ABBREVIATION DICTIONARY FOR TWITTER HATE SPEECH

Zainab Mansur[1], Nazlia Omar[2] and Sabrina Tiun[3]

[1]Department of Computer Science, Faculty of Sciences, Omar Al-Mukhtar University, Al Bayda, Libya

[2]Center for AI Technology, Universiti Kebangsaan Malaysia, Bangi Malaysia
[3]Center for AI Technology, Universiti Kebangsaan Malaysia, Bangi Malaysia

ABSTRACT

Informal methods of communication, like tweets, rely heavily on initialization abbreviations to reduce message size and time, making them difficult to mine and normalize using existing methods. Therefore, this present study compiled a lexicon repository to normalize the initialism abbreviations used in tweets in the English language. Several components were taken into consideration while compiling the repository. This included the Tweepy Python library, keyword list, small developed rules, and online dictionaries. A lexicon repository of 300 abbreviations and their complete forms was compiled. This will be used in an ongoing study to normalize Twitter hate speech data and to detect it.

KEYWORDS

Normalization, Hate Speech, Twitter, Abbreviation, Dictionary

1. INTRODUCTION

Social media users use informal language styles to communicate, making it difficult for natural language processing (NLP) tools to process the data [1]. Furthermore, the abundant use of abbreviations in informal writing poses a challenge in the text normalization process [2]. It is common for social media users to abbreviate words, especially on Twitter, in lieu of full names and words ([3]); [4]). Abbreviations are shortened versions of phrases or words, while an acronym is a phrase or word that was shortened by combining its parts [5]. [6] used abbreviations to normalize acronyms and abbreviations. Text mining algorithms produce more vector dimensions when using Tweets that contain many slang or abbreviated words [7]. Subsequently, extending all abbreviations and subclasses of these terms to their complete versions is necessary [5] [8]. Text normalization has been used to enhance multiple social media NLP tasks [9]. However, the currently available normalization methods fail to resolve and correct the informal abbreviation for tweets ([10]). An example of these abbreviations would be "RN" for “right now” and "OMG" for “oh, my God”. Every language contains its own set of distinct abbreviations. As such, social media platforms require dictionaries specific to each language [4]. This present study constructed a lexicon repository to normalize initialization abbreviations on Twitter. The compiled dictionary will be used to normalize Twitter hate speech detection data in an ongoing study. The rest of this study is organized as follows: Section 2 addresses similar studies, Section 3 describes how to compile a dictionary, and Section 4 gives the findings. A summary of this study and suggestions for future research are presented in Section 5.

David C. Wyld et al. (Eds): DBML, CITE, IOTBC, EDUPAN, NLPAI -2023
pp.261-269, 2023. IJCI – 2023 DOI:10.5121/ijci.2023.120219

2. LITERATURE REVIEW

A rigorous normalization process is required to manage the wealth of social media data and text created online. Linguistic resources, such as abbreviation dictionaries, could facilitate the normalization process. This section discusses studies similar to the objectives of this present study, i.e., creating dictionary resources to normalize abbreviations in social media texts. [3] built a colloquial dictionary of the Malay language using a context-aware system and dealt with ambiguity using the dictionary technique, where a term may have many meanings. The entries in the context dictionary were based on how frequently the times before and after the mistaken word appeared. As the dictionary contained a limited number of entries and a simple schema, it was constructed using the Python dictionary module.

[11] used the string replace method to expand abbreviations by first converting an abbreviation to lowercase and then searching it on four online dictionaries. However, there needs to be more normalization dictionaries that are (1) able to provide an adequate vocabulary that is linked to a webpage, (2) include internet slang and abbreviations, and (3) are frequently updated. [4] looked for free webpages on the Internet to extract sets of abbreviations. The meaning of each abbreviation was then manually checked on the available online dictionaries. According to the study, online English texts do not include social media terms. As such, slang term lexicons should be updated by manually compiling entries for each language.

[4] used a list of Spanish slang words carefully selected by [12] from 21,000 tweets using emotional hashtags. The Tweepy Python library was then used to develop a list of 200 slang words and their meanings. The words and their context were then manually revised.

Meanwhile, [13] created a context dictionary of Indonesian abbreviations and their multiple context-based meanings. Only words that appeared in complaint sentences were compiled. However, more research is needed to enhance the capacity of the algorithm to analyze English words and determine whether settings other than complaint data may be utilized. [14] developed a method based on the dictionary-based search and longest common subsequences (LCS) for Indonesian abbreviations and acronyms. However, the data collection process and the costs of maintaining the dictionary-based algorithm was high. Furthermore, the LCS performed poorly as many words have similar subsequent words. [7] developed a dataset via crowdsourcing to use the preceding or subsequent word as the keyword to normalize abbreviated Indonesian words. However, specific keywords are needed for reliable meaning identification.

Multiple studies have developed abbreviation dictionaries for social media text in different languages and for various tasks. Nevertheless, there need to be more dictionaries resources to normalize abbreviations for hate speech detection in Twitter texts. Hate tweets, there are no distinguishable characteristics between them and non-hate tweets[15]. Accordingly, It is possible that some abbreviations are only used in hate speech and would not be included in a list of non-hate speech abbreviations. Therefore, a separate list of hate speech abbreviations might be required to detect hate speech accurately. As such, this present study constructed a dictionary to help normalize the abbreviations in Twitter hate speech datasets.

3. METHODOLOGY

Multiple components such as the Tweepy Python library, a list of keywords, a set of rules, and online dictionaries were taken into consideration to compile such a resource. These components are discussed in greater detail in the following subsection. Figure 1 depicts the methodology that was used to compile the abbreviation dictionary.

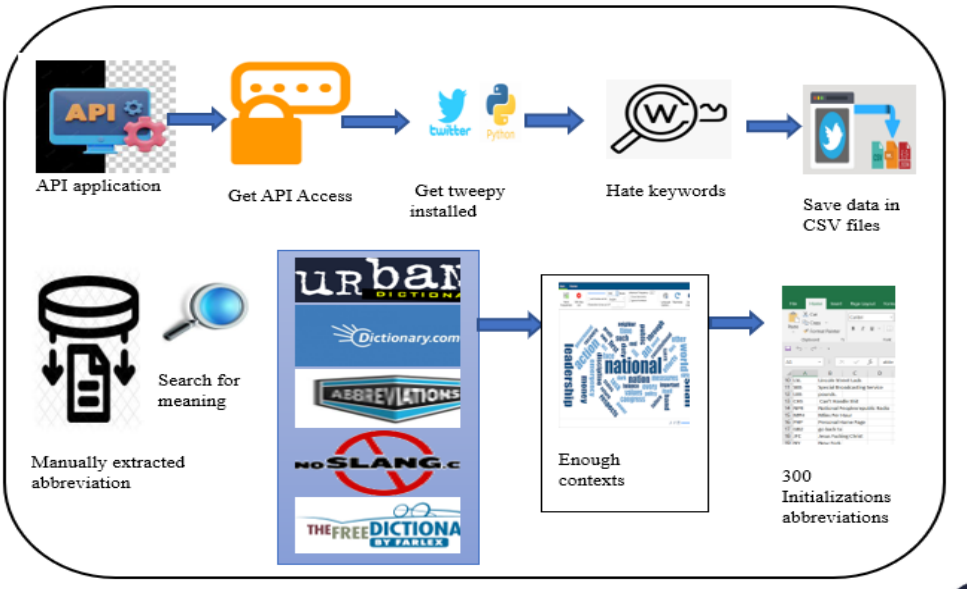

Figure 1. The method used to compile the abbreviation dictionary

A Twitter registration form was completed to create a new profile and gain access to the consumer key and secret (Figure 2). According to Twitter, each request must be authenticated through Open Authorisation (OAuth) (Figure 3). Most tweets were retrieved by using an open Twitter search API.

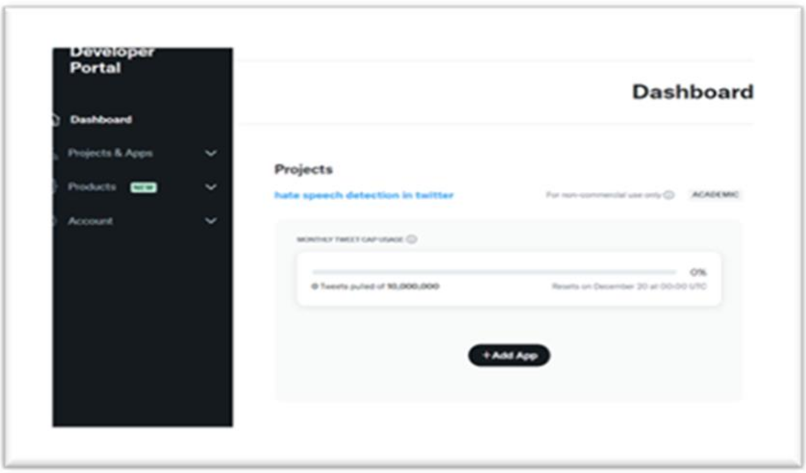

Figure 2. The Twitter API form in the developer account

Figure 3. The API access for authentication

The Tweepy Python library (http://www.tweepy.org/) was used to collect the tweets, excluding tweets written in languages other than English [16]. Abbreviations that appeared in hateful contexts were compiled using the exact keywords that [17] used. [17] provide 16000 annotated tweets that contain hate speech. Keywords similar to those used by [17] were used to collect data from Twitter. As seen in Table 1, as hate speech language on Twitter was the priority. However, a word may be hated for one but good for another. We collected tweets using the same hashtag of the benchmark dataset, which ensures the contextual text had to be the same for both offensive and inoffensive tweets. However, the annotation agreement of this dataset was 84%. Approximately 24506 tweets were downloaded using the hashtags associated with those terms. Figure 4 provides a sample of the tweets extracted from Twitter.

Table 1. A list of hashtags used in the [17] dataset

Hashtag terms
“MKR”, “asian drive”, “femi-nazi”, “immigrant”, “nigger”, “sjw”, “WomenAgainstFem- inism”, “blameonenotall”, “islam terrorism”, “notallmen”, “victimcard”, “victim card”, “arab terror”, “gamergate”, “jsil”, “racecard”, “race card”


```
I agree \n\nThis Tweet is everything https://t.co/xghRlBXcMZ', 'RT
@Pule_EM: The EFF needs to articulate its open border stance clearly. To a
South African that stays in a shack watching an undocumented…', '@NFLFilms
@demario__davis @Saints I hope you're real proud of yourself helping to bail
out a illegal immigrant who… https://t.co/oiKdSE9KtG', 'RT @ChuckCallesto:
BREAKING REPORT: Ted Cruz Proposes 'Win-Win' Way to Give Every Illegal
Immigrant $450,000, suggests giving them HUNTER B…', 'RT @PrentonTommy: I
wonder why the @ukhomeoffice is failing to stop the illegal immigrant
invasion https://t.co/h1QL0qmypI', 'RT @tbeaux29: @MVZexplorer @codeofvets
@iamajax @Cali_CC_2020 @dontcavitate @Chopstress @flowerPotzzz @ThomasP3120
@JRAYBRO1967 @b_gawk @33…', 'RT @SongsPhilosophy: On this day in
1970\n\nLed Zeppelin release "Immigrant Song" in the US\n\nLed
Zeppelin\nImmigrant Song \nLive 1972\n\n#Led…', 'RT @ChuckCallesto:
BREAKING REPORT: Ted Cruz Proposes 'Win-Win' Way to Give Every Illegal
Immigrant $450,000, suggests giving them HUNTER B…', '@737_Legacy
@resistwithlove1 Well being as how I called you uneducated, with no mention
of race based solely on you… https://t.co/7Eos41b25N', 'RT @Bethvanduyne:
HORRIBLE and of course this is happening. Illegal immigrants are lying to the
Biden Administration about their age, being…', '"Yery Noel Medina Ulloa was
busted Oct. 7 in Jacksonville when he was found covered in blood after
allegedly killin… https://t.co/DzQkfcl6e5', 'RT @ChuckCallesto: BREAKING
REPORT: Ted Cruz Proposes 'Win-Win' Way to Give Every Illegal Immigrant
$450,000, suggests giving them HUNTER B…', '@washingtonpost Biden
administration let a 23 yr old illegal immigrant from Honduras into the U.S.,
who used a fict… https://t.co/DgGaBMPe0Z', 'RT @WinWithWinsome: In what other
country could an immigrant woman rise from arriving with $1.75 to her
family's name, to serving as Lt. Go…', 'THOR: RAGNAROK (2017) Movie Clip -
God of Thunder | Marvel Studios HD https://t.co/gTS0Wnn6me via @YouTube…
https://t.co/GUbW4ZRoFP', 'RT @ChuckCallesto: BREAKING REPORT: Ted Cruz
Proposes 'Win-Win' Way to Give Every Illegal Immigrant $450,000, suggests
giving them HUNTER B…', 'RT @ChuckCallesto: BREAKING REPORT: Ted Cruz
Proposes 'Win-Win' Way to Give Every Illegal Immigrant $450,000, suggests
giving them HUNTER B…', '@MrsBuddystewar1 Being an immigrant with English as
```

Figure 4. A sample of the downloaded tweets

The extracted tweets were saved in the CSV file format. The abbreviations were then manually selected at random and checked across the online dictionaries listed in Table 2.

Table 2. Internet dictionaries that used to confirm the interpretation of the abbreviations

Dictionary Name	URL
Urban Dictionary	https://www.urbandictionary.com
DICTIONARY.COM	https://www.dictionary.com
ABBREVIATIONS	https://www.abbreviations.com
NOSLANG.COM	https://www.noslang.com
TheFreeDictionary	https://www.thefreedictionary.com/

These dictionaries contain many complete versions of one abbreviation. Furthermore, the meaning of an abbreviation can varies depending on its context [13]. As a result, a set of guidelines was created to select the most acceptable finished version for a particular abbreviation, where lexical variants appear in settings identical to their legal forms. [18]. Multiple studies have developed rule-based methods for a wide range of languages and tasks that yield impressive results [20, 22, 23]. As data analysis indicates that correlations can be detected based on context [24], the data was analyzed and used to develop the following rules:

a. If an abbreviation appears more than twice in a similar context, it is included.
   As an example, the following two tweets:” Bataao BC ab victim card khel raha hai when he is the one who called a fellow Muslim Islamist despite knowing the fâ€¦ https://t.co/pL52SOUEtQ, and “RT @tariq22qamar: Bataao BC ab victim card khel raha hai when he is the one who called a fellow Muslim Islamist despite knowing the fact thâ€¦”.
b. If an abbreviation lacks context, it is excluded.
   As an illustration, in the following two tweets: ” @A_Feranmi @madebycharles @DavidIAdeleke Backend o, everything. All while trying to shade NB for not being â€œtrustwoâ€¦ https://t.co/25SMXEP7K5”, and “@EazyH4 OMGGGG YOU A KING FR FR ðŸ’¯ #respect ðŸ¥ºðŸ¥ºðŸ¥º #NotAllMen”
c. If an abbreviation appears in a different context, it is excluded.
   As an example, these two different tweets: “@CM6982 @RagemanFC @Spadez86 @Nerdrotics Since when tf was BP political or SJW or whatever the fuck?”, “@conservatyler But a lot of BP are astroturfed by elites who can shelter themselves from consequences.”

It is significant to note that since we are interested in the context in which the abbreviations appear, pre-processing task is optional here as it is done in related works.

4. Results and Discussion

The abbreviations were randomly chosen from the extracted Twitter data and saved in a CSV file. These words and their contexts were manually reviewed according to the developed rules. A list of 300 abbreviations and their complete forms were saved in a CSV file as the resulting dictionary abbreviations. Figure 5 provides an excerpt of the complete form of the abbreviations.

	A	B	C	D	E
1	Abbr	Full			
2	JFC	Jesus Fucking Christ			
3	WHF	What the fuck			
4	MKR	Mega Kack Reiz			
5	UFC	Ultimate Fucking Championships			
6	JFK	Just fucking kidding			
7	DPS	Damage per second			
8	BGS	bubble guts			
9	CFC	Canâ€™t fucking cap			
10	FML	Fuck My Life			
11	FMD	Fuck My Day			
12	SMH	shaking my head			
13	WTF	well that fuck			
14	MGM	male genital mutilation			
15	BLK	Black			
16	LMFAO	Laughing my fucking ass off			

Figure 5. An excerpt of the initialism abbreviated words

However, a few existing works as [4] and [10], have collected English abbreviations from several web sources, which are different from our data source. The results of this research were evaluated with those of other comparable studies that built an abbreviation dictionary with data from Twitter (Table 3). This present study was found to have extracted more tweet instances than extant studies (Figure 6).

Table 3. A comparison with similar studies using twitter data

Reference	Abbreviation Count	Task	Language
[3]	Small in general	Text normalization	Malay
[19]	378	Text normalization	Indonesian
[12]	200	Sentiment analysis	Spanish
Proposed abbreviation dictionary	300	Hate speech detection	English

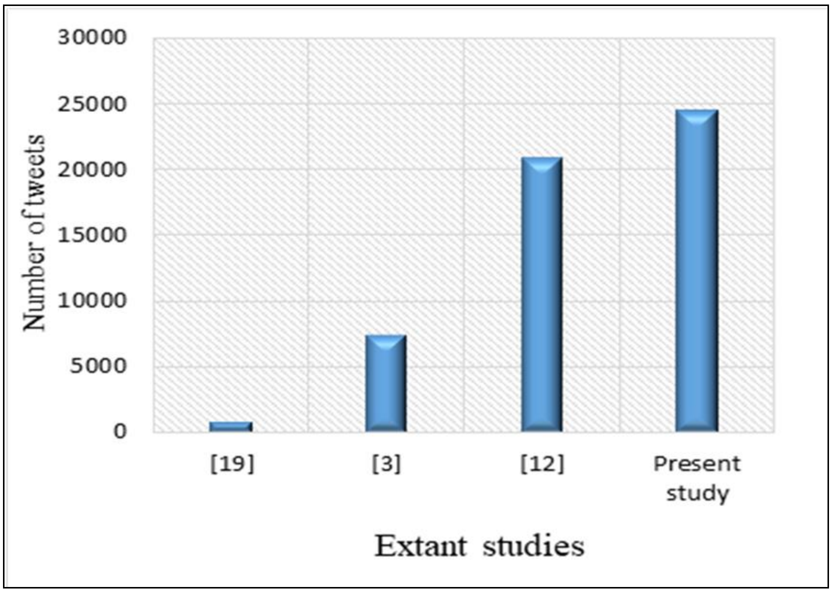

Figure 6. A comparison of tweets extracted from similar extant studies

However, the created abbreviation list will be examined on how well it can handle the initialized abbreviations when we develop the normalization model of hate speech text.On the other hand, a dictionary for a new abbreviation term should be created. As a result, regularly updating with a new abbreviation that cannot be found in our abbreviation list is necessary, which is available upon request.

5. Conclusion

This present study constructed a Twitter lexicon repository that normalizes initialism abbreviations. The dictionary was developed by using multiple strategies; such as collecting tweets from the Tweepy Python library, a hate speech keyword, a few developed rules, and online dictionary searches. The proposed method was then compared to similar methods that had been used to build abbreviation dictionaries for Twitter. The suggested method was distinct from those before in that it addressed the English language about hate speech. It also has more abbreviations than similar methods, except for one method that extracted 78 more words. The developed abbreviation dictionary will be used in a future study to normalize Twitter hate speech detection datasets.

Acknowledgements

This work was supported by Universiti Kebangsaan Malaysia and the Ministry of Higher Education (MOHE) under the grant FRGS/1/2020/ICT02/UKM/02/6 and TAP-K007009

References

[1] T. Baldwin and Y. Li, "An in-depth analysis of the effect of text normalization in social media," NAACL HLT 2015 - 2015 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. Proc. Conf., no. January 2015, pp. 420–429, 2015, doi: 10.3115/v1/n15-1045.
[2] D. L. Pennell and Y. Liu, "Normalization of informal text," Comput. Speech Lang., vol. 28, no. 1, pp. 256–277, 2014, doi: 10.1016/j.csl.2013.07.001.
[3] M. A. Saloot, N. Idris, and R. Mahmud, "An architecture for Malay Tweet normalization," Inf. Process. Manag., vol. 50, no. 5, pp. 621–633, 2014, doi: 10.1016/j.ipm.2014.04.009.

[4] H. Gómez-Adorno, I. Markov, G. Sidorov, J. P. Posadas-Durán, M. A. Sanchez-Perez, and L. Chanona-Hernandez, "Improving Feature Representation Based on a Neural Network for Author Profiling in Social Media Texts," Comput. Intell. Neurosci., vol. 2016, 2016, doi: 10.1155/2016/1638936.

[5] S. Noor, A. Noor, and S. Tiun, "Rule-based Text Normalization for Malay Social Media Texts," vol. 11, no. 10, 2020.

[6] Y. Wu, B. Tang, M. Jiang, S. Moon, and J. C. Denny, "Clinical Acronym / Abbreviation Normalization using a," 2013.

[7] D Sebastian, KA Nugraha, "Text Normalization for Indonesian Abbreviated Word Using Crowdsourcing Method," 2019 Int. Conf. Inf. Commun. Technol., pp. 529–532, 2019.

[8] A. M. Jaber, Word Sense Disambiguation for Clinical Abbreviations by Areej Mustafa Mahmoud Jaber, no. February. 2022.

[9] M. Arora and V. Kansal, "Character level embedding with deep convolutional neural network for text normalization of unstructured data for Twitter sentiment analysis," Soc. Netw. Anal. Min., vol. 9, no. 1, p. 0, 2019, doi: 10.1007/s13278-019-0557-y.

[10] J. Khan and S. Lee, "applied sciences Enhancement of Text Analysis Using Context-Aware Normalization of Social Media Informal Text," 2021.

[11] P. Sosamphan, V. Liesaputra, S. Yongchareon, and M. Mohaghegh, "Evaluation of statistical text normalisation techniques for twitter," IC3K 2016 - Proc. 8th Int. Jt. Conf. Knowl. Discov. Knowl. Eng. Knowl. Manag., vol. 1, no. January, pp. 413–418, 2016, doi: 10.5220/0006083004130418.

[12] G. Sidorov, M. I. Romero, I. Markov, R. Guzman-Cabrera, L. Chanona-Herńandez, and F. Vel asquez, "Detecci on autoḿatica de similitud entre programas del lenguaje de programaci on Karel basada en t ecnicas de procesamiento de lenguaje natural," Comput. y Sist., vol. 20, no. 2, pp. 279–288, 2016, doi: 10.13053/CyS-20-2-2369.

[13] N. Hanafiah, A. Kevin, C. Sutanto, Fiona, Y. Arifin, and J. Hartanto, "Text Normalization Algorithm on Twitter in Complaint Category," Procedia Comput. Sci., vol. 116, pp. 20–26, 2017, doi: 10.1016/j.procs.2017.10.004.

[14] D. Gunawan, Z. Saniyah, and A. Hizriadi, "ScienceDirect Normalization Normalization of of Abbreviation Abbreviation and and Acronym Acronym on on Microtext Microtext in in Bahasa Indonesia by Using and Longest Common Bahasa Indonesia by Using Dictionary-Based and Longest Common Subsequence Subseq," Procedia Comput. Sci., vol. 161, pp. 553–559, 2019, doi: 10.1016/j.procs.2019.11.155.

[15] Z. Zhang and L. Luo, "Hate speech detection: A solved problem? The challenging case of long tail on Twitter," Semant. Web, vol. 10, no. 5, pp. 925–945, 2018, doi: 10.3233/SW-180338.

[16] J. Roesslein, "Tweepy documentation!!!!!," vol. 3.6.0, 2016.

[17] Z. Waseem and D. Hovy, "Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter," pp. 88–93, 2016, doi: 10.18653/v1/n16-2013.

[18] B. Han, "Improving the Utility of Social Media with Natural Language Processing," no. February, 2014.

[19] N. Hanafiah, A Kevin, C Sutanto, Y Arifin, "ScienceDirect ScienceDirect Text Normalization Algorithm on Twitter in Complaint Category Text Normalization Algorithm on Twitter in Complaint Category," Procedia Comput. Sci., vol. 116, pp. 20–26, 2017, doi: 10.1016/j.procs.2017.10.004.

[20] U Nadia, N OMAR. "Malay Named Entity Recognition using Rule Based Approach". Asia-Pacific Journal of Information Technology and Multimedia. 2019;8(1):37-47.

[21] H Alshalabi, S Tiun, N Omar, E abdulwahab Anaam, Y Saif . BPR algorithm: New broken plural rules for an Arabic stemmer. Egyptian Informatics Journal. 2022, Vol. 23 No. 3, 363–371.

[22] NA Halid , N Omar, Malay Part Of Speech Tagging Using Ruled-Based Approach. Jurnal Teknologi Maklumat dan Multimedia Asia-Pasifik. 2017;6(2):91-107.

[23] S SAAD, K LATIFF U. Extraction of concept and concept relation for islamic term using syntactic pattern approach. Asia-Pacific Journal of Information Technology and Multimedia. 2018, Vol. 7 No. 2, 71 – 84.

AUTHORS

ZAINAB MANSUR received the B.Sc. de- gree (Hons.) in computer science from Omar Al-Mukhtar University, Libya, in 2003, and the M.S. degree in computer science from the Universiti Kebangsaan Malaysia (UKM), Malaysia, in 2011. She is currently a researcher at UKM. Her research interest falls under natural language processing, machine learning, text and Web mining, and sentiment analysis.

NAZLIA OMAR received the B.Sc. degree (Hons.) from UMIST, U.K, the M.Sc. degree from the University of Liverpool, U.K., and the Ph.D. degree from the University of Ulster, U.K. She is currently an Associate Professor at the Center for AI Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia. Her main research interest is in the area of natural language processing and computational linguistics.

SABRINA TIUN received the B.Sc. degree (Hons.) from Bradley University, USA, and the master's and Ph.D. degrees from Universiti Sains Malaysia. She is currently a Senior Lecturer with the Faculty of Information Science and Technology (FTSM), Universiti Kebangsaan Malaysia (UKM), Malaysia. Her research interests include natural language processing, computational linguistics, speech processing, and social science computing.