Template-Based Natural Language Generation in Interpreting Laboratory Blood Test

Opim Salim Sitompul, Erna Budhiarti Nababan, Dedy Arisandi, Indra Aulia, Hengky Wijaya

Abstract—**Results of blood test examination from medical test institutions are represented in medical abbreviation term-value pairs. The medical abbreviation terms are components of the blood examination being tested while the values are results of laboratory examination obtained. For experienced doctors, these terms and values are very familiar, by which they can easily determine whether those values are in normal conditions or indicating abnormalities. Those experienced doctors have in their mind what is a normal range for a given blood component being examined. However, this is not the case with young doctors in general. They should have some kinds of lookup table to compare each component value obtain from laboratory test with the corresponding acceptable (or normal) range. Furthermore, the abundance of laboratory test results generated every day will also accumulated to add more burdens for the young doctors. In this research, we propose a natural language generation (NLG) approach to help young doctors in interpreting results of blood test examination obtained from a laboratory. The NLG system is capable of giving some interpretation on the laboratory results and highlight importance indications for young doctors. Responses from young doctors and general practitioners, based on an assessment evaluated for readability, clarity, and general appropriateness of the textual interpretation show a promising result where the average percentage of naturalness measurement obtained was $\geq 90\%$.**

Index Terms—**natural language generation, blood test, hematology examination, young doctors, textual representation**

I. Introduction

ABILITY to explain results of a laboratory examination comes from day-to-day experiences as an outcome of accumulating knowledge and performing repeated tasks. The explanation is actually composed from interpretations of reference tables that usually accessible by medical workers. In general, the laboratory examination results are put in the form of term-value pairs. In the case of hematological examination for complete blood count (CBS) tests, for example, these term-value pairs are represented as Hgb-5.8 g/dl, Leu-1.5 $10^9/L$, etc. In a clinical laboratory, nurse or laboratory assistant is expected to summarized those term-value pairs in a tabular form to be easily read by doctors, as well as patients.

Manuscript received June 12, 2020; revised November 26, 2020. This research was funded by Kemenristekdikti Republik Indonesia through Lembaga Penelitian Universitas Sumatera Utara. The support was under the research grant DRPM Kemenristekdikti of Year 2019 under Contract Number: 225/UN5.2.3.1/PPM/KP-DRPM/2019.

Opim Salim Sitompul is with Department of Information Technology, Universitas Sumatera Utara, Medan, 20155, Indonesia, phone/fax: +6261822139; +6261822129; (e-mail: opim@usu.ac.id).

Erna Budhiarti Nababan is with Department of Information Technology, Universitas Sumatera Utara, Medan, 20155, Indonesia, (e-mail: ernabrn@usu.ac.id).

Dedy Arisandi is with Department of Information Technology, Universitas Sumatera Utara, Medan, 20155, Indonesia, (e-mail: dedyarisandi@usu.ac.id).

Indra Aulia is with Department of Information Technology, Universitas Sumatera Utara, Medan, 20155, Indonesia, (e-mail: indraaulia@usu.ac.id).

Hengky Wijaya was formerly a student at Department of Information Technology, Universitas Sumatera Utara, Medan, 20155, Indonesia, (e-mail: hengkywijaya1313@gmail.com).

However, results of the hematological examinations are given in tabular forms as they are with no explanation on whether the blood components are within normal, abnormal or critical values. To find out whether a blood component is normal, abnormal or critical, a physician should manually compare each value obtained with a list of normal range of values available, one after another individually. This manual procedure will consume much time, not to mention it is an error-prone task. Therefore, research works focusing on developing systems to interpret data into information in the form of textual representations, or conversion of numeric data into summary texts reports are becoming a necessity. The research works are ranging from natural language processing (NLP) to natural language generation (NLG).

Current reviews on clinical NLP systems were focused on generating structured information from unstructured free text, whereby many of the NLP systems are capable of processing clinical free text and generating structured output [1]. In order to model attributes and features, the clinical NLP systems usually are developed and evaluated according to annotations of word, sentence, or document level. Those annotations come from various types of document contents, document section types, named entities and concepts, and semantic attributes, which consists of patient status, report type, current medications, past medical history, discharge summary, diagnoses, symptoms, treatments as well as negation, severity, and temporality [2].

Likewise, concepts of generating textual representations using NLG approach has been widely described in [3]. Basically, numbers data are transformed into meaningful information in order to be read and understand by non-expert users, as well as users who have limited time in reading and understanding the whole data. However, a major drawback of those types of end-to-end (E2E) NLG approaches are that they were limited to small, delexicalized datasets [4]. This condition draws further research works into larger instances from new crowdsourced datasets in order to produce more complex outputs.

In turn, the NLG approach or narrative science in general is then oriented into health-related domain. Numerous research works on this topic had been conducted (*see* for example works by [5], [6], [7], [8]). In essence, the main purpose of those works is to educate common people in understanding data and to gain efficient access to information, thereby allowing them to make interpretation in order to take effective decisions [9].

In this research, we propose a system that could educate young physicians and medical workers to understand data obtained from laboratory examination of hematology patient by accompanying some interpretation into the generated summary text. The interpretation is concerning on whether the blood components are in normal, abnormal or critical condition. A template-based NLG approach is proposed to generate the summary text reports based on the results of patients hematological examination.

II. NLG for Textual Representation of Hematology Examination Result

Natural language generation is one of the natural language processing tasks to generate ordinary language from a machine representation system such as knowledge base or formal logic systems. It is like a translator that converts data into a natural language representation. In this term, NLG is one area of research that deals with automatic production of human-readable text suitable for certain applications [10].

To support natural language generation system efficiently, research works focused on two orthogonal methods: surface realization components and template-based techniques. The first approach uses reusable, general, and linguistic motivation, whereas the later uses simple and task-oriented approach [11].

Currently, NLG approach is commonly used in medical field to improve services as well as patient's safety for hospital pre-treatment [12].

A. Template-Based Natural Language Generation

A natural language generation system in which nonlinguistic inputs could be mapped directly without intermediate representation to the structure of its linguistic surface is known as template-based system. In this system, the whole phrases are extended from their lexical components in the form of Tree Adjoining Grammar (TAG) [13], [14].

As further described in [13], a template in the form of linguistic structure contains slots of relevant information that need to be filled in order to obtain well-formed output. Relevant information to be inserted into the slots are stored in a relevant information table.

B. Hematology Laboratory Examination

The results of laboratory tests can be expressed in three forms: quantitative (e.g. normal haemoglobin values of women are 12–16 g/dl), qualitative (e.g. positive or negative), and semiquantitative (qualitative results with positive or negative number: 1+, 2+, etc.). Complete blood count examinations conducted in most clinical laboratories use automated machines that include red cell indices as part of the profile [15].

Hematologic examination (hemogram) consists of examination of leukocytes, erythrocytes, hemoglobin, hematocrit, erythrocytes and platelets. A complete blood count consists of hemogram accompanying with differential leukocyte checks consisting of neutrophils, basophils, eosinophils, lymphocytes and monocytes [16].

III. Related Research

Research by [5] focused on effective approach of creating a system that generates a text of medical information reports for parents of premature babies. They analyze the signal and interpret electronic medical records (EMRs) data to identify the important events and the relationship between the events occurring in the EMR data. The NLG method is then deployed to convert the EMR data into a narrative text. Their research focused on the text produced by the system that could be understood by people who are not professionals in the medical field and the resulting report text only gives positive information about infant development.

Using time-series data obtained from continuous physiological signals and structured information about events from medical staffs, [9] developed a BT-45 prototype system to generate textual summaries on equipment settings and drug administration for an average of 45 minutes. The BT-45 system was developed based on the work from [17] which used raw data instead of AI knowledge base.

In [18] a natural language generation system called Report Generator was developed to produce summary reports of patient records. Those records are obtained from repository of patient medical records of cancer patients in the forms of narrative documents, such as letters, discharge reports, etc., as well as structured data such as test results, prescriptions, etc. The system design of the Report Generator follows a classical NLG pipeline architecture, which consists of a Content Selector, MicroPlanner and Syntactic Realiser.

Research on the creation of a system that could produce summary text from physician briefed notes and nurse structural documentation containing patient care plans for heart disease inpatient had been performed in [6]. This summary text is useful to educate patients in taking care of their health after hospitalization. It is also considered as an approach to educate them on what treatments are being performed during their inpatient period. The preparation of summary text begins with a process of building a graph to see the relationship between two input data (short note data written by doctors and nurses' structural documentation data). Subsequently, the selected information extracted from the graph was written into a summary text. The last process undertaken was the application of SimpleNLG system.

A similar research in creating a system for generating summary text of hospital patient data by combining information from two heterogeneous sources from doctors and nurse's documentation was also conducted in [8]. Their study focused on producing summary text by taking into consideration the complexities of medical terms. The first step is to extract written content of the medical document from both sources, and then to determine if there are any terms from the identified contents that belong to some simple terms without explanations, or complex terms that need explanation using some created metrics.

Research based on Natural Language Processing (NLP) together with Information Extraction (IE) techniques had been used to extract clinical knowledge from unstructured clinical information into structured clinical information. The NLP system called cTAKES was used in extracting useful and reliable clinical knowledge from EMRs written in Portuguese language from a hospital [19].

Moreover on medical domain, [20] proposed Patient Similarity Evaluation (PSE), a novel framework incorporating temporal information to medical concept embedding to measure the similarity between patient pairs. With this approach they can precisely predict the future health status of patients in advance.

The difference of this research with the previous research works is that in this study we implement Natural Language Generation to interpret the results of hematological examination of patients into the form of summary text using Template Generation System (TGen-System). TGen-System generates the template candidates (i.e. sentences with related slots) automatically which has been classified by considering the content of sentences.

IV. Methodology

The proposed system consists of two main components: the CBCI system that acts as the front-end interface for the user, and the TGen-System that acts as the back-end and supply classified template to the CBCI system. Detail illustration of the general architecture is shown in Fig. 1.

A. Data Set

Data used in this study are complete blood count (CBC) results obtained from clinical laboratories in the form of MS Word document files, describing blood components and features being examined. Data from patient's hematology laboratory examination was obtain from Hematology Analyzer, which is commonly used to measure blood samples. Using a monitor screen attached to the device, blood examination result is copied by a laboratory assistant by typing those data into predetermined MS Word template. A typical example of the template is shown in Fig. 2. This template consists of three tables, where the first table, located on the upper part of the template, records patient's personal data (such as name, age, gender, ward, date, and referring doctor). The second table, located in the middle, records the results of hematology examination (consists of four column: first column describes types or components of blood test, second column is the value obtained, third column is for units of each examination component, and the fourth column shows the reference of normal values). Last table is the lower part of the template which is used as signature columns of both laboratory assistant and head of laboratory. In this research, the MS Word document template is used as input to the system.

In a CBC, various components and features of human blood are measured, including hemoglobin which is protein containing iron molecule in red blood cells. Hemoglobin carries oxygen and carbon dioxide back and forth between the lungs and the body's tissues. Normal range values between adult male and female are slightly different and also varies among different ages. Other blood components such as blood sedimentation rate, hematocrit, and erythrocyte are also having different normal range values between male and female. Blood sedimentation rate is commonly called erythrocyte sedimentation rate (ESR) which measures how quickly the red blood cells settle at a bottom of a test tube containing blood sample. Erythrocyte gives red color in blood, while hematocrit measures the proportion of red blood cells in body (packed cell volume).

Leukocytes is another major component of blood that is also called white blood cells. This component is part of body immune system and its normal counts are no difference between male and female. Thrombocytes (or platelets) are cell fragments which circulate within the blood with relatively short life span of about 10 days. MCV (Mean corpuscular volume), MCH (mean corpuscular hemoglobin), and MCHC (mean corpuscular hemoglobin concentration) define the size (MCV) and hemoglobin content (MCH, MCHC) of red blood cells. These three contents are called red cell indices, useful in explaining the cause of anemias. Red cell indices can be calculated if the values of hemoglobin, hematocrit, and red blood cell count are known [15], [21].

The last components of a CBC are leukocytes type count which related to white blood cells (granulocytes) presented in percentages. First is eosinophil, proinflammatory white blood cell for immune system; basophil, the least common type of granulocyte, representing about 0.5% to 1% of circulating white blood cells; neutrophil, a type of white blood cell that protect human from infections; lymphocytes, include natural killer cells, cell-mediated, cytotoxic adaptive immunity, and humoral, antibody-driven adaptive immunity; and lastly monocytes, which in high level may indicate the presence of chronic infection, an autoimmune or blood disorder, cancer, or other medical conditions.

B. The CBCI System

Complete blood count information (CBCI) system is an information system to produce textual representation of complete blood examination. As shown in Fig. 1 the CBCI system acts as a front-end providing a simple interface where document of the examination result is read and textual representation is given as output. The generated output is based on selected template given from sentence templates database by looking-up related data described in the CBC data input.

Within the CBCI system there are four processes being conducted to produce the textual CBC report from MS Word document template.

1) Data Extraction: data to be extracted from the MS Word document are six patient's personal data fields and fourteen values of laboratory examination results, which are written on the first two tables of the template. An example of extracted data to be stored in temporary list structure is as follows:

['Patient name: A_7', 'Age: 52 YEARS', 'Sex: MALE', 'Ward: FLAMBOYANT 2', 'Date : 28 February 2017', 'Referring doctor: Confidential', '13.0', '12.8', '18', '243', '37.6', '4.81', '78.3', '27.0', '34.5', '2.0', '0.0', '86.0', '10.0', '2.0'].

After this process, the next process is to separate between the title fields and data values from the first part of the temporary list, and converting the string type numbers into floating point numbers, and lastly store the result into a final list, to get:

['A_7', '52', 'MALE', 'FLAMBOYANT 2', '28 February 2017', 'Confidential', '13.0', '12.8', '18', '243', '37.6', '4.81', '78.3', '27.0', '34.5', '2.0', '0.0', '86.0', '10.0', '2.0'].

2) Data Processing: this process is to find and determine the status of each data value of the blood components, whether it is in normal, decreasing abnormal, increasing abnormal, decreasing critical or increasing critical condition, based on a set of rules being setup by experts and stored in a knowledge base. Reference tables such as those shown

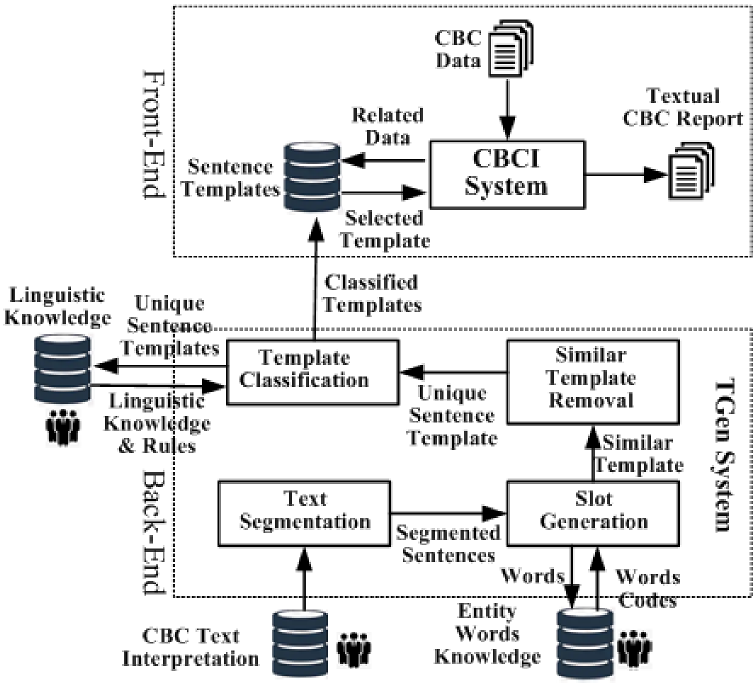

Fig. 1: Front-end and back-end of the textual CBC interpretation

in Table I, Table II, and Table III are needed in order to determine the blood condition status. The three tables show respectively the normal range, critical range, and set of rules used to determine blood condition.

3) Data Structuring: is processing of output data stored in the final list to determine and to build data structures of messages to be used in replacing slots in the sentence template. This process comprises of seven tasks, including preparing input data, declaring variable data, determining structure of opening sentence, determining structure of critical value description, determining structure of abnormal value description, determining structure of description sentence, and combining sentence data structure. An example of complete data structuring with various variables could be seen as in the following:
[['STATUSFOUND1': 'found', 'UNIT1': 'two', 'ABNORMALITY1': 'critical', 'STATUSFOUND2': found', 'UNIT2': 'four', 'ABNORMALITY2': 'abnormal', 'UNIT1': 'two', 'ABNORMALITY': 'critical', 'LEVEL1': 'decreasing', 'BLOODCOMPONENT1': 'hemoglobin and hematocrit', 'UNIT1': 'four', 'ABNORMALITY': 'abnormal', 'UNIT2': 'three', 'LEVEL1': 'increasing', 'BLOODCOMPONENT1': 'blood sedimentation rate, MCH and MCHC', 'UNIT3': 'one', 'LEVEL2': 'decreasing', 'BLOODCOMPONENT2': 'erythrocyte'], [['BLOODCOMPONENT': 'hemoglobin', 'level': '-', 'VALUEPOINT': '6.1', 'NORMALLIMIT': 'lowest', 'BLOODCOMPONENT': 'hematocrit', 'level': '-', 'VALUEPOINT': '23.6', 'NORMALLIMIT': 'lowest'], ['BLOODCOMPONENT': 'blood sedimentation level', 'level': '+', 'VALUEPOINT': '35', 'NORMALLIMIT': 'highest', 'BLOODCOMPONENT': 'MCH', 'level': '+', 'VALUEPOINT': '2.5', 'NORMALLIMIT': 'highest', 'BLOODCOMPONENT': 'MCHC', 'level': '+', 'VALUEPOINT': '6.0', 'NORMALLIMIT': 'highest', 'BLOODCOMPONENT': 'erythrocyte', 'level': '-', 'VALUEPOINT': '2.51', 'NORMALLIMIT': 'lowest']]]

Meanwhile, an example of detail interpretation of critical or abnormal values of hematology result is given such as:
[['BLOODCOMPONENT': 'hemoglobin', 'level': '-', 'VALUEPOINT': '2.3', 'NORMALLIMIT': 'lowest']]

The result of data structuring process is saved in a variabel of list data type called ODS (Output Data Structuring).

4) Microplanning: the main task of this process is called sentence generation that generate sentences to be presented into textual representation. In generating the intended sentence, microplanning interacts with the back-end TGEN-System.

C. The TGen-System

The core of the system lies on the template generation (TGen) system, which provide classified templates to the sentence template database in the front-end. The main tasks of TGen-System consist of four processes.

1) Text Segmentation: this process reads a CBC test interpretation database as input. The database is a knowledge base created by knowledge engineers from experts as a corpus of existing CBC text interpretation. An example of text interpretation is illustrated in the following:
"Based on patient's hematology examination, it is found one critical value and three abnormal values. The one critical value is critical decreasing, which is found in blood component hemoglobin. In addition, from three abnormal values there is one abnormal increasing value found in blood component leukocyte and two abnormal decreasing values found in blood components thrombocyte and hematocrit."

The segmentation process is a way to separate the corpus text into sentences using period character (".") as a delimiter. It could be seen from the illustration that there will be three segmented text to be save into a sentence list:

(a) "Based on patient's hematology examination, it is found one critical value and three abnormal values."

Patient Name : Confidential
Age : 52 YEARS
Sex : MALE

Ward : FLAMBOYAN 2
Date : 28 FEBRUARY 2017
Referring Doctor : Confidential

HEMATOLOGY EXAMINATION RESULT

EXAMINATION	RESULT	UNIT	NORMAL VALUES
COMPLETE BLOOD (CBC)			
Hemoglobin	13,0	g/dl	M=13.0 –18.0 F=12.0 – 16.0
Leukocyte	12,8	10^9/L	3.2 – 10.0
Blood Sedimentation Rate	18	mm/1 hr	M = 0 – 15 F = 0 – 20
Trombosite Amount	243	10^9/L	170 – 380
Hematocrite	37,6	%	M=40.0 – 50.0 F=35.0 – 45.0
Erythrocyte	4,81	10^{12}/L	M=4.4 – 5.6 F=3.8 – 5.0
MCV	78,3	Fl	80.0 – 100.0
MCH	27,0	Pg	28.0 – 34.0
MCHC	34,5	g/dl	32.0 – 36.0
Leukocytes type count			
Eosinophil	2	%	0 – 6.0
Basophil	0	%	0 – 2.0
Neutrophil	86	%	36.0 – 73.0
Lymphocytes	10	%	15.0 – 45.0
Monocytes	2	%	0.0 – 10.0

LAB ASSISTANT

Confidential

REGARDS,

Confidential

Fig. 2: Typical blood examination result in MS-Word format (adapted from Bahasa Indonesia)

TABLE I: Normal range of complete hematology ([16])

Examination	Unit	Sex	
		Male	Female
Hematocrit (Hct)	%	40-50	35-45
Hemoglobin (Hb)	g/dl	13-18	12-16
Erythrocytes (RBC)	10^{12}/L	4.4-5.6	3.8-5
Blood Sedimentation Rate	mm/l hr	< 15	< 20
MCV	fL	80-100	
MCH	Pg	28-34	
MCHC	g/dL	32-36	
Leucocyte	10^9/L	3.2-10	
Neutrophil (Segment)	%	36-73	
Eosinophil	%	0-6	
Basophil	%	0-2	
Lymphocytes	%	15-45	
Monocytes	%	0-10	
Trombosite Amount	10^9/L	170-380	

(b) "The one critical value is decreasing critical, which is found in blood component hemoglobin."

(c) "In addition, from three abnormal values there is one increasing abnormal value found in blood component leukocyte and two decreasing abnormal values found in blood components thrombocyte and hematocrit."

2) Slot Generation: the process is initiated by generating slot entities using regular expression (regex). Regex is used to find an entity or certain word(s) in a sentence list that corresponds to a certain tag slot type based on a list of entity slot data. Table IV lists tag slot types together with the entity slot data.

Subsequent task is assigning number to each tag slot so that microplanning process could precisely replace tag slot data in the template sentence with data obtained from data structuring process as well as to determine the type of sentence template. An example showing result of slot generation for the third sentence of the sentence list could be seen as in the following:

"In addition, from UNIT1 ABNORMALITY1 values there is UNIT2 LEVEL1 ABNORMALITY1 value found in blood component BLOODCOMPONENT1 and UNIT3 LEVEL2 ABNORMALITY1 found in blood components BLOODCOMPONENT2 and BLOODCOMPONENT3."

In this example, the uppercase words are slots to be replaced by corresponding values read.

3) Similar Template Removal: there are some possibilities that the template slots generated are duplicated since they are generated from many and possibly the same sentences. Therefore, the duplicate template slots will experience a process called similar template removal, which give unique sentence template as output.

4) Template Classification: all unique template slots obtain are then classified. The classification process is per-

TABLE II: Critical range values in hematology examination [22]

Examination	Unit	Limits	
		Lower	Upper
Hematocrit (Hct)	%	≤ 21	≥ 65
Hemoglobin (Hb)	g/dL	≤ 7.0	≥ 21
Platelet Count	$10^9/L$	≤ 20	≥ 1000
White Blood Cell Count	$10^9/L$	≤ 2.0	≥ 40

TABLE III: Rules to determine status of each blood component value

No.	Status	Rule
1	Normal	If blood component value resulted from hematology examination lies between lowest and highest ranges as in Tabel I.
2	Abnormal increasing	If blood component value resulted from hematology examination lies above highest range as in Tabel I.
3	Abnormal decreasing	If blood component value resulted from hematology examination lies below lowest range as in Tabel I.
4	Critical Increasing	If blood component value resulted from hematology examination lies above highest range as in Tabel II.
5	Critical Decreasing	If blood component value resulted from hematology examination lies below lowest range as in Tabel II.

TABLE IV: List of tag slot types with entity slot data

No.	Tag Slot	Entity Slot
1	"BLOODCOMPONENT"	"hemoglobin", "leukocyte", "bsr", "blood sedimentation rate", "thrombocyte", "hematocrit", "erythrocyte", "mcv", "mch", "mchc", "eosinophil", "basophil", "neotrophil", "lymphocyte", "monocyte"
2	"UNIT"	"one", "two", "three", "four", "five", "six", "seven", "eight" "nine", "ten", "eleven", "twelve", "thirteen", "fourteen"
3	"LEVEL"	increasing", "decreasing"
4	"ABNORMALITY"	"abnormal", "critical"
5	"STATUSFOUND"	"found", "exist"
6	"VALUEPOINT"	numeric
7	"NORMALLIMIT"	"lowest", "highest"

formed by consulting a database corpus of linguistic knowledge created by experts. From this database, linguistic knowledge as well as rules are applied to the template slots to give classified templates as output.

Input to TGen-System consists of two corpora, namely description of hematology examination result and details of critical or abnormal values of the hematology examination result. In the corpus of results descriptions, there are 20 examples of text description, containing 10 examples of description text with 3 sentences in each paragraph, and another 10 examples of description text contain 2 sentences in each paragraph, so there are 50 sentences in total. In addition, within the corpus of detail critical or abnormal values, there are 20 detail example sentences to be analyzed.

After the corpus of descriptions and corpus of details of the critical or abnormal value of the hematology examination results are analyzed by a series of TGen-System process, a set of sentence template is obtained. The number of sentence templates obtained from the TGen-System analysis consists of 10 double-opening sentence templates, 10 single opening sentence templates, 4 multiple description sentence templates, 7 single expression sentence templates, and 5 detail sentence templates.

TGen-System can support the CBCI-System performance in generating the interpretation of CBC Result in the medical report. In addition, text diversity and maintainability of the template can also be achieved by TGen-System. One example of the CBC result from a medical report given in Fig. 2 is shown in Fig. 3.

D. Evaluation of summary text naturalness level

System testing was performed by evaluating naturalness level of the sentences in summary texts through assessment of young doctors and general practitioners. The doctors were asked to fill and to evaluate a structured questionnaire. The questionnaire contains three aspects of evaluation, namely readability (understandable), clarity, and general appropriateness [23], [7], each with 0 to 5 weight scales. Subsequently, assessment was performed by experienced doctors to measure the accuracy of the summary text generated.

Patient Name	: A_7	Ward	: FLAMBOYAN 2
Age	: 52 Years	Date	: 28 February 2017
Sex	: MALE	Referring Doctor	: Confidential

HEMATOLOGY EXAMINATION RESULT

EXAMINATION	RESULT	UNIT	NORMAL VALUES
COMPLETE BLOOD (CBC)			
Hemoglobin	13.0	g/dl	P=13,0 –18,0 W=12,0 – 16,0
Leucocyte	12.8	10^9/L	3,2 – 10,0
Blood Sedimentation Rate	18	mm/1 hr	P=0 – 15 W=0 – 20
Thrombocyte Count	243	10^9/L	170 – 380
Hematocrit	37.6	%	P=40,0 – 50,0 W=35,0 – 45,0
Erythrocyte	4.81	10^{12}/L	P=4,4 – 5,6 W=3,8 – 5,0
MCV	78.3	Fl	80,0 – 100,0
MCH	27.0	Pg	28,0 – 34,0
MCHC	34.5	g/dl	32,0 – 36,0
Leukocyte Type Count			
Eosinophil	2.0	%	0 – 6,0
Basophil	0.0	%	0 – 2,0
Monocyte	2.0	%	0,0 – 10,0

DESCRIPTION TEXTS

Based on patient hematology examination result, it was found seven abnormal value(s). The seven abnormal value(s) consist of four abnormal decreasing value(s) that exist in blood component hematocrit, MCV, MCH and also lymphocyte and four abnormal increasing value(s) that exist in blood component leukocyte, blood sedimentaion rate and also neotrophil.

In detail, the above description could be explained as follows:
A. Abnormal Values
1. Hematocrit decrease 2.4 point from lowest normal limit.
2. MCV found decrease 1.7 point from lowest normal limit.
3. MCH experiencing decreasing 1.0 point from lowest normal limit.
4. Lymphocyte showed decreasing 5.0 point from lowest normal limit.
5. Leukocyte was detected increase 2.8 point from highest normal limit.
6. Blood sedimentaion rate increase 3 point from highest normal limit.
7. Neotrophil showed increasing 13.0 point from highest normal limit.

Lab. Assistant, Regards,

Confidential Confidential

Fig. 3: Summary of the patient's hematology laboratory examination in PDF format

Readability, clarity, and appropriateness aspects are intended to assess users' level of comprehension in reading sentences of summary texts, level of clarity of the summary texts, as well as whether sentences in the summary texts could enhance users' knowledge. The five evaluation scales used for each aspect are shown in Table V.

V. Result and Discussion

Result of naturalness from each evaluation aspect on 15 summary text examples generated by the system are evaluated based on the assessment of 25 young doctors. From the result shown in Figure 4, it can be seen that evaluation on the naturalness level of summary text could be described as follows:

1) Readability in terms of understandability and easiness in reading sentences of summary text by young doctors show that 96.5% agreed to say that the summary texts are understandable and only a small portion (3.5%) that say they are difficult to understand.
2) In term of clarity, results obtained from young doctors show that 97.0% say that the summary texts are clear and 3% say that the summary texts are not clear.
3) Results obtained from the young doctors on the general appropriateness of the summary texts show that 98.9% are appropriate and 1.1% consider those summary texts are not appropriate.

In addition to the evaluation from young doctors, this research is also evaluating the naturalness level of the summary text from 8 general practitioners as seen in Figure 5. From this figure, it could be seen that none of the two aspects of naturalness, i.e. readibility and clarity gave very difficult and very unclear results. The readability, clarity and appropriateness aspects of evaluation are 95.0%, 96.6%, and 87.5%, respectively. As could be expected, the difficulties in terms of naturalness considered by general practitioners are low, namely 5.0%, 3.4%, and 12.5%.

The average percentage obtained from both young doctors and general practitioners in terms of naturalness of the summary texts are 97.5% and 93.0%, respectively. These values show that the three evaluation aspects obtained from young doctor are higher than those from general practitioners. This is understandable since general practitioners have gained relatively more experience compared to young doctors, so that information in the form of short summaries are preferable. On the other hand, young doctors might

TABLE V: Evaluation of sentences in summary texts

	Readability	Clarity	Appropriateness
Scale 1	very difficult to understand	very unclear	very inappropriate
Scale 2	difficult to understand	unclear	inappropriate
Scale 3	adequate to understand	adequately clear	adequately appropriate
Scale 4	easy to understand	clear	appropriate
Scale 5	very easy to understand	very clear	very appropriate

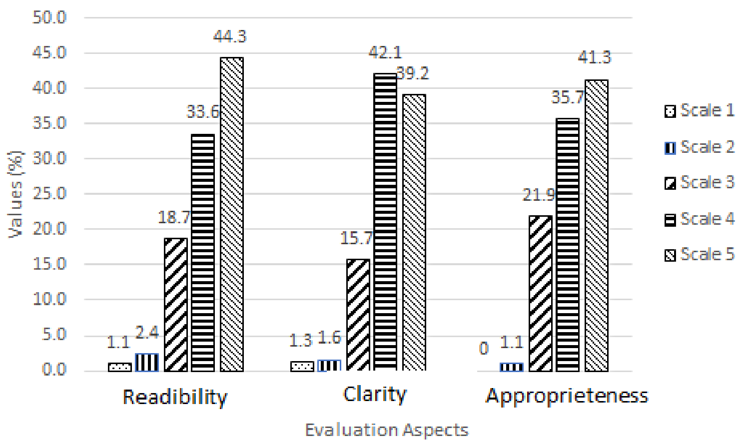

Fig. 4: Naturalness assesment of summary text by young doctors

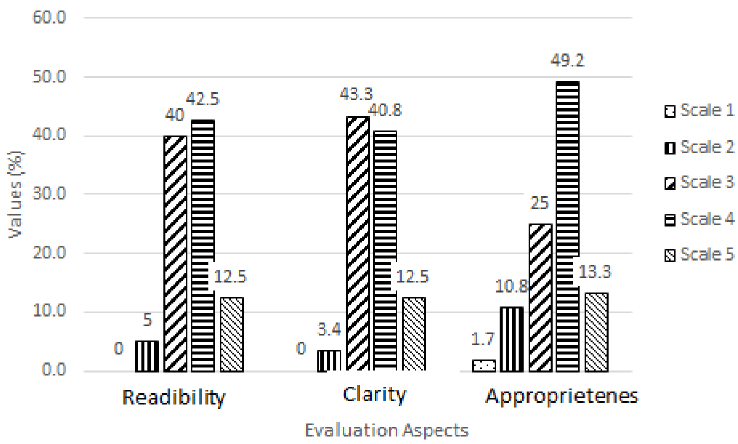

Fig. 5: Naturalness assesment of summary text by general practitioners

require more details in order to understand the summary texts.

VI. Conclusion and Future Work

The proposed template-based NLG implemented in the form of CBCI System and TGen-System is able to generate template candidates automatically by utilizing the linguistic knowledge of related experts. The summary texts generated based on these templates, fulfil three aspects of learning that could help young doctors in interpreting laboratory blood test results. The three learning aspects, namely readability (understandable), clarity (concise information), and appropriateness in delivering knowledge achieved considerably high results of 96.5%, 97.0% and 98.9%, respectively. This result also proved that the limitation of traditional template-based approach as mentioned in [11] could be minimized and the hematologic report is not only varied but also easier to understand.

Several future works could be suggested related to this research, such as using real or standard NLG system [14]. Enhancing results of summary text to be more informative, by comparing historical data obtained from previous patients' hematology laboratory results with new results obtained from laboratory examinations. Lastly, the system should be able to generate summary texts for higher level of target audiences such as medical specialists so that the information delivered is suitable to the appropriate users' level of knowledge.

References

[1] K. Kreimeyer, M. Foster, A. Pandey, N. Arya, G. Halford, S. F. Jones, R. Forshee, M. Walderhaug, and T. Botsis, "Natural language processing systems for capturing and standardizing unstructured clinical information: A systematic review," *Journal of Biomedical Informatics*, vol. 73, no. 1, pp. 14–29, 2017.

[2] S. Velupillaia, H. Suominen, M. Liakatae, A. Robertsa, A. D. Shahf, K. Morleya, D. Osborn, J. Hayes, R. Stewart, J. Downs, W. Chapman, and R. Dutta, "Using clinical natural language processing for health outcomes research: Overview and actionable suggestions for future advances," *Journal of Biomedical Informatics*, vol. 88, no. 1, pp. 11–19, 2018.

[3] E. Reiter and R. Dale, *Building Natural Language Generation System*. Cambridge: Cambridge University Press, 2000.

[4] O. Dusek, J. Novikovaa, and V. Riesera, "Evaluating the state-of-the-art of end-to-end natural language generation: The E2E NLG challenge," *Computer Speech & Language*, vol. 59, no. 1, p. 123–156, 2020.

[5] S. Mahamood and E. Reiter, "Generating affective natural language for parents of neonatal infants," in *Proceedings of the 13th European Workshop on Natural Language Generation (ENLG)*, 2011, pp. 12–21.

[6] B. Eugenio, A. Boyd, C. Lugaresi, A. Balasubramanian, G. Keenan, M. Burton, T. Macieira, K. Lopez, C. Friedman, J. Li, and Y. Lussier, "Patientnarr: Towards generating patient-centric summaries of hospital stays," in *Proceedings of the 8th International Natural Language Generation Conference*, 2014, pp. 6–10.

[7] I. Aulia, *Automatic Chart Interpreter System For Generating Health Surveillance Summaries Based On Indonesian Language*. Bandung: Telkom University, 2015.

[8] S. Archarya, B. Eugenio, A. Boyd, K. Lopez, R. Cameron, and G. Keenan, "Generating summaries of hospitalizations: A new metric to assess the complexity of medical terms and their definitions," in *Proceedings of the 9th International Natural Language Generation conference*, 2016, pp. 26–30.

[9] F. Portet, E. Reiter, A. Gatt, J. Hunter, S. Sripada, Y. Freer, and C. Sykes, "Automatic generation of textual summaries from neonatal intensive care data," in *Lectures Notes in Artificial Intelligence: Artificial Intelligence in Medicine, AIME 2007*, pp. 789–816.

[10] O. Biran, *Data-Driven Solutions to Bottlenecks in Natural Language Generation*. University of Columbia, New York: Dissertation, 2016.

[11] S. Busemann and H. Horacek, "A flexible shallow approach to text generation," in *Proceedings of the 9th International Workshop on Natural Language Generation*, 1998, pp. 238–247.

[12] A. Schneider, P. Vaudry, A. Mort, C. Mellish, E. Reiter, and P. Wilson, "MIME - NLG in pre-hospital care," in *Proceedings of the 14th European Natural Language Generation Workshop (ENLG)*, 2013, pp. 152–156.

[13] K. v. Deemter, E. Krahmer, and M. Theune, "Real versus template-based natural language generation: A false opposition?" *Association for Computational Linguistics*, vol. 31, no. 1, pp. 15–23, 2005.

[14] T. Becker, "Practical, template–based natural language generation with tag," in *Proceedings of the 6th International Workshop on Tree Adjoining Grammar and Related Frameworks (TAG+6)*, 2002, pp. 80–83.

[15] P. R. Sarma, "Red cell indices," in *Clinical Methods, 3rd edition: The History, Physical, and Laboratory Examinations*, H. Walker, W. Hall, and J. Hurst, Eds. Boston: Butterworths, 1990, ch. 152, p. 720.

[16] F. Herawati and R. Andrajati, *Pedoman Interpretasi Data klinik (in Bahasa)*. Jakarta: Ministry of Health Republic of Indonesia, 2011.

[17] E. Reiter, "An architecture for data-to-text systems," in *Proceedings of the 11th European Workshop on Natural Language Generation (ENLG-07)*, Schloss-Dagstuhl, Germany, 2007, p. 97–104.

[18] D. Scott, C. Hallett, and R. Fettiplace, "Data-to-text summarisation of patient records: Using computer-generated summaries to access patient histories," *Patient Education and Counseling*, vol. 92, no. 2, p. 153–159, 2013.

[19] M. Lamy, R. Pereira, J. C. Ferreira, F. Melo, and I. Velez, "Extracting clinical knowledge from electronic medical records," *IAENG International Journal of Computer Science*, vol. 45, no. 3, pp. 488–493, 2018.

[20] Z. Lin and D. Yang, "Medical concept embedding with variable temporal scopes for patient similarity," *Engineering Letters*, vol. 28, no. 3, pp. 651–662, 2020.

[21] F. Habibzadeh, M. Yadollahie, M. Roshanipoor, and M. Haghshenas, "Derivation of blood hemoglobin concentration from hematocrit: A simple method for rural areas," *Archives of Iranian Medicine*, vol. 4, no. 3, pp. 120–122, 2001.

[22] ARUP Laboratories, "Critical values list," University of Utah Department of Pathology. [Online]. Available: https://www.aruplab.com/files/resources/testing/ARUP_Critical_Values.pdf

[23] A. Belz and E. Reiter, "Comparing automatic and human evaluation of NLG systems," in *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*. Trento, Italy: Association for Computational Linguistics, 2006. [Online]. Available: https://www.aclweb.org/anthology/E06-1040

Opim Salim Sitompul completed his PhD on Information Science from Universiti Kebangsaan Malaysia, Selangor in 2005. He is a professor at Department of Information Technology, Universitas Sumatera Utara, Medan, Indonesia. He is a member of IAENG since 2017.

Erna Budhiarti Nababan completed her PhD on Science and Management Systems from Universiti Kebangsaan Malaysia, Selangor in 2010. She is currently a senior lecturer at Department of Information Technology, Universitas Sumatera Utara, Medan, Indonesia. She is a member of IAENG since 2017.

Dedy Arisandi is currently a lecturer at Department of Information Technology, Universitas Sumatera Utara, Medan, Indonesia

Indra Aulia is currently a lecturer at Department of Information Technology, Universitas Sumatera Utara, Medan, Indonesia

Hengky Wijaya was formerly a student at Department of Information Technology, Universitas Sumatera Utara. He is currently the Platform Management Officer at PT Bank Syariah Mandiri, Jakarta, Indonesia