

Insights extraction on cross-cultural interaction through astronomy online labs using data analytics

M. Bakri¹, Norhaslinda Kamaruddin², M. Hamiz³, P. Marlia⁴, A. H. S. Nurhasmiza⁵, Z. Othman⁶, N. A. S. Nilam⁷

^{1,3}Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Malaysia

²Advanced Analytics and Engineering Centre, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Malaysia

^{4,5}Language Academy, Faculty of Social Science and Humanities, Universiti Teknologi Malaysia, Malaysia

⁶Department of Geoinformation, Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia, Malaysia

⁷Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Malaysia

Article Info

Article history:

Received Jan 29, 2019

Revised Apr 13, 2019

Accepted May 19, 2019

Keywords:

Data analytics

Cross-cultural interaction

Visualization

Astrology education

Dialogical inquiry

ABSTRACT

Dialogical inquiry on astrology offers the participants to gain not only the intellectual and technological knowledge on the subject matter but also the social benefit of the interaction. Different cultures and values may pose as a challenge for adaptation when the participants start to collaborate in order to complete the group work. Hence, multiple sessions of cross-cultural interaction through Astronomy Online Labs had been conducted to give the participants a standardized platform to discuss and communicate. However, it is imperative to observe the content and frequency of the interaction to ensure both parties (Local and Non-local) benefited from such interaction. The interaction had been recorded and analysed to give us some insights for the improvement of the future participants' engagement. The visualization techniques such as word cloud, word forest, timeline as well as Venn diagram approaches had been used and it is observed that the participants are actively communicating with the Non-local slightly dominating the session. It is hoped that the analysis tool can be embedded in the platform that it can provide dynamic analysis on the go while interaction happens so the moderator can steer the interaction to the intended topic.

Copyright © 2019 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Norhaslinda Kamaruddin,
Advanced Analytics and Engineering Centre,
Faculty of Computer and Mathematical Sciences,
Universiti Teknologi MARA,
40450 Shah Alam, Selangor, Malaysia.
Email: norhaslinda@tmsk.uitm.edu.my

1. INTRODUCTION

Collaborative learning is an educational approach to teaching and learning activities that involves group of students working together to solve a problem, complete a task or create a product. It involves two or more students to be grouped together and discuss any related task given by the instructor. The activities are varied, but what can be completed during the session are problem solving, study teams, debates or any discussion regarding the subject matter. Traditionally, students will be in the same place and learn through their peers face-to-face. However, with technology nowadays, students from different countries and cultures can collaborate together in a group with the help of Internet [1, 2]. The main purpose of collaborative learning is to enhance social skills of students since they need to interact with the members in the group which are coming from different background [3]. The interaction between students might help them to gain confidence as they get the support from the team members [4]. Vygotsky [5] believed that children can

improve with the right interaction with knowledgeable peers, He also suggested that the better way to examine intelligence is on the application of the knowledge whether the children can independently solve the problem or needing the adult's intervention [6]. The concept is known as the Zone of Proximal Development (ZPD).

One of the ways to promote collaborative learning is through dialogical inquiry. Dialogical inquiry allows expert-student interactions that involve a mix of authoritative talks and dialogic discourses with the former being more appropriate for experts to introduce new ideas to students and the latter being more suitable for students to share and discuss their views with expert [7, 8]. Such approach also can be extended to student-student interactions from different cultures to enrich the multidimensional experience of learning astrology. The participants can gain intellectual knowledge in astrophysics concepts and technological skills in handling the Astronomy Online Labs application. Moreover, the participants are exposed to new cross-cultural interaction that they can benefit from conversing with participants who have different culture and values for social enhancement. It is believed that the exposure can elevate the participants multifaceted learning experiences at the cognitive and emotional level [9-11].

However, the conversation must be conducted in English, which for local participants English is not their mother tongue. Some may feel ashamed, lack of self-confidence, awkward, and many debilitating effects are experienced when using the second language as a medium of communication. Researchers reach a consensus that anxiety may affect the performance of the students [12-15]. This effect may be noticed during the cross-cultural discussion where the native speakers are more confident and have the tendency to dominate the interaction. Such notion is supported by Gunawardena et al. [16-17] who suggested that the language barriers for non-native speakers may cause handicap to equal participation.

Based on Law and Nguyen-Ngoc finding [18], cross-cultural interaction can be measured using three main factors, namely; cognitive, affective and social. Cognitive refers to the ability on the groups perform the given task using their intellectual knowledge and skills. The effect of the task to the participants' feeling of either positive emotions (joy, satisfied, calm) or negative emotions (frustrated, stressed, upset) can be grouped as affective factor. Moreover, social factor encompasses the participation of the participants in collaborating to complete the task [19-20]. Hence, in this work, students' discussions are quantitatively measured to see the knowledge transferred among the participants during the discussion. Four different data analytics techniques are applied to the text data to unearth the insights that may give us further details on the discussion, namely: word cloud [21-22], word forest [23], timelines [24] and Venn diagram [25]. From the understanding, a dynamic conversation analysis can be proposed so it can provide dynamic analysis on the go while interaction happens so the moderator can steer the interaction to the intended topics.

This paper is organized in the following manner. Section 2 describes the research methodology employed in this work. The different data analytics techniques are described in detail. The experimental set-up, result and discussion are provided in Section 3. This paper is concluded in Section 4 with summary and future direction of the research.

2. RESEARCH METHOD

There are 31 local and 10 non-local participants joined the study. The local participants consist of 31 first year students from Universiti Teknologi Malaysia who enrolled for Field Astronomy course of SGHU1203 whereas the non-local participants are from Space Research Centre, Leicester University on voluntary basis. Hence, the number of non-local is smaller compared to local participants because the local participants are compulsory to join the discussion for the purpose of earning marks for their coursework.

Table 1. The total number of participants for each group and session

Group	Session 1		Session 2		Session 3		Session 4	
	L	NL	L	NL	L	NL	L	NL
1	4	2	3	2	7	1	4	1
2	3	3	1	1	x	x	3	1
3	3	3	3	2	x	x	6	2
4	4	2	x	x	5	2	5	2
5	3	3	3	2	x	x	x	x

Four live sessions had been conducted through the collaborative learning platform on 9th and 20th March 2018 respectively to discuss about one group task on 'Introduction to Orbit' and an individual assignment using the video conferencing tool. The participants are divided into 5 groups and each group must interact in all the four sessions. The participants from UTM are denoted as Local (L) and on contrary the participants from Leicester University are labelled as Non-local (NL) participants. The participants acquired

scientific knowledge as well as the skills in collaborative learning, effective communication and critical thinking. Based on Table 1, it shows that for each session, small groups of students were formed in order to discuss the task given. During the discussion, they can verbally communicate or use chatting function to adopt the collaborative learning. The x symbol in Table 1 shows the missing data that are corrupted after the data has been collected. For each session, as per discussed, there are minimum of 2 and maximum of 8 students that had been participated in one group. The overall process of research methodology is shown in Figure 1.

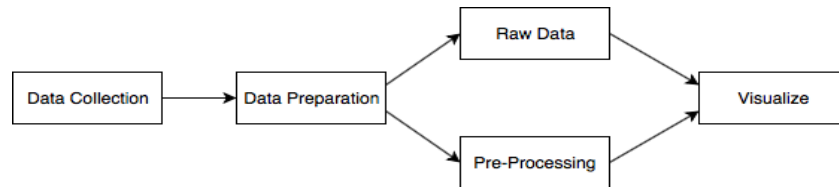


Figure 1. Research methodology process

As explained in the previous section, the data collection phase is done once the fourth session for each group is finished. Supposedly there are videos with audio and some text conversation being recorded, however, some of the data are corrupted, hence, only text conversation is being analysed in this paper. The data is prepared by converting all the text conversation data into small letters to ease the comparison between the data in database. Once the data has been converted, the process to visualize the data is divided into raw and cleaned data. The reason to divide the data into raw and cleaned is because we are trying to see the differences if some unnecessary words are removed. The raw data includes all the text as what has been conversed in the during the session. For the cleaned data, the pre-processing involved in order to clean the data started with removing all the symbols from the text conversation. Then, the data is tokenized to compare with the database of stop words. If the stop word exists, it will be removed from the data and the existing data is appended into a cleaned textfile. The visualization process is conducted once the data are processed. There are four types of visualization representation selected to represent the raw and cleaned text conversation data which are word cloud [21-22], word forest [23], timeline [24] and also the Venn diagram [25].

3. RESULTS AND ANALYSIS

In this section, the results of research in visualization diagrams are presented with relevant discussion. Table 2 illustrated the summary of the overall data used for the analysis. It consists of 12,634 words that comprises from 5 groups of students' discussions. The words are also segregated based on the speaker of either local or non-local to see the interaction and the cultural assimilation between these two groups.

Table 2. Summary of the Data Overall

	Total	Unique word		Total	Unique word
Raw	12634	1740	Non-local	4443	946
			Local	8191	1393
Cleaned	6831	1622	Non-local	2317	841
			Local	4514	1282

3.1. Word cloud

Word cloud visualization is used to display most occurrence words in the text data [21-22]. The size of the words in the visualization represents its frequency in the document. The more frequent the words are used, the bigger size it becomes. The representations will be able to highlight words that are commonly used in the conversation. The word clouds are generated for overall data for both local and non-local participants as shown in Figure 2 and Figure 3. For each category, two word-clouds are generated, one using the raw data and another one after removing stop words using python NLTK library. The process is then repeated for each individual group. Based on the word clouds, it can be observed that 'think' is the most used words by the non-local participants and 'ok' is the most used words by the local participants. It shows that non-locals love

to express their opinions on the discussed topic or subject while the locals will mostly agree with the given opinions.



Figure 2. Word cloud for (a) Overall raw data (b) Cleaned data



Figure 3. Word cloud for (a) Non-local (b) Local cleaned data

3.2. Word forest

Word Forest visualization is a text visualization technique which combines word cloud and text network visualization [23]. The visualization able to show significant relationships between the words in the textual data. The word forests are also generated for the same data category as in the word cloud visualization. From the word forest in Figure 4, it shows that the participants are discussing the tasks given in the assignment. Some of the most used phrases in the discussion are ‘I think’, ‘do you’, ‘we can’ and ‘we have to’. Other phrases are related to the formula being used to solve the given tasks.

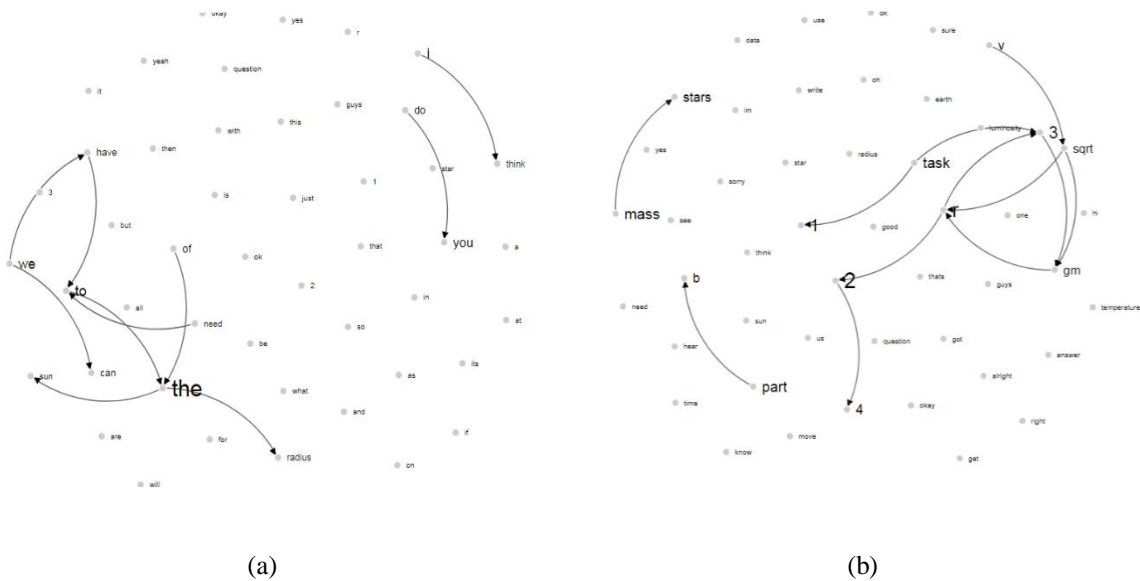


Figure 4. Word forest for (a) Overall raw data (b) Cleaned data

3.3. Timelines

The purpose of timeline visualization is to show the participation of local and non-local participants in the discussion. The orange boxes represent the local while the blue boxes represent the non-local. Arrangement of the boxes is based on the conversation and the word count is used to calculate the length of the boxes in the visualization.

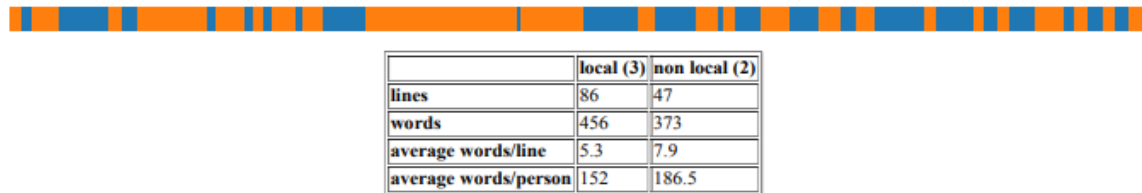


Figure 5. Example for timeline for group 1 session 2

Figure 5 shown the output of timeline visualization for Session 2 of Group 1. The number of participants for each category is shown in the table header inside the bracket. From the timeline generated, local participants dominate the first half of the discussion and the non-locals dominate from the middle until the end of the discussion. Average words per line and average words per person are also calculated to see how active the participant is in the discussion. From the results obtained, non-locals have more words per person compared to the locals in 14 out of 15 sessions even though they are in smaller number in all sessions.

3.4. Venn diagram

Venn diagram is a visualization technique used to show logical sets of elements as circles within an enclosing rectangle (the universal set). Common elements between the sets are being represented by intersections of the circles. Elements that will be chosen for the visualization is the unique words used in the discussion. Top 30 words used by local and top 30 words used by non-local will be selected as the elements of the Venn Diagram. Using Venn diagram, we are able to see overlapping word used by local and non-local. Then, the words are segregated into 4 categories to further dissect the students' discussion. The categories are astronomy, social, administrative and unclassified. The astrology class is labeled for words related to astrology; for instance: 'radius', 'velocity' and 'sun'. Since participants in this study are basically stranger to each other especially for intergroup communication (Local vs Non-local) salutation and greeting are the conversational starting point. Possible words for this class are 'hello', 'hi' and 'thank you'. The administrative class refers to the word of asking about the task, such as 'can', 'do' and 'not sure'. Subsequently, any words that are not segregated in the previous categories will be labeled as unclassified.

We proposed a Venn diagram approach to see the words uttered by the participants based on their cluster of either Local and Non-local. The overlapping region can be used to observe the similar words used by both groups whereas the word populated in the group region are the unique words used by that particular group. Figure 6 shows the Venn diagram generated using the cleaned data for overall discussion. It presents the 30 most frequently used words for both Local and Non-local participants. The left and right regions of the Venn diagram show the unique words used by Local and Non-local participants whereas the middle Venn diagram are the words used by both groups.

Based on the word segregation being done based on the output of venn Diagram, some interesting insights can be observed. From the 30 words that are frequently used by each group of participants, 17 are used by both and 13 are unique for each group. Table 3 classifies the word used by the Local, Non-Local and both groups according to the word labels.

From Table 3, the astrology words are used only 20.93% as opposed to the social and administrative words with frequency used of 34.88% and 32.56% respectively. The lowest word class usage percentage is unclassified class with 11.63%. The result show that the participants are focusing on being courteous in order to be accepted by the peers in their communication and more concern with the task at hand of solving the assignment given. Hence, it is not surprising if both word labels are being uttered more than two quarter of the conversation with 67.44%. It is also observed that Non-Local participants are being more vocal on astrology words whereas Local participants seems to be more agreeable with social word uttered of 18.6%. Based on the result, the Non-Local participants are more dominant in the astrology field as compared to the Local participants. A moderator may need to be utilized to facilitate and steer the discussion towards the astrology discussion rather than just merely being nice to each other.

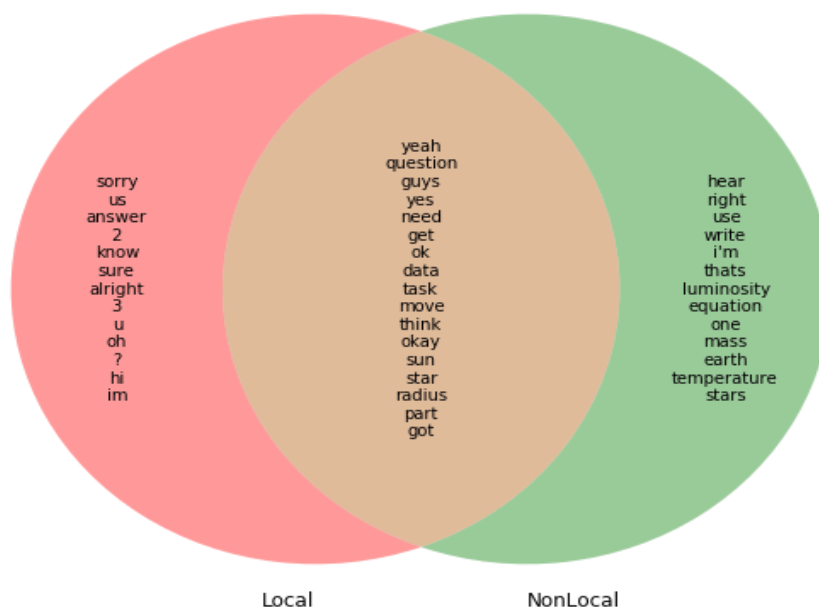


Figure 6. Venn diagram for overall raw data

Table 3. Summary of the data group 2

Word label	Local	Non-local	Both
Astrology	-	luminosity, equation, mass, earth, temperature, stars	sun, star, radius
Social	sorry, us, sure, alright, im, u, hi, oh	right, i'm	yeah, guys, yes, ok, okay
Administrative	answer, know	hear, use, write,	question, need, get, data, task, move, think, got, part
Unclassified	2, 3, ?,	thats, one	-

4. CONCLUSION

This present work studied the findings of students’ performance and capacity building in a cross-cultural computer-supported collaborative learning through Astronomy online labs (C3AOL). Results from the analysis indicates that both local and non-local students are able to mutually understand each other, e.g. the word ‘ye’ as typed by local participants, may be clarified as ‘yeah’ by non-local participants. The positive attitudes and the willingness of students to engage with the native speakers were also showed that the interaction of cross-cultural exists.

The findings from this project indicated that the Malaysian students can perform well and accepted by their counter partner in collaborative learning via the virtual session. They viewed that they can developed and cooperated well during team work, together with obtained self-confidence, self-independence and cultural awareness. However, it is obvious that the Non-local participants dominate the astrology discussion and the Local participants seems to have high acceptance level and agreeableness. To foster good discussion, a moderator may be the favourable solution to steer the discussion to be going in the right track and boosting the confidence of the local participants to be more engaging and actively inquiring. With this tool it is hope that cross-cultural with computer-supported online course will become more available, easily accessible and popular in line with the rapid development of information and communication technologies. Such approach can be embedded in learning analytics assessment system to tailor to individual needs [26-27].

ACKNOWLEDGEMENTS

The authors would like to thank to the Research Council, UK, Ministry of Education Malaysia (KPM) for providing financial support through the Newton Ungku Omar Fund (Vote No. R.K. 130000.7341.4B296) and Universiti Teknologi Malaysia (UTM) for assisting in the research management service and to conduct the work published in this paper.

REFERENCES

- [1] J. G. S. Goldie, "Connectivism: A Knowledge Learning Theory for the Digital Age?", *Medical Teacher*, vol. 38(10), pp. 1064-1069, 2016.
- [2] M. Warschauer, "Computer-mediated Collaborative Learning: Theory and Practice", *The Modern Language Journal*, vol. 81(4), pp. 470-481, 1997
- [3] A. Chioran, "Why is Collaborative Learning Important?", [Online] Available at: <https://www.nuiteq.com/company/blog/why-is-collaborative-learning-important> [Accessed 28 Jan. 2019].
- [4] M. Laal and S. M. Ghodsi, "Benefits of Collaborative Learning", *Procedia-social and Behavioral Sciences*, vol. 31, pp. 486-490.
- [5] L. S. Vygotsky, "Mind in Society: The Development of Higher Psychological Processes", Harvard University Press, 1980
- [6] W. Crain, "Theories of Development: Concepts and Applications", 6th ed. Upper Saddle River, NJ:Prentice Hall, 2010
- [7] H. Bound, "Developing Quality Online Dialogue: Dialogical Inquiry", *International Journal of Teaching and Learning in Higher Education*, vol. 22(2), pp. 107-124, 2010
- [8] A. M. Sharp, "The Community of Inquiry: Education for Democracy", *Thinking: The Journal of Philosophy for Children*, vol. 9(2), pp. 31-37, 1991.
- [9] D. H. Schunk, "Self-efficacy and Academic Motivation", *Educational Psychologist*, vol. 26(3-4), pp. 207-231, 1991.
- [10] A. Bandura, "Social Cognitive Theory of Personality", *Handbook of Personality*, vol. 2, pp. 154-196, 1999.
- [11] M. J. Elias, J. E. Zins and R. P. Weissberg, "Promoting Social and Emotional Learning: Guidelines for Educators", ASCD, 1997
- [12] E. Horwitz, "Language Anxiety and Achievement", *Annual Review of Applied Linguistics*, vol. 21, pp. 112-126, 2001.
- [13] M. Azher, M. N. Anwar and A. Naz, "An Investigation of Foreign Language Classroom Anxiety and Its Relationship with Students' Achievement", *Journal of College Teaching and Learning*, vol. 7(11), pp. 33, 2010.
- [14] E. M. Phillips, "The Effects of Language Anxiety on Students' Oral Test Performance and Attitudes", *The Modern Language Journal*, vol. 76(1), pp. 14-26, 1992.
- [15] P. D. MacIntyre, "How does Anxiety Affect Second Language Learning? A Reply to Sparks and Ganschow", *The Modern Language Journal*, vol. 79(1), pp. 90-99, 1995.
- [16] C. Gunawardena, "Culture and Online Education", In M. Moore & W. Anderson (Eds.), *Handbook of Distance Education*, Mahwah, NJ: Lawrence Erlbaum Associates, Publishers, pp. 753-775, 2003.
- [17] C. Gunawardena and D. LaPointe, "Social and Cultural Diversity in Distance Education", In T. Evans, M. Haughey & D. Murphy (Eds.), *International Handbook of Distance Education*, London, UK: Emerald Group Publishing Limited, pp. 51-70), 2008.
- [18] A. V. Nguyen-Ngoc and E. L. C. Law "Evaluation of Cross-cultural Computer-supported Collaborative Learning: Preliminary Findings for iCamp Challenges", In *EdMedia: World Conference on Educational Media and Technology*, pp. 1887-1896, 2007.
- [19] G. Conole and P. Alevizou, "A Literature Review of the Use of Web 2.0 Tools in Higher Education", *A Report Commissioned by the Higher Education Academy*, 2010.
- [20] P. Van den Bossche, W. H. Gijsselaers, M. Segers and P. A. Kirschner, "Social and Cognitive Factors Driving Teamwork in Collaborative Learning Environments: Team Learning Beliefs and Behaviors", *Small Group Research*, vol. 37(5), pp. 490-521, 2006.
- [21] F. Heimerl, S. Lohmann, S. Lange and T. Ertl, "Word Cloud Explorer: Text Analytics based on Word Clouds", In *47th Hawaii International Conference on System Sciences*, pp. 1833-1842, 2014.
- [22] S. Lohmann, F. Heimerl, F. Bopp, M. Burch and T. Ertl, "Concentri Cloud: Word Cloud Visualization for Multiple Text Documents", In *19th International Conference on Information Visualisation*, pp. 114-120, 2015.
- [23] M. Bakri, S. Z. Z. Abidin, N. Omar, M. Hamiz and A. Razali "Word-forest Visualization of Discussed Topics in Social Media Comments", *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, vol. 10(1-6), pp.109-112, 2018
- [24] O. Hoeber, L. Hoeber, M. El Meseery, K. Odoh and R. Gopi, "Visual Twitter Analytics (Vista) Temporally Changing Sentiment and the Discovery of Emergent Themes within Sport Event Tweets", *Online Information Review*, vol. 40(1), pp. 25-41, 2016.
- [25] R. Talib, M. K. Hanif, S. Ayesha and F. Fatima, "Text Mining: Techniques, Applications and Issues", *International Journal of Advanced Computer Science and Applications*, vol. 7(11), pp. 414-418, 2016.
- [26] M. Hamiz, M. Bakri, N. Kamaruddin and A. Mohamed, "Assessment Analytic Theoretical Framework Based on Learners' Continuous Learning Improvement", *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 11(2), pp. 682-687, 2018.
- [27] A. Wahab and N. Kamaruddin, "Brain Developmental Disorders' Modelling based on Preschoolers Neuro-Physiological Profiling", *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12(2), pp. 542-547, 2018

BIOGRAPHIES OF AUTHORS

M. Bakri is a lecturer in Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Melaka, Kampus Jasin, Merlimau, Melaka, Malaysia.



Assoc. Prof. Ts. Dr. Norhaslinda is currently holds a post of associate professor in Faculty of Computer and Mathematical Sciences, University Teknologi MARA, Shah Alam, Selangor, Malaysia.



M. Hamiz is a lecturer in Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Cawangan Melaka, Kampus Jasin, Merlimau, Melaka, Malaysia.



Assoc. Prof. Dr. Marlia Puteh is currently an Associate Professor at Language Academy, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia.



A.H.S. Nurhasmiza is a lecturer at Language Academy, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia.



Z. Othman is a senior lecturer at the Geoinformation Department, Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia, Johor, Malaysia.



N.A.S. Nilam is a lecturer at Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia.