Available: http://www.ebizmba.com/articles/news-websites

[16] J. A. Iglesias, A. Tiemblo, A. Ledezma, and A. Sanchis, "Web news mining in an evolving framework," *Inf. Fusion*, vol. 28, pp. 90-98, 2016.

[17] Baskar, S., & Dhulipala, V. R., “M-CRAFT-Modified Multiplier Algorithm to Reduce Overhead in Fault Tolerance Algorithm in Wireless Sensor Networks”, Journal of Computational and Theoretical Nanoscience,2018, 15(4), 1395-1401.

[18] J. Yoon, "Detecting weak signals for long-term business opportunities using text mining of Web news," *Expert Systems with Applications*, vol. 39, pp. 12543-12550, 2012/11/15/ 2012.