Bulletin of Electrical Engineering and Informatics
Vol. 10, No. 1, February 2021, pp. 365~373
ISSN: 2302-9285, DOI: 10.11591/eei.v10i1.2504

Visualizing stemming techniques on online news articles text analytics

Nurul Atiqah Razmi, Muhammad Zharif Zamri, Sharifah Syafiera Syed Ghazalli, Noraini Seman
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Malaysia

Article Info

Article history:

Received Mar 31, 2020
Revised Jun 16, 2020
Accepted Jul 23, 2020

Keywords:

Lancaster stemmer
Porter stemmer
Stemming
Text analytics
Visualization

ABSTRACT

Stemming is the process to convert words into their root words by the stemming algorithm. It is one of the main processes in text analytics where the text data needs to go through stemming process before proceeding to further analysis. Text analytics is a very common practice nowadays that is practiced toanalyze contents of text data from various sources such as the mass media and media social. In this study, two different stemming techniques; Porter and Lancaster are evaluated. The differences in the outputs that are resulted from the different stemming techniques are discussed based on the stemming error and the resulted visualization. The finding from this study shows that Porter stemming performs better than Lancaster stemming, by 43%, based on the stemming error produced. Visualization can still be accommodated by the stemmed text data but some understanding of the background on the text data is needed by the tool users to ensure that correct interpretation can be made on the visualization outputs.

Corresponding Author:

Noraini Seman,
Department of Computer Science,
Faculty of Computer & Mathematical Sciences,
Universiti Teknologi MARA,
Shah Alam, Selangor, Malaysia.
Email: aini@tmsk.uitm.edu.my

1. INTRODUCTION

Text analytics has gained a great deal of attention and development in the last decade due to the tremendous growth of text data from the electronic media. A study from IDC has stated that the volume of text data is expanding by 50 times within the period of 2010 to 2020, thanks to the changes and advancements in electronics technology [1]. Text analytics or text mining refers to the process of knowledge discovery from text (KDT) from sources such as structured source-RDBMS and unstructured sources-XML, JSON, word documents and images [2]. It is practiced for data from different areas such as social media, electronic news, websites, blogs, electronic health record, academics, etc [3]. In text analytics, the pre-processing of text data is a very critical process. In the pre-processing phase to stem text data, when heavy stemming algorithms are applied on text data and the algorithms aggressively direct the removal of affixes, this results in incorrect stem words. The aggressive algorithms reduce the under-stemming errors but increase the over stemming. Under stemming is the condition where words that refer to the same root are not reduced to the same stem word, while over-stemming is the condition where words are reduced to the same stem word even though they have different meanings. Due to these stemming errors, the end analytics on the stemmed text may not be valid [4, 5]. In this study, the visualization process is a means to evaluate the performance of text data pre-processing. In this study, the Voyant tool is utilized from its webpage [6]. This tool is a widely-used tool for text data visualization due to its user-friendly interface and the ease to

interpret its resulted outcome for public usage [7, 8]. This study utilizes text data on the contents of petrol prices and blockchain- two popular topics in the country. Text analytics has been applied on petrol prices field for petrol forecasting-whether prices are going to increase or decrease [9]. Meanwhile, blockchain topic is also emerging in the country as its application grows in various industries [10, 11].

Text analytics approach includes multidisciplinary approaches that cover the process of text data acquisition and retrieval, text pre-processing, the analysis of the text itself, the extraction of information, visualization process and machine learning approaches for identification and classification [12, 13]. In the pre-processing of text data, the complex structure of the text data is pre-processed in order to be used in further analysis [3]. The common tasks in text pre-processing are tokenization, filtering, lemmatization, and stemming of the words. In this study, two steps of pre-processing are performed, which are filtering and stemming. Filtering is the process where some words - usually the words with no content addition such as the prepositions and conjunctions are being removed from the text data. The removed words are known as 'stop words'. As for stemming, it is the process to transform words into their root words by applying the stemming algorithms [14]. The filtering of the text data is conducted manually by removing the stop words, followed by stemming of the text data manually and by Python NLTK tool [15]. Meanwhile, stemming can be defined as the removal of affixes (prefixes, suffixes) from a word to generate its root word for analytical purpose [16]. It is an operation that relates morphological variants in words by the removal process of affixes; either the prefixes or suffixes, and producing a root form of word called a stem that most of the time, similarly approximates the root morphemes [17, 18]. Its main objective is to lessen the distinctive syntactic structures/word types of a word. There are two famous techniques in stemming, which are Porter and Lancaster. Porter stemming is the most generally utilized stemmer while Lancaster stemming is known as one of the most aggressive stemming algorithms.The stemming algorithms can be classified into three main categories, which are affix removal method, statistical method and mixed methods [19, 20].

The objective of this study is to compare the stemming error as resulted from Porter Stemmer vs Lancaster Stemmer, and thus to identify which one of them has better stemming performance. Next, the resulted stemmed text data are visualized, and the resulted visualizations are compared to identify any significant difference between the visualization outputs of manually stemmed text data, Porter stemmed, and Lancaster stemmed.

- Porter stemmer

This most popular algorithm for stemming of text data in the English Language has been proposed in 1980 [21]. The initial algorithm is based on the idea that the suffixes in English Language are constructed from smaller and simpler suffixes. At present, this algorithm consists of about 60 rules. When a rule is applied and accepted, the suffix is removed appropriately, and the next rule is then applied. Porter's general rule for removing a suffix is as follows [22]:

<condition><suffix>→<new suffix>

This means that if a word ends with the suffix and the stem before "suffix" satisfies the given condition, "new suffix" will replace "suffix". For example, a rule (m>0) EED→ EE means if the word has at least one vowel, consonant and ending EED, replace the ending to EE. Therefore, "agreed" turn into "agree". Meanwhile, for "feed", the word is remained since asides from the 'eed', it only contains one consonant and no other vowel [18, 21, 23]. Porter algorithm provides a simple approach for conflation that works well in practice and is applicable to a range of languages. It generally produces less error rate than other stemmers [21, 24, 25].

- Lancaster stemmer

The Lancaster Stemmer which is also known as Paice and Husk Stemmer is a conflation based iterative algorithm, developed by Chris D. Paice in 1990 [23, 26]. This algorithm consists of a table containing 120 rules provided in a rule file. Each rule specifies either removal or replacement of an ending. The stemming rules are terminated if they are not applicable for a word, or if the word starts with a vowel and there are only two letters left, or if a word starts with a consonant and there are only three characters left. If not, the stemming rules are applied and the process repeats. For example, the word 'agreed' that starts with a vowel is reduced correctly to 'agree'. However, this algorithm has the tendency to produce excessive stemming. The word 'agree' is under-stemmed into 'agr' while 'are' is under-stemmed into 'ar'. It is a very heavy algorithm, but at the same time it is easy to be implemented where the iteration of every rule takes care of removal and replacement of characters in a document as per rules applied.

The objective of this study is to compare the stemming error as resulted from Porter Stemmer vs Lancaster Stemmer, and thus to identify which one of them has better stemming performance. Next, the resulted stemmed text data are visualized, and the resulted visualizations are compared to identify any significant difference between the visualization outputs of manually stemmed text data, Porter stemmed and Lancaster stemmed.

2. RESEARCH METHOD

The methodology of this study is as shown in Figure 1. There are four designated phases: (i) data collection, (ii) data extraction, (iii) data processing, and (iv) visualization.

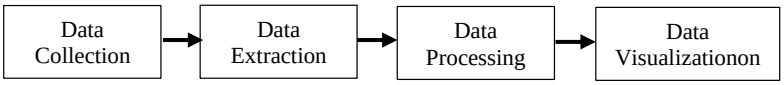


Figure 1. The general flow of text visualization

This study applies a sample of 10 text documents that are selected randomly from online newspaper articles. The topics covered in the articles are related with petrol prices and blockchain technology in Malaysia. Five articles are related to petrol prices issues and five articles are related to blockchain issues. Only five sentences are extracted from each of the article to limit the size of this study.In pre-processing, the stop words in the documents are filtered out. Stop words are the list of words that do not give contextual meaning and added information to the whole document, such as the words 'a', 'the', 'and', 'while', 'is' and etc. These words are irrelevant to the context, but most of the times appear frequently in documents that may affect the representation of the text data by visualization. Next, manual stemming is conducted, followed by stemming by Porter method and Lancaster method which Python NLTK [15]. The 2D visualization of the stemmed text data is then conducted. The performance of the stemming techniques are evaluated based on the stemming errors and the visualization outputs.

3. RESULTS AND ANALYSIS

3.1. Stemming by Porter method vs Lancaster method

All documents are uploaded into Python NLTK as separate documents. Excerpts of the resulted stemmed articles are shown in the following Figure 2 (a) and 2 (b). From Figure 2 (a) and 2 (b), significant differences can be seen from the result. Porter stems words into longer roots, while Lancaster stems words into much shorter root words. For example, the original words from the article - 'saving' is reduced into the correct root word 'save' by Porter but is reduced into 'sav' by Lancaster. This is an under-stemming error where the word is not reduced to the correct word. Other examples are 'similar' reduced to 'simil', 'Africa' reduced to 'Afric', and 'China' reduced to 'chin' by Lancaster. However, errors are also observed for Porter stemming. The summary of stemming errors for both stemmers on the 10 articles, as compared to our manual stemming as the baseline is presented in the following Table 1.

(a) (b)

Figure 2. (a) Example of stemmed text result of Porter Stemmer of Lancaster Stemmer, (b) Example of stemmed text result

For all the 10 articles tested, Lancaster stemming results in higher stemming error compared to Porter stemming. This is resulted from the aggressiveness of Lancaster that under-stems words, resulting in

the words being reduced to incorrect root words. The average stemming error for Porter is 21.89%, while Lancaster is having almost 10% higher error than Porter at 31.23%. As a comparison of the percentage of average stemming error, from the result, Porter stemming performs at 43% better than Lancaster stemming in term of less stemming error. The resulted stemmed texts from both stemmers are then visualized to compare on the visualization result from different stemming outputs.

Table 1. Comparison of stemming errors between Porter and Lancaster stemming

Article No.	Stemming Error	
	Porter	Lancaster
Article 1	22.33%	29.13%
Article 2	16.95%	31.36%
Article 3	25.00%	31.03%
Article 4	24.66%	35.62%
Article 5	27.97%	32.20%
Article 6	11.35%	18.44%
Article 7	22.41%	32.76%
Article 8	24.31%	34.72%
Article 9	21.77%	32.65%
Article 10	22.14%	34.35%
Average	21.89%	31.23%

3.2. Visualization outputs

The stemmer tool summarizes the number of unique words in the articles. For manual stemmed articles, the text contains 464 unique words. For Porter stemmed articles, there are 493 unique words while for Lancaster stemmed articles, there are 499 unique words. The manual stemmed articles contain much lesser number of unique words than the later. This is caused by the stemming variation from Porter and Lancaster stemming. For example, the word 'countries' is manually stemmed as 'country', but by Porter stemming, it is stemmed as 'countri'. This results in the Porter stemmed articles to obtain two nouns for 'country' which are 'country' and 'countri'. For Lancaster stemming, as example, the word 'africa' is stemmed as 'afric' while the 'african' is stemmed as 'afr', resulting in two nouns for 'africa' which are 'afric' and 'afr' whilst both words are manually stemmed into 'africa'. In short, Porter and Lancaster stemming increase the unique words in text data as compared to manual stemming.We refer to the statistical summary on the frequency and the relative frequency values of each word. Whilst frequency is the count of appearances of every word, the relative frequency is the fraction of times of occurrences of a word [27]. In this study, it indicates the count of appearances of words relative to the number of words in a segment of a corpus. The segments are the fraction of sentences separation, either by punctuation marks or by separators. In the case of this study, the visualization tool segments the corpus by the serial number of the uploaded articles which is from 1 to 10.

$$Relative\ frequency = \frac{\text{Frequency of words in a segment}}{\text{Total number of words in a segment}} \quad (1)$$

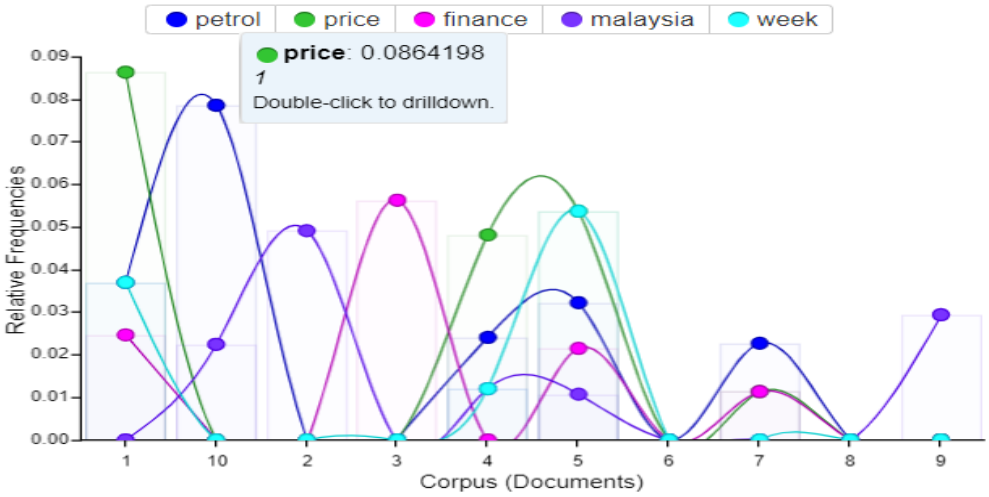
The statistical summary of frequency and relative frequency of the top 10 words in the articles are shown in Table 2.

Table 2. The relative frequency of words from different stemming techniques

Terms	Manual		Porter		Lancaster	
	freq	Relative frequency	freq	Relative frequency	freq	Relative frequency
price/prices/pric	17	0.086420*	13	0.061728	17	0.061728
baobab	6	0.084507*	6	0.084507*	13	0.070423*
petrol	17	0.078652*	17	0.078652*	10	0.078652*
blockchain	8	0.065574	8	0.065574*	8	0.065574*
africa	4	0.056338	4	0.056338	-	-
finance/financi/fin	9	0.056338	5	0.056338	6	0.056338
Hellogold	4	0.056338	4	0.056338	4	0.056338
university/univers	5	0.054348	5	0.053763	5	0.053763
price/prices/pric**	-	0.053763	4	0.049383	4	0.049383
week/weekli	9	0.053763		-	-	-
increase	6	-	6	0.049383	5	0.049383
country	6	-		-	4	0.049180

Note: * Top three rankings based on the relative frequency.** Appearance of word in second segment

The result shows that by manual stemming, the word 'price' acquires the highest relative frequency which is 0.086420, followed by 'baobab' (0.084507) and 'petrol' (0.078652). However, by Porter and Lancaster stemming, the top three words with the highest relative frequency do not include the word 'price'. The top three words from Porter stemming are 'baobab' (0.084507), 'petrol' (0.078652), and blockchain (0.065574), while the top three words from Lancaster stemming are 'petrol' (0.078652), 'baobab' (0.070423) and 'blockchain' (0.065574). This shows some inconsistency between the relative frequency representation for manually stemmed articles over Porter and Lancaster stemmed articles since the word with the highest relative frequency from manual stemming is ranked forth for both Porter and Lancaster stemming. The resulted relative frequency is visualized in term of the document trend at every segment of the articles as presented in Figure 3 (a) while the frequency value is visualized in the word cloud as in Figure 3 (b).



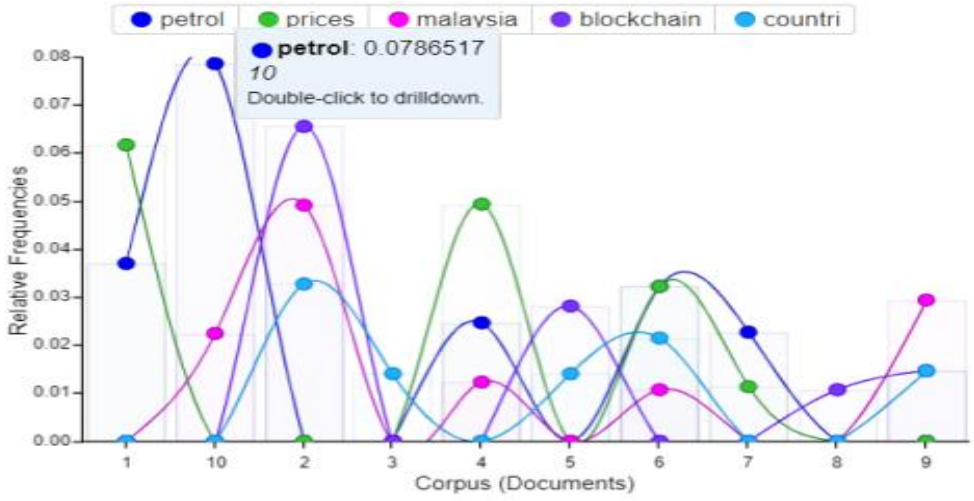
(a)

(b)

Figure 3. (a) Visualization in term of the document trend of manually stemmed articles, (b) Visualization of corpus word cloud of manually stemmed articles

The visualizations in Figure 3 (a) and 3(b) shows that from the manually stemmed article, 'price' occupies the highest relative frequency at the first segment of the text data and followed by petrol in the tenth segment. The words with the highest relative frequency over the 10 articles are 'petrol', followed by 'price', 'finance', 'malaysia' and 'week'. This indicate that the articles are mainly explaining on the petrol prices in Malaysia that are set by the Government of Malaysia on the weekly basis, and how does the petrol price

setting is influenced by the financial condition in Malaysia. Two words appear as the key words in the word cloud – 'price' and 'petrol'. The result from manual stemmed articles are compared with the visualizations of Porter stemmed articles as in Figure 4 (a) and 4 (b).



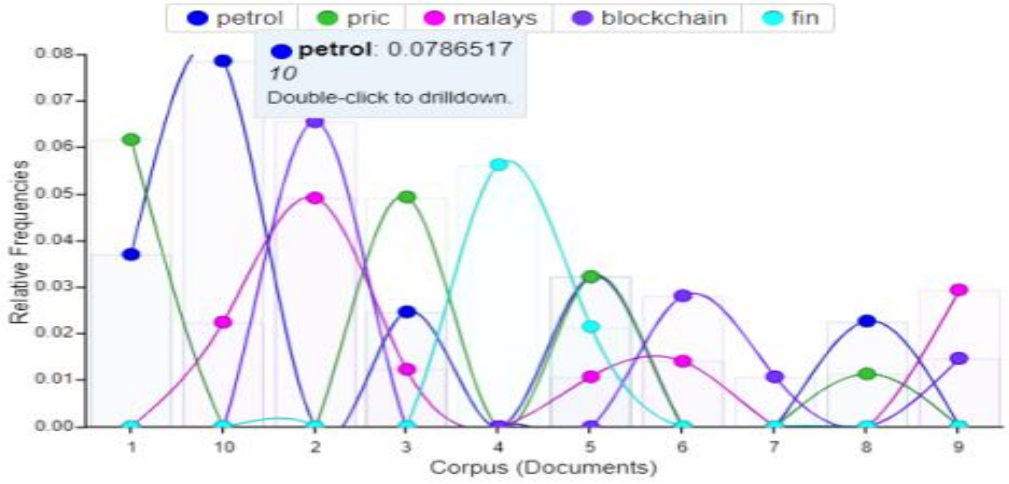
(a)

(b)

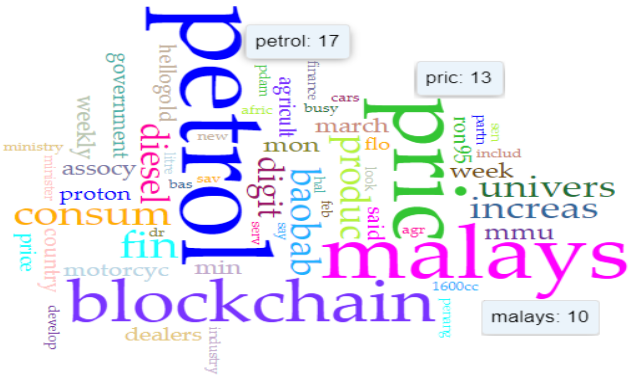
Figure 4. (a) Visualization of document trend using Porter stemmed articles, (b) Word cloud visualization using Porter stemmed articles

For Porter stemmed articles, the visualizations in Figure 4 show words with the highest relative frequency are 'petrol' in the tenth segment, followed by 'prices', 'malaysia', 'blockchain' and 'countri'. Same with the manually stemmed articles, the top two words with the highest frequency are 'petrol' and 'prices' that signify the topic of the articles on petrol prices issue. It is followed by 'malaysia', that signifies that the article is discussing on petrol prices issues in Malaysia. The next highest frequency words are 'blockchain' and 'countri', that indicates blockchain technology as having influence on petrol prices in Malaysia and for the word 'countri', the implication of this word is not clear whether it is explaining on the Malaysia country only or it regards other countries as well since the word comes in singular noun. From the word cloud, 'petrol' word retains its original frequency at 17 counts, but the size of 'prices' has decreased to 13 counts due to occurrences of under-stemming that separate the root word 'price' from word 'prices'. Since the size of 'prices' reduces, 'Malaysia' word appears to be more significant.

In Figure 5 (a) and 5 (b) is similar with Porter stemmed articles, the highest relative frequency word is 'petrol' at the tenth segment, followed by 'pric', 'malays', 'blockchain', and 'fin' (financial). In the word cloud, the significant size of 'petrol', 'pric', 'malays', and 'blockchain' can be seen. Similar with the result from Porter stemmed articles, the visualization may signify that the articles are discussing on petrol prices issues in Malaysia which influences or being influenced by the blockchain technology.



(a)



(b)

Figure 5. (a) Visualization of document trend using Lancaster stemmed articles, (b) Word cloud visualization using Lancaster stemmed articles

From the study as outlined in this paper, it has been found that Porter stemming performs better than Lancaster stemming, based on the stemming error produced by 43% better. The visualization on the three type of stemming techniques has resulted in the following words being visualized with the highest relative frequency and are the key words of the articles. The following words are in their order of relative frequency from the highest to the fifth highest relative frequency.

Manual : 'petrol', 'price', 'finance', 'malaysia' 'week'
Porter : 'petrol', 'prices', 'malaysia', 'blockchain' 'countri'
Lancaster : 'petrol', 'pric', 'malays', 'blockchain', 'fin'

From the manually stemmed text data, the main topic of the text data is on the petrol price setting in Malaysia that is affected by the financial situation of the country and is being reviewed and set by weekly basis. As observed here, the term 'blockchain' is visualized as less significant. This is due to the lack of counts of the term 'blockchain' in its related articles – it is being mentioned once to two times per article. On the other hand, by Porter and Lancaster stemming, the word 'blockchain' is also visualized as one of the most frequent word. The reason is due to by these two stemmers, the word 'finance' and 'week' are being under-stemmed. The original word 'financial' is reduced to 'financi' by Porter while reduced to 'financ' by Lancaster. The original word 'weekly' is reduced to 'weekli' by both stemmers that results in this 'weekli' word failed to be combined with the root word 'week' in the frequency counting. This has led to the missed-out information that the articles are stressing on the weekly set-up of the petrol prices. Another interesting finding is that the words 'Malaysia' and 'Malaysian' have been reduced to 'Malays', that indicate Lancaster stemming is not able to differentiate between 'Malaysia' – a country, and 'Malay' – a race. It is critical in term of stemming articles related to the country of Malaysia, or stemming articles related to the races in Malaysia, whereby Malaysia is a multi-racial country and the resulted stemming by Lancaster can result in

misguided visualization outputs on these matters. The word ‘financial’ has also been reduced to only ‘fin’, that anticipates the users to guess out its root word as they can be, for example, ‘fine’, ‘finish’, ‘final’, and etc. As for the relative frequency result as shown in Table 2, however, both Porter and Lancaster stemmed articles have not obtained ‘price’ as having the highest relative frequency. This is due to the root word ‘price’ have been stemmed to incorrect stems such as ‘prices’ and ‘pric’, that makes the statistical information incorrect.

4. CONCLUSION

As a conclusion from this study, in term of the stemming performance from the result of stemming errors, Porter stemming performs at 43% better than Lancaster stemming, but the current stemming errors are still high which are more than 20% of the words in the articles. As for the visualization process, visualization can still be accommodated by the stemmed text data by the two stemmers, since some of the visualization is correctly visualized as compared to the output from the manually stemmed text data. However, some understanding on the background of the articles is needed for by tool users to ensure that correct interpretation can be made on the visualization outputs. The study can be further continued to compare between Porter stemming and Lancaster stemming for Malay or English Languages, related to the issues in the country of Malaysia (as Lancaster stemming fails to stem ‘Malaysia’ correctly) and in term of the issues on multi races in Malaysia (as Lancaster stemming stems the word ‘Malaysia’ into ‘Malays’) to see the impact of the stemming error that may result in significant differences in the visualization output. Other than that, there are still potentials for the algorithms to be improved for English Language to increase the stemming accuracy and thus, the visualization output accuracy.

ACKNOWLEDGEMENTS

The authors would like to thank Universiti Teknologi MARA (UiTM) for the research funding and support via grant number 600-IRMI/PERDANA 5/3 BESTARI (107/2018).

REFERENCES

[1] J. Gantz and D. Reinsel, “THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East,” pp. 1-16, December 2012.
[2] R. Feldman and I. Dagan, “Knowledge Discovery in Textual Databases (KDT),” *First Int. Conf. Knowl. Discov. Data Min.*, vol. 95, pp. 112–117, 1995.
[3] Z. Zainol, M. T. H. Jaymes, and P. N. E. Nohuddin, “VisualUrText: A Text Analytics Tool for Unstructured Textual Data,” *Journal of Physics: Conference Series*, vol. 1018, no. 1, 2018.
[4] A. Joshi, N. Thomas, and M. Dabhade, “Modified Porter Stemming Algorithm,” *International Journal of Computer Science and Information Technologies*, vol. 7, no. 1, pp. 266–269, 2016.
[5] C. D. Paice, “An Evaluation Method for Stemming Algorithms,” in *SIGIR ’94*, London: Springer London, pp. 42–50, 1994.
[6] “Voyant Tools.” [Online]. Available at: https://voyant-tools.org/. [Accessed: 11-May-2019].
[7] M. Burghardt, J. Pörsch, B. Tirlea, and C. Wolff, “WebNLP: An Integrated Web-Interface for Python NLTK and Voyant,” *Proceeding 12th KONVENS*, pp. 1–5, September 2015.
[8] M. E. Welsh, “Review of Voyant Tools,” *Collab. Librariansh.*, vol. 6, no. 2, pp. 96–97, 2014.
[9] X. Li, W. Shang, and S. Wang, “Text-based crude oil price forecasting: A deep learning approach,” *International Journal of Forecasting*, 2018.
[10] L. S. Sankar, M. Sindhu and M. Sethumadhavan, "Survey of consensus protocols on blockchain applications," *2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, pp. 1-5, 2017.
[11] M. N. Saadat, S. Abdul Halim, H. Osman, R. Mohammad Nassr, and M. F. Zuhairi, “Blockchain based crowdfunding systems,” *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 15, no. 1, p. 409-413, July 2019.
[12] S. Dang and P. H. Ahmad, “Text Mining : Techniques and its Application,” *IJETI International Journal of Engineering & Technology Innovations*, vol. 1, no. 4, pp. 22–25, 2014.
[13] S. Sinclair and G. Rockwell, “Text Analysis and Visualization,” *A New Companion to digital humanities*, pp. 274–290, 2015.
[14] M. Allahyari *et al.*, “A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques,” 2017.
[15] “Python NLTK Stemming and Lemmatization Demo.” [Online]. Available: http://text-processing.com/demo/stem/. [Accessed: 11-May-2019].
[16] M. Sankupellay and S. Valliappan, “Malay-Language Stemmer,” *Sunway Academic Journal*, vol. 3, pp. 147–153, 2006.
[17] W. B. Frakes, V. Tech, and C. J. Fox, “Strength and Similarity of Affix Removal Stemming Algorithms,” *ACM SIGIR Forum*, New York, NY, USA, vol. 37, no. 1, pp. 26-30, 2003.

[18] D. A. Hull and G. Grefenstette, "A Detailed Analysis of English Stemming Algorithms," *Rank Xerox Research Centre*, pp. 1-16, 1996.
[19] M. Anjali and G. Jivani, "A Comparative Study of Stemming Algorithms," *Int. J. Comput. Technol. Appl.*, vol. 2, no. 6, pp. 1930–1938, 2011.
[20] M. Hadni, A. Lachkar and S. A. Ouatik, "A new and efficient stemming technique for Arabic Text Categorization," *2012 International Conference on Multimedia Computing and Systems*, Tangier, pp. 791-796, 2012.
[21] P. Willett, "The Porter stemming algorithm: then and now," *Program: electronic library and information systems*, vol. 40, no. 3, pp. 219-223, 2006.
[22] M. F. Porter, "Celebrating 40 Years Of ICT In Libraries, Museums And Archives An Algorithm For Suffix Stripping," *Program: electronic library and information systems*, vol. 14, no. 3, pp. 211–218, 2006.
[23] S. R. Sirsat, S. R. Sirsat, D. V. Chavan, D. Hemant, S. Mahalle, and P. N. Mahavidyalaya, "Strength and Accuracy Analysis of Affix Removal Stemming Algorithms," *International Journal of Computer Science and Information Technologies*, vol. 4, no. 2, pp. 265–269, 2013.
[24] M. Lennon, D. S. Peirce, B. D. Tarry, and P. Willett, "An evaluation of some conflation algorithms for information retrieval," *Journal of information Science*, vol. 3, no. 4, pp. 177–183, 1981.
[25] W. Ben and A. Karaa, "A New Stemmer To Improve Information Retrieval," *International Journal of Network Security & Its Applications*, vol. 5, no. 4, pp. 143–154, 2013.
[26] J. Köhler, S. Philippi, M. Specht, and A. Rüegg, "Ontology based text indexing and querying for the semantic web," *Knowledge-Based Systems*, vol. 19, no. 8, pp. 744–754, Dec. 2006.
[27] S. Dean, B. Illowsky, and D. Ph, "Sampling and Data : Frequency, Relative Frequency, and Cumulative Frequency Frequency Table of Student Work Hours w/Relative Frequency," pp. 1–6, 2010.