

Relationship Analysis on the Experience of Hospitalised Paediatric Cancer Patient in Malaysia using Text Analytics Approach

Zuraini Zainol¹, Puteri N.E. Nohuddin², Nadhirah Rasid³, Hamidah Alias⁴, A. Imran Nordin⁵

Department of Computer Science, Universiti Pertahanan Nasional Malaysia

Faculty of Science and Defence Technology, Sungai Besi Camp 57000 Kuala Lumpur, Malaysia¹

Institute of Visual Informatics, Universiti Kebangsaan Malaysia, Bangi 43600 Selangor^{2,3,5}

Department of Paediatrics, Faculty of Medicine, Universiti Kebangsaan Malaysia, Cheras 56000 Kuala Lumpur⁴

Abstract—The purpose of this study is to analyse the keyword relationships of paediatric cancer patient's experiences whilst being hospitalised during the treatment session. This study collects data through 40 days of observations on 21 paediatric cancer patients. A combination of text analytical visualizations such as network analysis map and bubble graph to analyse the data is applied in this study. Through the analysis, keywords such as “cri” (crying), “lay” (laying), “sleep” (sleeping) and “watch” (watching) are found the most common activities that have been experienced by paediatric cancer patients when they were hospitalised. Based on this observation, it can be argued that these activities can be represented as the experience that they have whilst being in the hospital. Based on the findings, hospitalised paediatric cancer patient's experience is limited due to the treatment protocol that requires them to be attached to intravenous line. Therefore, most of their activities are focused in bed such as sleeping, playing with their mobile, watching video and so on in bed. This study also offers a novel approach of transforming cancer patient data into useful knowledge about keyword relationship in paediatric cancer patient's experience during their stay in the hospital. The incorporation of these two text analytics offers insights for researchers to understand the interesting hidden knowledge in the collection of unstructured data, and this information can be used by medical providers, psychologists, games designers and others to develop any applications that can assist their difficulties and ease their pain while warded in the hospital.

Keywords—Patient experience; paediatric; keyword relationship analysis; bubble graph; text network analysis

I. INTRODUCTION

Millions of people have been diagnosed with cancer around the world affecting people regardless of age. Every year about 300,000 new cancer related cases are diagnosed in children and teens under the age of 19 every year [1]. Similarly, Ferlay et al. [2] report that 10.3% of Malaysians are at risk to die before the age of 75 years old because of cancer related illness and 10% of cancer sufferers are children [3]. Many factors could be the cause for these unfortunate children to have cancer for example genetic syndromes, low immune system and high doses of ionizing radiation are known cause of childhood cancer. Moreover, environmental factors including infection and immune response could be of risk factors [4, 5].

Today, the advancement of medical technology and the latest findings in drugs for cancer have contributed to the betterment of paediatric cancer management. The survival rates of child cancer patients are increasing but the diagnosis and the treatments provided can be extremely stressful and the treatment protocols planned for them can affect the life of a child and their family [3]. Available cancer treatments such as chemotherapy, surgery, radiotherapy, bone marrow transplantation and immune therapy amongst others, all of which may have an affect towards children's physical and psychological well-being [6, 7]. The physical and psychological affect could be caused by the medical and physical effects, psychological effects, and cognitive and neuropsychological effects.

Typically, paediatric cancer patients are required to be hospitalised during the treatment session. The unpleasant feeling to stay in the hospital is argued as one of the most frequently stated negative experience with the journey of paediatric cancer patients. Indeed, hospitalisation itself is a stressful situation, and a patient's stress level could increase significantly as the length of hospitalisation increases. Psychosocial effects could manifest as increased levels of anxiety, and concerns about dying which could lead to depression [8].

Based on the findings, hospitalised paediatric cancer patient's experience is limited due to the treatment protocol that requires them to be attached to intravenous line. Therefore, most of their activities are focused in bed such as sleeping, playing with their mobile, watching video and so on in bed. This study also offers a novel approach of transforming cancer patient data into useful knowledge through keyword relationship analysis in paediatric cancer patient's experience during their stay in the hospital. The incorporation of these two text analytics offers insights for researchers to understand the interesting hidden knowledge in the collection of unstructured data. Therefore, this information can be used by medical providers, psychologists, games designers and others to develop any programs or applications that can assist their difficulties and ease their pain while being warded in the hospital.

In this paper, the authors describe a work done to analyse the experience of hospitalised paediatric cancer patient in

Malaysia. The relationship analysis in the text analytics is used to analyse the qualitative data collected through ethnography study done at the paediatric oncology ward at Hospital Canselor Tuanku Mukhriz, Cheras, Kuala Lumpur, Malaysia. The results from the analysis would help to visualise the experience that paediatric cancer patients have whilst being warded in the hospital. The related work on paediatric cancer patients, patients experience in the hospital, relationship analysis and text analytics are described in Section II. The methodology of the study is presented in Section III. The results and the discussion of findings are presented in Section IV. Finally, Section V concludes this study.

II. BACKGROUND AND RELATED WORKS

Cancers in children are quite different from cancers affecting adults [9]. They tend to occur in different parts of the body, they look different under the microscope and respond differently to treatment. With the advancement of cancer treatments, many cancers can be cured. Doctors incline to practice the word 'remission' rather than the word 'cured'. Remission means there is no evidence of cancer following treatment. However, sadly, in some cases a cancer returns months or years later. In another research, it was reported that younger patients with cancer have greater fear of cancer recurrence (FCR) due to unpredictable disease outcome that leads to depression and anxiety [10]. Most pediatric cancer patients face with many kinds of experience, especially in the ward. Thus studying the experiences of pediatric patients are also important as authors are able to propose better services or care from human assistance like book reading and talking companion [11, 12] or technological assistance such as computers and games [13].

Patient experience is an increasingly important factor for improving services and designing products in many sectors like healthcare [14], computer gaming [13] and customer hospitality [15]. In this study, it focuses patients' experiences in healthcare treatments. Boote, Telford, and Cooper [16] pointed the prominence definitions of patient experience by defining terms 'disease' which is a physiologic and clinical abnormality and 'illness' which is defined as the subjective experience of the patients. While healthcare professionals can be considered the experts in disease, patients and healthcare consumers can be considered experts in illness which are the experiences of having the disease [17]. Boote et al. also argued that when both healthcare professionals and patients' perspectives are considered together, the analysis may offer a synergized knowledge which could offer new insights towards improving the healthcare industry. Studying patient experience has proven to be useful, in a number of areas. Initially, at the individual level in which private hospitals with patient-centered care has been championed for some years within the healthcare industry [18] and there is evidence that it can improve outcomes for patients, including patient safety, patient satisfaction and anxiety reduction [19]. Another area is that investigating patient experience can lead to a product design, for example computer games specifically designed and developed to assist patients endure the pains during their treatments [20, 21]. This paradigm of patient-physician experiences and interaction focuses on the inclusion of the

patient in decisions about their care, and in relating the evidence about disease to their experience of illness in order to help make the most appropriate and also personalised treatment decisions for that individual.

Thus, a number of research used statistical method to analyse the patient experience data [14]. Nonetheless, data mining (DM) techniques were also commonly used in healthcare. In healthcare systems, massive data is available and data mining techniques are suitable to be used as a tool to analyze those mass data like classification such as rule set classifiers, decision tree algorithms, neural network architecture, neuro-fuzzy, and Bayesian network structure discovery techniques. DM is a well-known method to extract non-trivial knowledge from raw data in relation to strategic decision makings, prediction and insights. There are many techniques in DM such as Classification, Clustering, Association Rules, Bayesian Networks, and Decision Tree [22-25]. Galatas *et al* introduced a novel hybrid technique of feature selection and Naïve Bayes to determine patient satisfaction as significant indicators for decision makings and improving quality services of healthcare [14]. Another study was developed and tested a learning-based text-mining approach to facilitate analysis of patients' experiences of care and develop an explanatory model illustrating impact on Health-related quality of life (HRQOL) [26]. In hospitality domain, DM techniques are used to analyze customer behavior and predict the possible behavior of expected clients by using Classification, Regression, Link analysis, and Segmentation [14].

According to [27], Text Mining (TM) can be defined as an analytic process that is designed to explore the unstructured text documents in search of useful information and knowledge hidden from a large amount of text resources. TM has been widely applied in many application domains such as military knowledge [28-31], business analytics [32-34], documents analysis [35-37], social media issues [38-40], etc. In this research, TM which is one of DM method is explored as to find new knowledge through examining qualitative data of patient experiences while having chemo treatment [20, 21]. Through automation, TM approaches offer high throughput systems which can analyse much larger document collections than would be feasible to inspect manually. The advantages of this technique are two-fold: it is more likely to (i) detect non-trivial patterns, and (ii) lead to more statistically significant results. However, text interpretation requires both linguistic and domain knowledge, which make its automation a challenging task.

TM is corresponding to DM in its approach to knowledge extraction through identifying and analysing text patterns [41]. Where DM deals with data held within structured databases, TM approaches aim to find patterns in unstructured textual data. Unlike structured data, textual data and natural language require complex analysis in order to understand its content. Therefore, in textual data, normally context based keywords are communicated to the reader through the language used and the assumption of some background knowledge used of the users for interpretation. This context is difficult to ascertain in an automated process. Terms extraction and relationship analysis are used to extract knowledge from sets of

unstructured text data. A research is conducted to investigate the relationships between verses and chapters at the keyword level in a Malay translated Tafseer. A combination technique of TM and network analysis is developed to discover non-trivial patterns and relationships of verses and chapters in the Tafseer. This is achieved through keyword extraction, keyword-chapter relationship discovery and keyword- chapter network analysis [42].

Another study reviewed substantial sets of cancer publications which involved expeditious growth of biomedical text [43]. This has lead the grown of TM techniques which are used to extract information about cancer diagnostics, treatment and prevention from the unstructured biomedical text and focused on presenting basic concepts of TM algorithms, tools and data sets [43]. These findings could theoretically assist many researchers to elect appropriate TM tools and datasets. Discussions on applying TM techniques to support cancer systems biology research are also been reviewed.

III. FRAMEWORK OF DOCUMENT AND KEYWORD RELATIONSHIP ANALYSIS

This study applies the enhancement of the proposed framework by Rasid et al. [21]. The optimised framework is applied to generate keyword relationships from the paediatric cancer patient observation forms which is recorded during their hospitalisation. The observation of paediatric cancer patients was conducted at a randomly chosen time-slot during the day; for example, in the morning (7 am to 3pm), afternoon (3pm to 11pm) and in the evening (11pm to 7 am). Fig. 1 shows the framework of document and keyword relationship analysis (FDKRA). The framework comprises of 2 main components: (i) Document Pre-processing and Keyword Extraction Analysis module and (ii) Keyword Relationship Visualisation module to present the experiment findings. Document and Keyword Extraction Analysis is the main analysis engine for extraction and ranking of documents and keywords / term relationships from the paediatric cancer dataset. Finally, the discovered document and keyword relationships are presented using two visualisation types including network graph and bubble graph.

In this study, the data was collected through the shadow observations method conducted in the paediatric oncology ward, 4D at Pusat Perubatan Universiti Kebangsaan Malaysia (PPUKM), Hospital Canselor Tuanku Muhriz, Cheras, Kuala Lumpur. The shadow observations were aim to collect paediatric cancer patient's activities/routines while they are hospitalised. All information gathered from the shadow observations are recorded using the observation form. The recorded data are then transcribed in the Excel format for data analyzing. The observation form is completely anonymous which only contains number of observation, patient's age, gender, days in ward, bed number, time, and treatment. The

dataset underwent text pre-processing to ensure correct format for different terms and keywords before conducting text analysis.

The next element in the FDKRA describes the representation of the document, which refers to patients and keyword findings and analysis. Based on the keyword pattern found in the documents, the related patterns within a group are discovered. The relationship of the keywords should represent the content of patients' observation. This study is a part of the second component of FDKRA, focussing on grouping of selected keywords and visualising them as important keywords from experiences among paediatric cancer patients.

A. Document Pre-Processing and Keyword Extraction Module

This module consists of documents pre-processing and keyword extraction. The dataset that comprises of 21 plain text documents need to be prepared. In text analysis, the document pre-processing is the most important step as the raw dataset is often incomplete, inconsistent, contains many errors, etc. Poor data quality will affect the accuracy of TM results. The pre-processing helps to improve the quality of data and also, the accuracy and effectiveness of text analysis. This module is developed based on the concept of TF-IDF, counting and ranking the words in the given content, followed by selecting words that occur more than the threshold. This module generates a DKM-TFIDF, which is a $m \times n$ matrix that represents text documents (observation data) versus terms (frequent keywords). DKM-TFIDF tracks the term frequency for each term in all the observations. Thus, DKM-TFIDF can become a very large sparse matrix, depending on the number of documents and number of terms in each observation. DKM-TFIDF representation is a method to represent the documents as numeric structures. Representing text as a numerical structure is a common starting point for TM and analytics, such as search and ranking, creating taxonomies, categorisation, document similarity and text-based machine learning.

B. Keyword Relationship Visualization Module

The module consists of three different types of visualisations: (i) Word Cloud, a graphical representation of keyword frequency. Keywords are usually single words, with the importance of each keyword is differentiated with font size or colour; (ii) Text Network Analysis Graph illustrates relationships between survey documents and keywords. Keywords are displayed as round nodes and lines are used to represent the relationships between them; and (iii) Bubble Graph exhibits each document and its keywords. Each document is presented in a bubble and keywords associated with the documents are represented in sub-bubbles, following the DKM-TFIDF. Sub-bubbles are differentiated with colours and sizes.

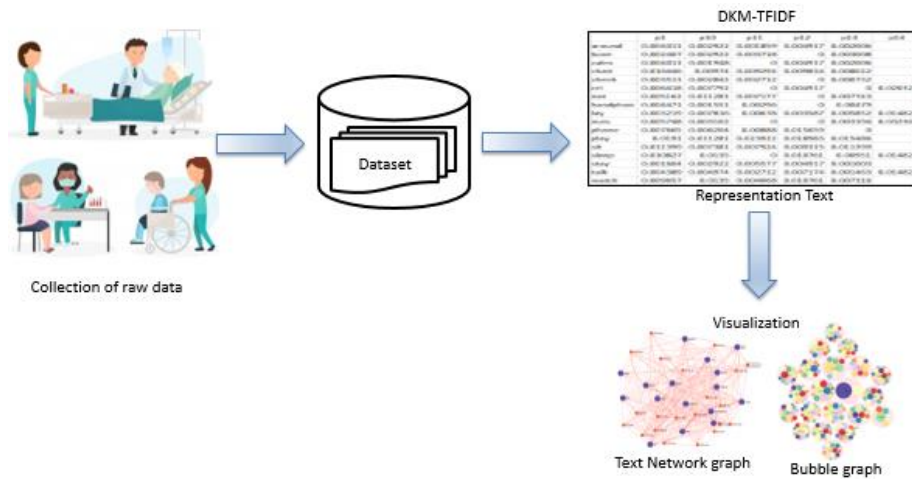


Fig. 1. Framework of Document and Keyword Relationship Analysis (FDKRA).

IV. RESULTS AND DISCUSSION

This study applies the enhancement of the proposed. This section presents the experimental results using FDKRA. In this experiment, a set of 21 patients from the transcription of observation data is applied as the input. Each patient corresponds to a single document. The transcribed data is prepared as plain text documents. Text pre-processing is an important step in most text mining techniques and applications. It prepares the input data for consequent analysis. Low quality of text data affects the accuracy of TM results. Most text-based documents are often very noisy containing typos, errors and multiple acronyms for the same word. The pre-processing task can improve both the quality of data and accuracy, and effectiveness of text mining.

As shown in Fig. 2, the transcribed observation data is converted into a collection of text documents or corpus using the “tm” package in R. The text document is cleaned by removing numbers, symbols, punctuation marks, whitespace, etc., and converted all text into lowercase for standardisation. This is to ensure that multiple form of keywords such as “sleep” or, “SLEEP” are treated similarly in the experiment. After that, the words are tokenized by breaking up the text into discrete words. The next step is to remove all stop words (e.g., prepositions, pronouns, conjunctions, etc.) and reduced the existing words to their stems. This is to ensure that only the root of the word is presented in the document keyword matrix (DKM). In this study, some frequent keywords such as “mother”, “father” and “patient” are removed as these words

are found common in the dataset. Besides that, the custom stop words including “get”, “can”, “im”, etc. are also removed from the dataset. These stop words are specific to the dataset that may not contain value to the dataset. In this experiment, the SnowballC package is applied for text document stemming. Such words “boring”, “bored” and “bores” will be reduced to “bore” after stemming.

Most of TM tasks require data to be represented in the form of a matrix or vector – document term matrix (DTM) or term document matrix (TDM). The DTM or TDM describes the frequency of terms/keywords that occur in a collection of documents. As illustrated in Table I, the Document Keyword Matrix (DKM) consists of 475 terms (correspond to columns) extracted from 21 text documents (correspond to rows) with 84% sparsity. Sparsity refers to the threshold of relative document frequency for a term. The table below shows that 84% of the row entries in DKM contains zero entries. In other words, most keywords on DKM do not appear in most of text documents. For example, the keywords “small” (1), “son” (1), “still” (1), “teach” (1) and “switch” (1) are marked as zero in most text documents. Therefore, these less frequent keywords need to be removed.

Table I shows the results of removing sparse in DKM, with 38 keywords extracted from 21 documents with 31% sparsity which means that 31% of the entries contains zeroes (0). In other words, a 38 x 21 matrix is created representing 38 unique keywords and 21 text documents.

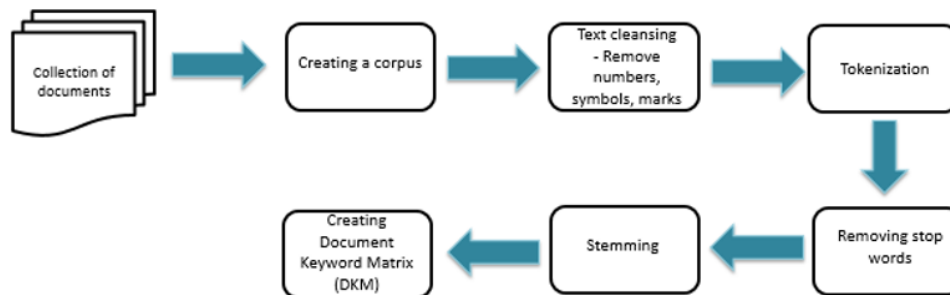


Fig. 2. Overview of Text Pre-Processing Process.

TABLE. I. PARTIAL DKM AFTER REMOVING 31% SPARSITY

Docu- ments	Keywords									
	accompan	around	back	bed	Blanket	bore	bring	stand
p1	3	10	6	97	11	6	1	0
p2	0	1	0	6	2	2	0	1
p3	3	6	2	41	5	8	2	1
p4	4	7	2	53	4	11	3	0
....
....
p21	1	0	0	2	0	0	0	1

Fig. 3 shows the visualisation of most frequently used keywords within the text documents, in the form of a word cloud. Word cloud is easy to understand and readable as it provides multiple choices of colours that symbolizing the keywords with different sizes. The size of keywords corresponds to the frequency of the terms. The larger font size corresponds to a higher frequency value. Based on our observation, the keywords “bed” (671), “sit” (348), “watch” (296), “sleep” (213), “play” (205), and “chair” (197) have the highest frequencies in the dataset.

After the text documents are pre-processed, the DKM is transformed into a TF-IDF representation, which highlights important terms in the survey. The term-weighting statistics is applied for identifying important keywords in a collection of text documents. Each keyword is assigned a weight, which represents its importance in the text document (Table II). The infrequent keywords with less weight are discarded. Thus, the list of terms in a text document can be arranged according to its importance. In this experiment, a subset of the most important terms is selected as keywords. The DKM-IDF also becomes an input for the network analysis map and bubble graph.

Table II shows the results of removing sparse in DKM-TFIDF, with 16 keywords extracted from 21 documents with 30% sparsity. In other words, a 16 x 21 matrix is created representing 16 unique keywords and 21 text documents. The new DKM-TFIDF becomes an input for the network analysis graph and bubble graph for keywords and patients.

Although the DKM-TFIDF consists of the summary of patients and their related keywords, it lacks of representing the visual data, particularly mapping the relationship between keyword and patients. Therefore, the DKM-TFIDF is transformed into a text network analysis. The text network analysis plots a text as a network graph where the nodes on the graph represents the specific keywords and patients (subjects). Fig. 4 illustrates the text network visualisation of the 21 patients with all the keywords. The blue nodes represent the 16 keywords such as “around”, “bore”, “calm”, “chair”, “check”, “cri”, “eat”, “handphon”, “lay”, “nur”, “play”, “sit”, “sleep”, “stay”, “talk” and “watch”. On the other hand, the red nodes represent the 21 patients in the observation session. The light brown lines represent the linkages between patients and keywords. The thickness of colours for each connection is represented by the value of the occurrence of keywords in related documents, as per the DKM-TFIDF. For example, it

can be clearly seen that the node (keyword) “cri” is linked to a group of nodes (patients) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 15, 16, 17 and 20. These connections represent the node (keyword) “cri” is a common activity that had been experienced by almost all paediatric cancer patients. This relates to the observation that most of the paediatric cancer’s patients are crying while receiving their chemotherapy treatments in ward.

Fig. 5 visualises a bubble graph that displays multiple bubbles of 21 patients with its important keywords developed from the DKM-TFIDF. This graph is developed using a Java Script. Each patient is represented in a pink bubble together with its related keywords. Each pink bubble (patient) may contain a group of multicolour sub-bubbles (selected important keywords). The size of sub-bubbles is dependent on the weight of a keyword (patient) and its associated important keywords. The legend on the left hand side contains 16 different coloured boxes indicating the type of keywords. The details of sub-bubbles can be further visualised by clicking on the selected bubble (patient).

For example, the bubble graph for patient 14 can be further visualised (see Fig. 6). The bubble that visualize patient 14 consists of five small multi-coloured sub-bubbles that relate to important keywords such as “nur” (nurse), “cri” (cry), “sleep”, “talk” and “lay”. A quick analysis of the bubble text visualization for keywords (“nur”-“cry”-“sleep”-“talk”-“lay”) shows that there is an important link between them. This relates to the observation that “The patient woke up from sleeping and crying and the nurse come to him. When his mother got back, he is just lying on the bed and talking to his mother”.



Fig. 3. Word cloud of the DKM with 100 Keywords.

TABLE II: PARTIAL DKM-TFIDF AFTER REMOVING 30% SPARSITY

Documents	Keywords									
	around	bore	calm	chair	Check	cri	eat	watch
p1	0.004011	0.002407	0.004011	0.016446	0.003511	0.006418	0.005162	0.009657
p2	0.004555	0.009100	0.001948	0.004550	0.001820	0.013277	0.004550	0.000000
p3	0.005081	0.006775	0.000847	0.010162	0.002471	0.001694	0.007628	0.012973
p4	0.004894	0.007691	0.01398	0.011187	0.002040	0.002797	0.014397	0.014282
....
....
p21	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.098079	0.000000

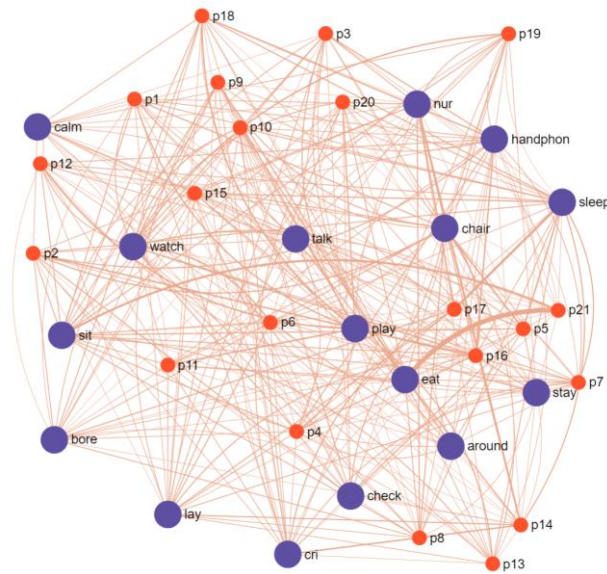


Fig. 4. DKM-TFIDF Text Network Analysis Map for Selected Important Keywords.

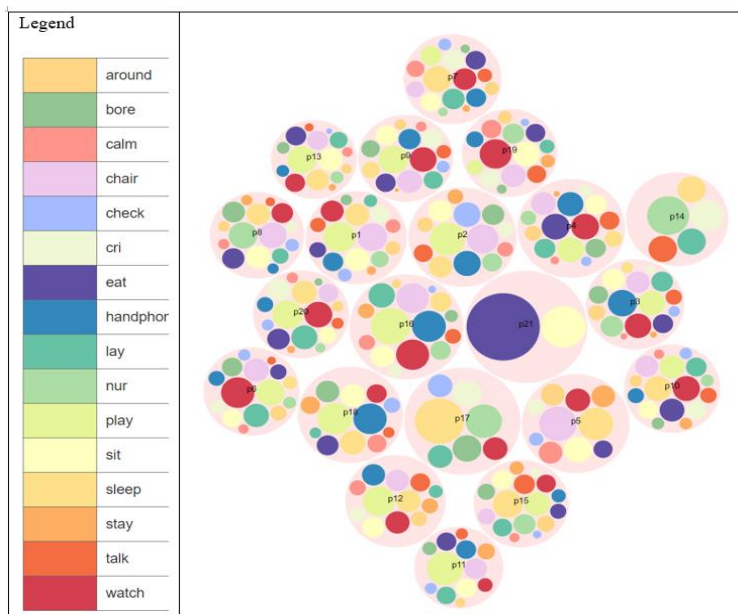


Fig. 5. Visualization of a Bubble Graph for 21 Patients and 16 Keywords.

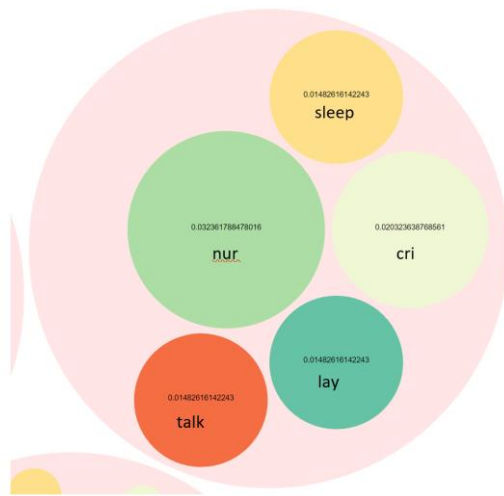


Fig. 6. Visualization of a Bubble Graph for Patient 14 with 5 main Keywords.

V. CONCLUSION

In this paper, the FDKRA framework was used to enable the collection and processing of the content of paediatric cancer patient's experience using two main components: (i) Document Pre-processing and Keyword Extraction Analysis module and (ii) Keyword Relationship Visualisation module to present the experiment findings. The results suggest that by using the keyword relationship visualisation module allows a clear presentation of unstructured data into a meaningful information. In our work, the collected observation data is hard to be interpreted given that the nature of the data is very unstructured. However, authors can present insight of the collected data to interpret patient's experience during their hospitalisation. The authors also argue that this strategy can be used for other qualitative study to allow researchers to have a quick understanding of the pattern of the collected data. This will ensure that the researchers would not miss any important pattern when they conduct the qualitative analysis.

ACKNOWLEDGMENT

The authors wish to thank the Universiti Pertahanan Nasional Malaysia (UPNM) and Pusat Keselamatan Siber (PKS), UPNM for funding and supporting this research. The authors also acknowledge the participation and cooperation received from patients and staffs from the paediatric oncology ward 4D, Hospital Canselor Tuanku Mukhriz (HCTM) Universiti Kebangsaan Malaysia and a little contribution from the FRGS Grant, Ministry of Education Malaysia (Grant No: FRGS/1/2016/ICT04/UKM/02/3) that has been contributed for the data collection phase.

REFERENCES

- [1] W. H. Organization, "International childhood cancer day: Much remains to be done to fight childhood cancer," Lyon, France: World Health Organization. Retrieved February, vol. 16, pp. 1-2, 2016.
- [2] J. Ferlay, H. R. Shin, F. Bray, D. Forman, C. Mathers, and D. M. J. I. j. o. c. Parkin, "Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008," vol. 127, no. 12, pp. 2893-2917, 2010.
- [3] L. Penkman, L. Scott—lane, and W. J. J. o. p. o. Pelletier, "A psychosocial program for pediatric oncology patients: a pilot study of "The Beaded Journey"," vol. 24, no. 2, pp. 103-115, 2006.
- [4] F. A. Cangerana Pereira, A. P. Mirra, M. d. R. Dias de Oliveira Latorre, and J. V. De Assunção, "Environmental Risk Factors and Acute Lymphoblastic Leukaemia in Childhood," *Revista Ciencias de la Salud*, vol. 15, no. 1, pp. 129-144, 2017.
- [5] P. A. Buffler, M. L. Kwan, P. Reynolds, and K. Y. Urayama, "Environmental and genetic risk factors for childhood leukemia: appraising the evidence," *Cancer investigation*, vol. 23, no. 1, pp. 60-75, 2005.
- [6] I. Hamzah, A. I. Nordin, H. Alias, N. Rasid, and H. J. A. S. L. Baharin, "Game Design Requirements Through Ethnography Amongst Pediatric Cancer Patients," vol. 24, no. 3, pp. 1567-1570, 2018.
- [7] K. Enskär, L. J. N. C. von Essen, and Y. People, "Physical problems and psychosocial function in children with cancer," vol. 20, no. 3, 2008.
- [8] I. Hamzah, A. I. Nordin, N. Rasid, and H. Alias, "Understanding Hospitalized Pediatric Cancer Patients' Activities for Digital Games Design Requirements," in *Proc. 5th Int. Visual Inform. Conf., (IVIC) 2017*: Springer, pp. 552-558.
- [9] CHOP. "Pediatric Cancers Differ from Adult Cancers, and Need Different Treatment Plans, Say CHOP Experts." <https://www.chop.edu/news/pediatric-cancers-differ-adult-cancers-and-need-different-treatment-plans-say-chop-experts> (accessed 26 July 2019)
- [10] J. V. Crist and E. A. Grunfeld, "Factors reported to influence fear of recurrence in cancer patients: a systematic review," *Psycho-Oncology*, vol. 22, no. 5, pp. 978-986, 2013.
- [11] N. Salahieh. "Storytelling Can Be Therapeutic for Children Battling Cancer." https://uscstoryspace.com/2017-2018/salahieh/Fall_Midterm/ (accessed 19 July, 2019).
- [12] "Chemo Companions." <http://chemocompanions.org/> (accessed 26 July, 2019).
- [13] M. Ghazisaeidi, R. Safdari, A. Goodini, M. Mirzaiee, and J. Farzi, "Digital games as an effective approach for cancer management: Opportunities and challenges," *J. Edu. Health Promotion*, vol. 6, 2017.
- [14] G. Galatas, D. Zikos, and F. Makedon, "Application of data mining techniques to determine patient satisfaction," in *Proc. 6th Int. Conf. on Pervasive Tech. Related to Assistive Environments, 2013*: ACM, p. 41.
- [15] I. Khan, R. J. Garg, and Z. Rahman, "Customer service experience in hotel operations: an empirical analysis," *Procedia-Soc. and Behav. Sci.*, vol. 189, pp. 266-274, 2015.
- [16] J. Boote, R. Telford, and C. Cooper, "Consumer involvement in health research: a review and research agenda," *Health policy*, vol. 61, no. 2, pp. 213-236, 2002.
- [17] M. P. Pomey, D. P. Ghadiri, P. Karazivan, N. Fernandez, and N. Clavel, "Patients as partners: a qualitative study of patients' engagement in their health care," *PLoS one*, vol. 10, no. 4, p. e0122499, 2015.
- [18] I. Gabutti, D. Mascia, and A. Cicchetti, "Exploring "patient-centered" hospitals: a systematic review to understand change," *BMC health services research*, vol. 17, no. 1, p. 364, 2017.
- [19] C. Doyle, L. Lennox, and D. Bell, "A systematic review of evidence on the links between patient experience and clinical safety and effectiveness," *BMJ open*, vol. 3, no. 1, p. e001570, 2013.
- [20] N. Rasid, P. N. E. Nohuddin, H. Alias, I. Hamzah, and A. I. Nordin, "Using data mining strategy in qualitative research," in *Int. Visual Inf. Conf. (IVIC), 2017*, pp. 100-111.
- [21] N. Rasid, P. N. E. Nohuddin, Z. Zainol, I. Hamzah, H. Alias, and A. I. Nordin, "Experience Mining Through Ethnography Study Among Pediatric Cancer Patients in Malaysia," *Adv. Sci. Lett.*, vol. 24, no. 3, pp. 1562-1566, 2018.
- [22] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.
- [23] M. H. Dunham, *Data mining: Introductory and advanced topics*. Pearson Education India, 2006.
- [24] S. Mourya and S. Gupta, *Data mining and data warehousing*. Alpha Science International, Ltd, 2012.

- [25] P. N. E. Nohuddin, Z. Zainol, A. S. H. Lee, A. I. Nordin, and Z. Yusoff, "A Case Study in Knowledge Acquisition for Logistic Cargo Distribution Data Mining Framework," *Int. J. of Adv. Appl. Sci., (IJAAS)*, vol. 5, no. 1, pp. 8-14, 2018.
- [26] R. Wagland et al., "Development and testing of a text-mining approach to analyse patients' comments on their experiences of colorectal cancer care," *BMJ Qual Saf*, vol. 25, no. 8, pp. 604-614, 2016.
- [27] Z. Zainol, P. N. E. Nohuddin, M. T. H. Jaymes, and S. Marzukhi, "Discovering "interesting" Keyword Patterns in Hadith Chapter Documents," in *Int. Conf. Information Commun. Tech. (ICICTM)*, pp. 104-108, 2016.
- [28] Z. Zainol, P. N. E. Nohuddin, W. M. U. Noormanshah, and M. H. A. Hijazi, "Visualization of Context-Based Keyword Pattern Cluster Analysis on Tacit Knowledge Among Officer Cadets at Universiti Pertahanan Nasional Malaysia (UPNM)," *Adv. Sci. Lett.*, vol. 24, no. 3, pp. 1550-1554, 2018.
- [29] Z. Zainol, S. Marzukhi, P. N. E. Nohuddin, W. M. U. Noormanshah, and O. Zakaria, "Document Clustering in Military Explicit Knowledge: A Study on Peacekeeping Documents," in *Proc. 5th Int. Visual Inform. Conf. (IVIC)*, 2017, pp. 175-184.
- [30] P. N. E. Nohuddin and Z. Zainol, "Discovering Explicit Knowledge using Text Mining Techniques for Peacekeeping Documents," *Int. J. Business Inf. Sys. (IJBIS)*, 2020, in press.
- [31] S. Marzukhi, N. H. Mohammad Daud, Z. Zainol, and O. Zakaria, "Framework of Knowledge-Based System for United Nations Peacekeeping Operations Using Data Mining Technique," in *Proc. 4th Int. Conf. Information Retrieval Knowl. Man., (CAMP)*, pp. 18-22, 2018.
- [32] A. S. H. Lee, Z. Yusoff, Z. Zainol, and V. Pillai, "Know your hotels well! -- An Online Review Analysis using Text Analytics," *Int. J. Eng. & Tech., (IJET)*, vol. 7, no. 4.31, pp. 341-347, 2018.
- [33] K. Berezina, A. Bilgihan, C. Cobanoglu, and F. Okumus, "Understanding satisfied and dissatisfied hotel customers: text mining of online hotel reviews," *J. Hospitality Marketing & Man.*, vol. 25, no. 1, pp. 1-24, 2016.
- [34] N. F. Ibrahim, X. Wang, and H. Bourne, "Exploring the effect of user engagement in online brand communities: Evidence from Twitter," *Comp. in Human Behavior*, vol. 72, 2017.
- [35] P. N. E. Nohuddin, Z. Zainol, F. C. Chao, M. T. James, and A. Nordin, "Keyword based Clustering Technique for Collections of Hadith Chapters," *Int. J. Islamic Appl. Comp. Sci. Tech., (IJASAT)*, vol. 4, no. 3, pp. 11-18, 2015.
- [36] W. M. U. Noormanshah, P. N. E. Nohuddin, and Z. Zainol, "Document Categorization Using Decision Tree: Preliminary Study," *Int. J. Eng. & Tech., (IJET)*, vol. 7, no. 4.34, pp. 437-440, 2018.
- [37] Z. Zainol, M. T. H. Jaymes, and P. N. E. Nohuddin, "VisualUrText: A Text Analytics Tool for Unstructured Textual Data," *J. Phys.: Conf. Ser.*, vol. 1018, no. 1, p. 012011, 2018.
- [38] Z. Zainol, S. Wani, P. N. E. Nohuddin, W. M. U. Noormanshah, and S. Marzukhi, "Association Analysis of Cyberbullying on Social Media using Apriori Algorithm," *Int. J. Eng. & Tech., (IJET)*, vol. 7, no. 4.29, pp. 72-75, 2018.
- [39] D. Mouheb, M. H. Abushamleh, M. H. Abushamleh, Z. Al Aghbari, and I. Kamel, "Real-Time Detection of Cyberbullying in Arabic Twitter Streams," in *Proc. 10th IFIP Int. Conf. New Tech. Mobility Security (NTMS)*, 2019, IEEE, pp. 1-5.
- [40] H. Mohamed, S. Marzukhi, Z. Zainol, T. M. T. Sembok, and O. Zakaria, "Semantic-based Social Media Threats Detection". *Proc. 12th Int. Conf. Ubiquitous Information Man. and Comm., (IMCOM 18)*, , 2018.
- [41] M. W. Berry and J. Kogan, "Text Mining. Applications and Theory," Wiley, 2010.
- [42] S. Chua and P. Nohuddin, "Relationship Analysis of Keyword and Chapter in Malay-Translated Tafseer of Al-Quran," *J. Telecommunication, Electronic and Comp. Eng. (JTEC)*, vol. 9, no. 2-10, pp. 185-189, 2017.
- [43] F. Zhu et al., "Biomedical text mining and its applications in cancer research," *J. Biomed. Informa.*, vol. 46, no. 2, pp. 200-211, 2013.