Research Article

Development of data-to-text (D2T) on generic data using fuzzy sets

Lala Septem Riza[1*], Muhammad Ridwan[1], Enjun Junaeti[1] and Khyrina Airin Fariza Abu Samah[2]

Department of Computer Science Education, Universitas Pendidikan Indonesia, Bandung, Indonesia[1]
Universiti Teknologi Mara Cawangan Melaka Kampus Jasin, Malaysia[2]

Abstract
Data-to-Text (D2T) is an option for translating non-linguistic data into textual form. However, along with technological developments, the various fields of data and the variety of users are one of the focuses that must be considered in the development of D2T. This study aims to develop a D2T system with input in the form of general data so that it can receive data from any field or domain, whether the data have header information, data types, rules or not. Then fuzzy rule based systems are used to interpret data in general. The system developed can produce information in the form of data summaries, newest data information, and predictive information. It is carried out in the R programming language by utilizing several available packages. Experiments are carried out by measuring the level of readability of the news generated, computation time, and comparing the results with related research. The experimental results show that the information generated is proven to represent the data provided and can be understood by the level of students even at the elementary school level, and the computation time is quite good.

Keywords
Data-to-text, Natural language generation, Machine learning, General purpose, General corpora, Fuzzy rule based system, Time-series analysis, Linear regression, Knuth-morris-pratt.

1.Introduction
Along with current technological developments, the availability of information is increasing, especially information in the form of non-linguistic data or numerical data. Non-linguistic data can be presented in text or linguistic form to make it easier to retrieve information [1]. So the D2T system is developed with the ability to produce information in textual form with input in the form of non-linguistic data or numeric data [2]. The data used can be obtained from various data sources such as recordings of a sensor, event logs, and other data sources that are recorded periodically. D2T is part of the Natural Language Generation (NLG) system where D2T translates data into text by assuming that the data used is fundamentally true and accurate [3]. An NLG system is divided into at least four main parts, namely (macro planning, micro planning, linguistic realization, and presentation), where each section has its own sub-sections, such as in macro planning there are subsections of content planning, text planning, and Rhetorical Structure Theory (RST) and in microplanning, there is lexicalization [4].

D2T architecture is almost similar to NLG which is divided into four main stages, i.e., signal analysis, data interpretation, document planning, microplanning and realization [2]. It is a solution that can be used to translate non-linguistic data to the public without eliminating the meaning contained in the data, with the aim of making information in the form of text that is presented easier to understand compared to non-linguistic data. There are at least two main problems that must be considered in the construction of a corpus. First, the various types of information stored in a corpus, so that the resource aspect is an important factor and must be considered. Second, the various types of users, so that the corpus that is built must be able to cover the various needs of each user [5].

Various studies have been carried out to build a D2T system for each specific area, for example, in the climatology field there is a Forecast Generator ("FoG") system that can convert weather maps into forecasts in sentence form with natural language processing [6], then there is D2T Weather Prediction (DWP) system which can produce climatological news summaries and weather for one month and

---

*Author for correspondence

provides prediction information for the next day [7], besides that there is "SumTime-Mousam", this application can generate textual oceanic weather forecasts for offshore oil rigs [8]. In the health sector, the "BabyTalk" system was introduced with the intent to produce text summaries from 45 minutes of neonatal data which would later be used as supporting material for the decision to present the modalities that occurred at that time [9], the "BT-Nurse" system summarizes events during the nursing shift, based on the results of the patient's electronic medical record [10]. In the economic field, "Knowledge-Based Report Generator" can produce stock reports based on product stock data (non-linguistic) of a market [11].

In this study, we are aimed to develop an application of D2T to generate news based on general data with a period of hours, daily, weekly, monthly, or even yearly but restricted only to exact data and time series. The General data in question is data that is not bound to any domain, whether the data has an identity in the form of header information, category or not, so that the corpus and the resulting output are as general as possible. In this research, a D2T system will be built that is not tied to any field and can receive input in the form of general data and then produce the most general output possible so that the system built is expected to be a solution to the problems previously mentioned. To achieve this, the development of the D2T system uses machine learning methods [12], such as Linear Regression [13, 14], Knuth-Morris-Pratt (KMP) [15, 16], Pearson correlation, and statistical tools for data processing, analysis, and prediction [17].

2.Methods
Basically, the D2T Model is divided into four main parts, namely signal analysis, data interpretation, document planning, microplanning and realization [2]. In this research, the development of a D2T model for general data was carried out, so that it focused on modifications of signal analysis and data interpretation as illustrated in "*Figure 1*". This model describes a summary of the stages of system development.

2.1General data handler
In this process, the dataset is checked, whether there is a header in the dataset, otherwise the header name becomes the default (e.g., "DateTime", "v2", "v3", "v4", etc.). Then, the next process is parameters consisting of names, types (numerical/categorical), rules (crisp/fuzzy), and alternate which will function to replace parameter names on the final display of the text as shown in *Table 1*.

Tabel 1 Configurations of parameters

	ColName	Type	Rule	Alternate
1	CloudCoverage	Numeric	crisp	Cloud Coverage
2	Temperature	Numeric	fuzzy	NA
3	WindSpeed	Numeric	crisp	Wind Speed
4	WindDirection	Numeric	crisp	Wind Direction
5	Rainfall	Numeric	fuzzy	NA

2.2Signal analysis
Basically, this process is utilized to transform numerical data into discrete patterns. In this research, statistical tools such as Min, Max, Mean, and Machine Learning are applied. Linear Regression is used to determine the trend of a parameter [13, 14, 18], KMP is used to look for the same data pattern as this week's data pattern [19], while time series analysis and exponential smoothing are used in determining predictions for further data [20]. This process is carried out to look for discrete patterns of data, which will be further processed at a later stage, for more details described in *Table 2*.

In addition, in this process there is an Extreme Event tracing process, first looking for the difference between the data on index i and i+i, then adding up the difference according to whether the difference is negative (decreasing) or positive (increasing), then the highest sum of differences between a negative and positive number, stored in the form of a list, the information stored are namely: the number of increases/decreases, the initial and final indices of the increase/decrease, and the duration of the increase/decrease.

Whereas data is categorized as a repeated event if the number of rows of data of the same value consecutively exceeds n rows of data, the user can determine the value of n by changing the value. However, if the user does not define it, the default value is used, namely: for data with a period of hours, the minimum limit used is 6 hours. For data with a daily timeframe, the minimum limit used is one week

or 7 days. For data with a monthly period, the minimum limit is used, namely one quarter or the equivalent of 4 months, and for data with an annual period, the minimum limit is used for one quarter or 3 months.

All of these processes are applied to numerical data, for categorical data a special process is applied,

namely motive discovery, where the system will look for whether there is a data pattern that is the same as the data pattern number (*n-i*) until n, where n is the index of data for today, while i is the parameter limit such as weekly, quarterly, quarter as used in the repeated event process.
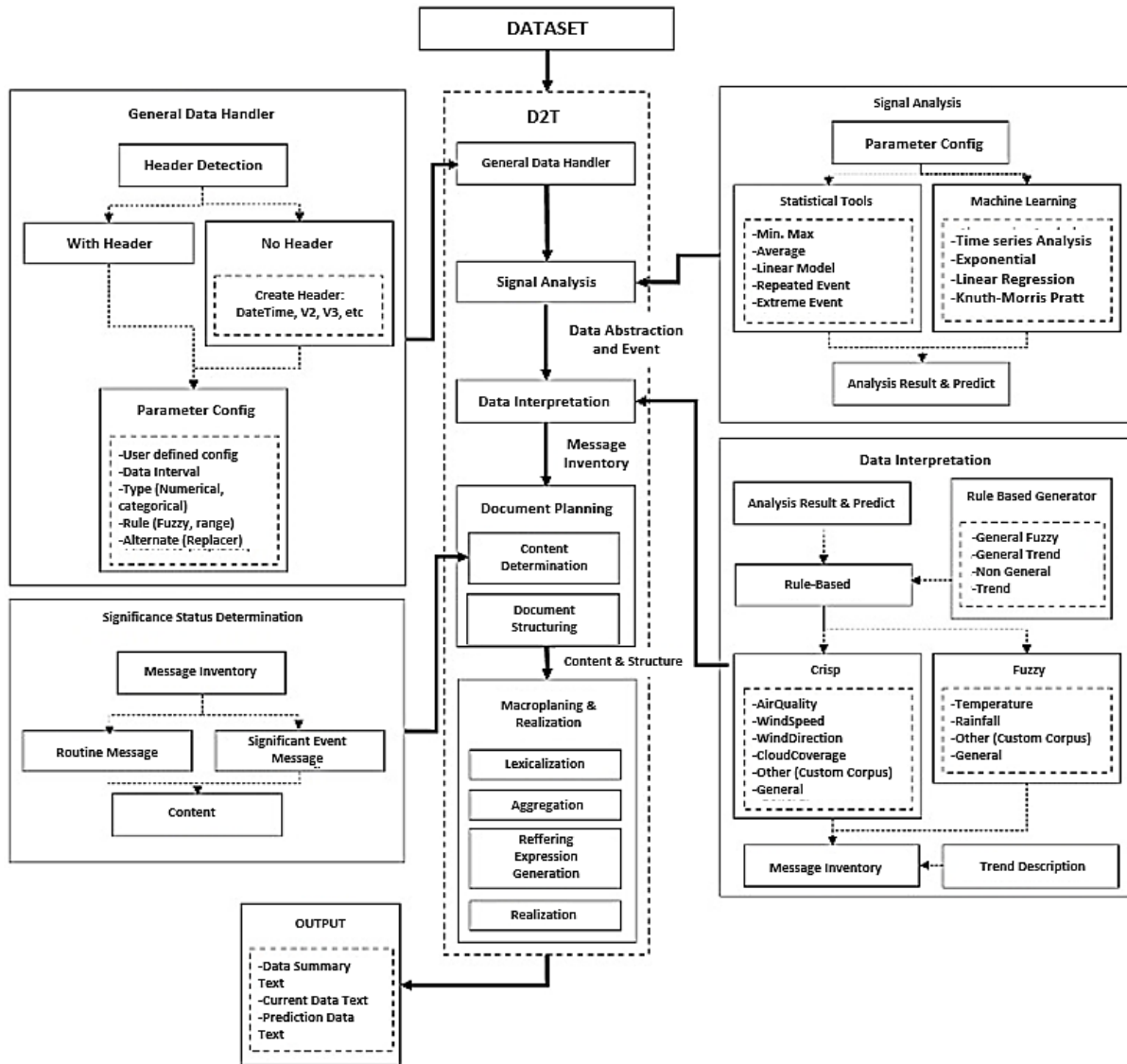
Figure 1 The proposed D2T Model for general data in this research

Table 2 Model signal analysis results

	ColName	MaxDate	MaxValue	MaxIndex	MinDate	MinValue	MinIndex	SumValue	Average	Trend
1	V2	09/01/2015 00:00	14.65	177	05/01/2002 00:00	8.28	17	2174.52	10.35	+
2	V3	06/01/2018 00:00	13.03	210	06/01/2002 00:00	6.91	18	2086.52	9.93	+
3	V4	09/01/2015 00:00	22.20	177	05/01/2001 00:00	12.87	5	3517.32	16.74	+
4	V5	09/01/2015 00:00	15.08	177	05/01/2001 00:00	5.62	5	2040.77	9.71	+
5	V6	06/01/2018 00:00	10.52	210	06/01/2002 00:00	4.71	18	1512.70	7.20	+
6	V7	05/01/2009 00:00	3.94	101	05/01/2010 00:00	1.17	113	603.21	2.87	+
7	V8	11/01/2016 00:00	3.03	191	05/01/2002 00:00	1.06	17	280.79	1.33	+
8	V9	06/01/2014 00:00	11.26	162	09/01/2001 00:00	4.89	9	1759.47	8.37	+
9	V10	12/01/2013 00:00	11.44	156	12/01/2001 00:00	5.67	12	1858.65	8.85	+

2.3 Data interpretation

It is focused on mapping the discrete patterns and events produced by signal analysis into messages and relationships defined by developers. Because the news that will be generated is in the form of general news, this stage is divided into two main processes, namely, rule-based interpretation using the crisp set and interpretation using the fuzzy rule based systems. However, for general data that doesn't have a rule as a reference for interpretation, there is a special process called general fuzzy generator and general trend generator, where data of general type will use the rule-based concept by using fuzzy sets. The general data will be interpreted based on the corpus with a membership value which is a modification of the fuzzy membership function on the general trend generator as can be seen in *Figure 2*, where the membership value for each parameter is highly dependent on the value range. The maximum and minimum values are obtained from the signal analysis stage [21].

In addition to general data, there are data interpretations that have been defined in Config Parameters, including air quality [22], wind speed [23], wind direction [24], and cloud coverage [25] using the crisp membership function while for temperature [26], rainfall [27], and parameters that have not been identified (general), data interpretation will be carried out using the fuzzy membership function. In addition, users can customize this process by changing the value or adding new parameters to be interpreted.
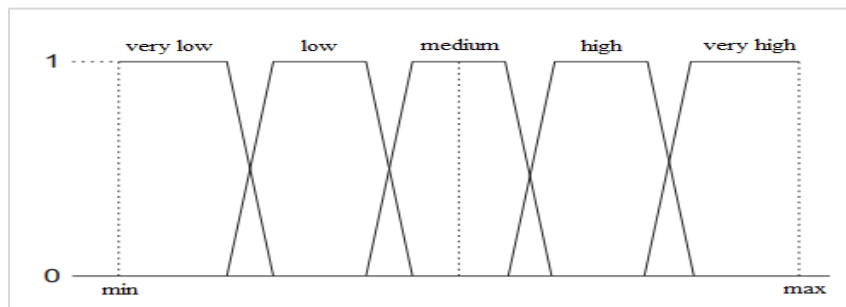

Figure 2 Fuzzy membership function for general parameter

2.4 Document planning

It is a process to determine whether events will be mentioned in the text along with the document structure. This process involves selecting content (content determination) and forming a text structure (document structuring) [28]. For the content selection process, it is done by dividing the content into two groups, namely routine message and significant event message [7], while document structuring is done by creating a scheme based on the target text created [7].

385

At this stage, the user can customize what will be displayed in the text. For summary content and current data description, content is selected by grouping the two previous groups, whereas for predictive information only uses the routine message group.

2.5 Microplanning and realization

It is a step to generate natural language in text based on the previous steps. At this stage, there are at least four things that need to be done, namely, lexicalization, aggregation, referring expression generation, and structure realization. In the lexicalization stage, a representation process is carried out between data changes, for example, "turn progressively to", "decrease to", "keep stable at", and so on [6]. The aggregation stage is carried out when connecting several messages into one unit using simple conjunction referring to contrast value [29]. Referring expression generation is done by generating randomly based on the corpus made [6].

Implementation of structure realization is carried out by arranging all content into a predetermined structure [2], the next process is to replace the parameter names with alternate values found in the config parameters.

3. Results

Experiments were carried out by generating news with 12 test-cases as in *Table 3*. For example, the raw data of CL_NHM as illustrated in *Table 4*, it can be seen that we have six attributes (i.e., date and time, cloud coverage, temperature, wind speed, wind direction, and rainfall) in numeric. By running the system and inputting these data, we will obtain the paragraphs in text illustrating the data behaviors.

Each experimental result was measured on four aspects, namely, reliability and computation time [26−29], as well as the comparison of outputs. General systems are compared with systems in a specific field, also a validation of representative text. The reliability test is performed using the readability analyzer application and for the computation time it is done using the system.time() function in R, while representative text validation is done by comparing the information with data visualization.

After doing some experiments using datasets as presented in *Table 3*, we obtain results as paragraphs. For example, the paragraphs generated by the D2T system for the CL_NHM data in *Table 4* is presented in *Figure 3*.

Table 3 Data collection: dataset code and their sources for experiments

Dataset code	Dataset	Source and detail
KB_WH	(Buy) March 2018	Bank Indonesia website (https://www.bi.go.id/) rupiah exchange rate against foreign currencies with a monthly range, from 1 January 2020 to 1 June 2018, using headers
KB_NH	(Buy) April 2018	Bank Indonesia website (https://www.bi.go.id/) rupiah exchange rate against foreign currencies with a monthly range, from 1 January 2020 to 1 June 2018, without headers
KB_NHM	(Buy) Mei 2018	Bank Indonesia website (https://www.bi.go.id/) Rupiah exchange rate against foreign currencies with a monthly range, from 1 January 2002 to 1 June 2018 with Corpus customization, without using headers
SD_WH	July 2016	Daily interval data from the website https://solargis.com. For one year starting from 01 July 2016 to 31 December 2016 using a header
SD_NH	August 2016	Daily interval data from the website https://solargis.com. For one year starting from 01 August 2016 to 31 December 2016 without using a header
SD_NHM	September 2016	Daily interval data from the website https://solargis.com. For one year from 01 September 2016 to 30 September 2016 with Corpus customization, without using headers
CL_WH	1 January 2016 - 31 December 2017	Climatological data on the website www.MeteoGalicia.gal, for one year in the 2016-2017 period using headers
CL_NH	1 January 2016 - 31 December 2017	Climatological data on the website www.MeteoGalicia.gal, for one year in the 2016-2017 period without using a header
CL_NHM	1 January 2016 - 31 December 2017	Climatological data on the website www.MeteoGalicia.gal, for one year in the 2016-2017 period with Corpus customization, without using headers

Dataset code	Dataset	Source and detail
AQ_WH	2016-2017	Air quality data on the website www.MeteoGalicia.gal, for one year in the 2016-2017 period using headers
AQ_NH	2016-2017	Air quality data on the website www.MeteoGalicia.gal, for one year in the 2016-2017 period without using a header
AQ_NHM	2016-2017	Air quality data on the website www.MeteoGalicia.gal, for one year in the 2016-2017 period with Corpus customization, without using headers

Table 4 The raw data of CL_NHM (Climatological data)

DateTime	CloudCoverage	Temperature	WindSpeed	WindDirection	Rainfall
07/06/2016 00:00	40.8	21.3	5.47	315	0
07/07/2016 00:00	20.9	20.1	6.41	315	0
07/08/2016 00:00	27.2	19.5	7.02	315	0
07/09/2016 00:00	23.2	19.1	5.94	315	0
07/10/2016 00:00	77.5	18.7	5.44	180	0
...	...	...	...	...	...
07/02/2017 00:00	12.6	18.9	7.34	315	0
07/03/2017 00:00	13.3	21.8	4.86	315	0
07/04/2017 00:00	18.7	24	6.59	225	0
07/05/2017 00:00	81.1	19.2	8.35	225	0
07/06/2017 00:00	58.3	17.9	6.48	315	0

Climatology NEWS

Regarding to the daily data, between 07/06/2016 00:00 to 07/06/2017 00:00, with parameters: Cloud Coverage, Temperature, Wind Speed, Wind Direction, and Rainfall. It can be seen that, Temperature trend is decreased and Rainfall trend is constant but the rest is increased. Temperature parameter is higher than last week's data, but the rest parameters are lower than last week's data. There were some repeating value more than 36 days: Wind Direction stayed constant at no rain during 6 Jul - 18 Aug 2016. Cloud Coverage fluctuated rapidly (increased 90.9 points and decreased 87.6 points), Wind Speed fluctuated significantly (increased 20.05 points and decreased 27.03 points), and Rainfall fluctuated significantly (increased 67.2 points and decreased 64.1 points). For the past 7 days , no equivalent patterns were found for each categorical parameters. Rainfall appears to have a highest impact to all variable with moderate relationship in average.

Today data show that: Cloud Coverage in mostly cloudy condition. Temperature in warm condition. Wind Speed in light Breeze condition. Rainfall in no rain condition. Rainfall reached their lowest value on this day.

Based on prediction result, it's expected that tomorrow sky will be light rain although it's covered by partly cloudy sky. Followed by temperature which decreased to warm. Cloud Coverage will shifted progressively to partly cloudy. Temperature will steady at warm. Wind Speed will steady at light Breeze. Rainfall will normaly move to light rain.

Figure 3 Texts generated by the D2T system in the 6th experiment using climatological data with corpus customization (CL_NHM)

4.Discussion
Overall, based on the result, the amount of content is certainly increasing, but textually this application is not as good as the DWP output [7] in his explanation, because the concept of this application is built for general data so that it can generate news based on any data as long as the data follows the input data format, while in DWP research input data must be the same as that in the study (parameters). In the DWP study, there are two input data, namely climatology and air quality, so that the content that appears has two parts. Whereas in research conducted by Ramos et al. [27, 28, 29] , the text was built only for air quality and wind speed. In Goldberg's research [6], there is a message about wind and snow separately whereas in the Gkatzia study it is only communicated

regarding the state of the sky [30]. This shows that the D2T that has been built before is only for specific data, not general, so there is a possibility that the previous system will not work if given other time-series data. Moreover, in the aspect of reliability, an assessment is carried out based on the Flesch Reading Ease Score obtained using the ReadibilityAnalyzer tool on the www.datayze.com website and the www.Readabilityformulas.com site, so that the results are in *Table 5*. The results of the Computation Time are obtained by running the system.time() function in R language, so that the results are as in *Table 6*. It can be seen that in general speaking the computation times of the developed D2T system is quite fast.

Table 5 Readability aspect measurement results

Dataset code	Flesch reading ease score (Readability formulas)	Flesch reading ease score (Datayze)
CL_WH	79.3	67.65
CL_NH	90.3	82.27
CL_WHM	87.5	79.58
KB_WH	60.3	50.6
KB_NH	89.5	84.38
KB_NHM	91.5	86.77
SD_WH	82.1	78.1
SD_NH	83.7	74.3
SD_NHM	74.9	90.2
AQ_WH	87.2	81.5
AQ_NH	81.3	86.6
AQ_NHM	72.4	75.7
Average	81,7	78,1
Overall Average	80.0	

Table 6 Computation time measurement results

Dataset code	Running time (s)
CL_WH	2.01
CL_NH	1.98
CL_WHM	2.14
KB_WH	1.84
KB_NH	2.08
KB_NHM	2.08
SD_WH	1.98
SD_NH	2.03
SD_NHM	2.25
AQ_WH	2.44
AQ_NH	2.63
AQ_NHM	2.79
Average	2.235

5.Conclusion and future work

The development of a D2T system for general data using machine learning and other approaches (i.e., KMP and statistical tools) has been done, where it can work without input in the form of numeric data tables of any kind. This is one of the advantages, because any input system whether the data has information in the form of headers, data types, rules, or not, the system will still be usable. From the overall results of the experiments conducted, the output of the system is proven to represent the given data. This study obtained an overall average score of 80.0 in the aspect of reliability, which means that the output of this system is classified as a reading category that is very easy to understand even by the level of elementary school students. Meanwhile, in the aspect of Computation Time, the average computation time is 2,235 seconds. For further research, the development of corpus can be carried out so that the type of output produced becomes more varied, especially in the content determination section, it is hoped that the machine learning algorithm can be used for content selection in the output text. Moreover, UI / UX will be improved so that users can more freely customize the application.

Acknowledgment
None.

Conflicts of interest
The authors have no conflicts of interest to declare.

References
[1] Gerstl P. Linking linguistic and non-linguistic information. Data & knowledge engineering. 1992; 8(3):205-22.
[2] Reiter E. An architecture for data-to-text systems. In proceedings of the eleventh European workshop on natural language generation (ENLG 07) 2007 (pp. 97-104).
[3] Gkatzia D, Lemon O, Rieser V. Data-to-text generation improves decision-making under uncertainty. IEEE Computational Intelligence Magazine. 2017; 12(3):10-7.

[4] McDonald DD. Natural language generation. Handbook of Natural Language Processing. 2010; 2:121-44.

[5] Soehn JP, Zinsmeister H, Rehm G. Requirements of a user-friendly, general-purpose corpus query interface. Proceedings of the LREC Workshop Sustainability of Language Resources and Tools for Natural Language Processing. 2008 (pp. 27-32).

[6] Goldberg E, Driedger N, Kittredge RI. Using natural-language processing to produce weather forecasts. IEEE Expert. 1994; 9(2):45-53.

[7] Riza LS, Putra B, Wihardi YA, Paramita B. Data to text for generating information of weather and air quality in the R programming language. Journal of Engineering Science and Technology. 2019; 14(1):498-508.

[8] Reiter E, Sripada SG, Robertson R. Acquiring correct knowledge for natural language generation. Journal of Artificial Intelligence Research. 2003; 18:491-516.

[9] Portet F, Reiter E, Gatt A, Hunter J, Sripada S, Freer Y, Sykes C. Automatic generation of textual summaries from neonatal intensive care data. Artificial Intelligence. 2009; 173(7-8):789-816.

[10] Hunter J, Freer Y, Gatt A, Reiter E, Sripada S, Sykes C, Westwater D. BT-Nurse: computer generation of natural language shift summaries from complex heterogeneous medical data. Journal of the American Medical Informatics Association. 2011; 18(5):621-4.

[11] Kukich K. Design of a knowledge-based report generator. In meeting of the association for computational linguistics 1983 (pp. 145-50).

[12] Carbonell JG, Michalski RS, Mitchell TM. An overview of machine learning. Machine Learning. 1983; 1:3-23.

[13] Riza LS, Handian D, Megasari R, Abdullah AG, Nandiyanto AB, Nazir S. Development of R package and experimental analysis on prediction of the CO2 compressibility factor using gradient descent. Journal of Engineering Science and Technology. 2018; 13(8):2342-51.

[14] Riza LS, Nasrulloh IF, Junaeti E, Zain R, Nandiyanto AB. gradDescentR: An R package implementing gradient descent and its variants for regression tasks. In international conference on information technology, information systems and electrical engineering 2016 (pp. 125-9). IEEE.

[15] Riza LS, Rachmat AB, Munir TH, Nazir S. Genomic repeat detection using the knuth-morris-pratt algorithm on r high-performance-computing package. International Journal of Advances in Soft Computing and its Applications. 2019; 11(1):94-111.

[16] Riza LS, Firmansyah MI, Siregar H, Budiana D, Rosales-Pérez A. Determining strategies on playing badminton using the Knuth-Morris-Pratt algorithm. TELKOMNIKA Telecommunication Computing Electronics and Control. 2018; 16(6):2763-70.

[17] Riza LS, Anwar FS, Rahman EF, Abdullah CU, Nazir S. Natural language processing and levenshtein distance for generating error identification typed questions on TOEFL. Journal of Computers for Society. 2020; 1(1):1-23.

[18] Atilgan A, Tanriverdi C, Yucel A, Oz H, Degirmenci H. Analysis of long-term temperature data using Mann–Kendall trend test and linear regression methods: the case of the southeastern Anatolia region. Scientific Papers Series a Agronomy LX. 2017:455-62.

[19] Régnier M. Knuth-Morris-Pratt algorithm: an analysis. In international symposium on mathematical foundations of computer science 1989 (pp. 431-44). Springer, Berlin, Heidelberg.

[20] Ostertagova E, Ostertag O. Forecasting using simple exponential smoothing method. Acta Electrotechnica et Informatica. 2012; 12(3):62-6.

[21] Castillo-Ortega R, Marín N, Martinez-Cruz C, Sánchez D. A proposal for the hierarchical segmentation of time series. application to trend-based linguistic description. In IEEE international conference on fuzzy systems (fuzz-IEEE) 2014 (pp. 489-96). IEEE.

[22] Vallero DA. Fundamentals of air pollution. Academic Press; 2014.

[23] https://www.unc.edu/~rowlett/units/scales/beaufort.html. Accessed 20 May 2018.

[24] http://snowfence.umn.edu/Components/winddirectionanddegreeswithouttable3.htm. Accessed 20 May 2018.

[25] http://www.theweatherprediction.com/habyhints/189/. Accessed 20 May 2018.

[26] Belz A. Probabilistic generation of weather forecast texts. In human language technologies 2007: the conference of the north american chapter of the association for computational linguistics; proceedings of the main conference 2007 (pp. 164-71).

[27] Ramos-Soto A, Bugarín A, Barro S. Fuzzy sets across the natural language generation pipeline. Progress in Artificial Intelligence. 2016; 5(4):261-76.

[28] Reiter E, Dale R. Building applied natural language generation systems. Natural Language Engineering. 1997; 3(1):57-87.

[29] Ramos-Soto A, Bugarín A, Barro S. On the role of linguistic descriptions of data in the building of natural language generation systems. Fuzzy Sets and Systems. 2016; 285:31-51.

[30] Gkatzia D, Lemon O, Rieser V. Natural language generation enhances human decision-making with uncertain information. arXiv preprint arXiv:1606.03254. 2016.

Lala Septem Riza received PhD in Computer Science from Universidad de Granada, Spain in 2015. He is currently working in the Departement of Computer Science Education, Universitas Pendidikan Indonesia, Indonesia. He is teaching Machine Learning, Big Data Platform, and Statistical Data Science. His research interests are in Machine Learning, Data Science, and Educations.
Email: lala.s.riza@upi.edu

Muhammad Ridwan, received undergraduate from the study program of computer science in Department of Computer Science Education, Universitas Pendidikan Indonesia. During his study, he is active in many student activites, and he is an assistant in a laboratory in the department. The final project is about Data to Text and Natural Language Generation.
Email: just.muhammadridwan@student.upi.edu

Enjun Junaeti was born in Majalengka City, West Java Province, Indonesia on December 20, 1985. She completed her bachelor's degree at Universitas Pendidikan Indonesia in 2008 and obtain a master's degree in Mathematics at Bandung Institute of Technology in 2011. She is currently a lecturer in the Department of Computer Science Education, University of Indonesian Education. She is currently being worked on the simulation and modelling of the student's behaviours and teacher's actions in the learning process based on Artificial Intelligence, Machine Learning, and Big Five Personality.
Email: enjun@upi.edu

Khyrina Airin Fariza Abu Samah. She is a senior lecturer from Universiti Teknologi MARA (UiTM), Melaka Jasin Campus. She has 13 years of working experience in the Information Technology field in the semiconductor industry before joining UiTM. She has Diploma, Bachelor's Degree and Master's Degree in Computer Science and PhD in Information Technology. Her research interest in Artificial Intelligent, Operational Research, Algorithm Analysis, Clustering and Optimization, Evacuation Algorithm, Internet of Things (IoT) and Sentiment Analysis.
Email: khyrina783@uitm.edu.my