A Processing Model Using Natural Language Processing (NLP) For Narrative Text Of Medical Record For Producing Symptoms Of Mental Disorders

Sri Mulyana
Intelligent Systems
Research Laboratory,
Universitas Gadjah Mada
Yogyakarta, Indonesia
smulyana@ugm.ac.id

Sri Hartati
Intelligent Systems
Research Laboratory,
Universitas Gadjah Mada
Yogyakarta, Indonesia
shartati@ugm.ac.id

Retantyo Wardoyo
Intelligent Systems
Research Laboratory,
Universitas Gadjah Mada
Yogyakarta, Indonesia
rw@ugm.ac.id

Subandi
Faculty of Psychology
Universitas Gadjah Mada
Yogyakarta, Indonesia
subandi@ugm.ac.id

***Abstract*—Mental health is still a serious problem in Indonesia. Data on basic health research in 2013 showed the prevalence of serious mental disorders in Indonesia reached 1.7 permil. The ratio of professional psychiatrists with the number of sufferers is 0.47 per 100000 patients which means that they still do not meet the requirements of WHO. This has an impact that not all people with mental disorders can be handled by professional psychologists directly. Therefore, the development of a case-based reasoning system (CBR) to help diagnosing types of mental disorders can be an option to overcome these gaps.**

In most health care places, an initial examination of mental patients is carried out by medical professionals who are not experts. At a certain level, the patient cannot express the symptoms that are felt, so the examiner can make narrative texts about the patient's condition. On the other hand, text documents are not structured that are commonly processed in CBR systems, so they cannot be directly applied. Therefore a text processing model to become symptoms is required.

In this paper, the text medical record processing model using Natural Language Processing (NLP) is discussed to produce symptoms of mental disorders. This model is a part of the development of a case-based reasoning computer system to help diagnosing the types of mental disorders and their management.

Keywords—text processing, natural language processing, mental disorders, symptoms, medical record

I. Introduction

In Indonesia, mental health is a serious issue with the report obtained from the Database of Health Research of 2013 putting the prevalence of serious mental disorder to be 1.7 per mil with 1-2 of 1,000 individuals having the problem. This condition was exacerbated by the minimal professional psychologists and mental health service facilities in various parts of the country, making many patients lack proper treatment. Therefore, a computer reasoning system was developed as an alternative to bridge the gap.

Most of the professional medical personnel such as specialists could only be found in big cities with none in the majority of the medical facilities existing in local areas. However, other medical personnel does not have the ability required to make anamnesis for patients like specialists in mental disorder.

The steps involved in diagnosing patients with a mental disorder in clinical medical practice are as shown in Fig. 1.

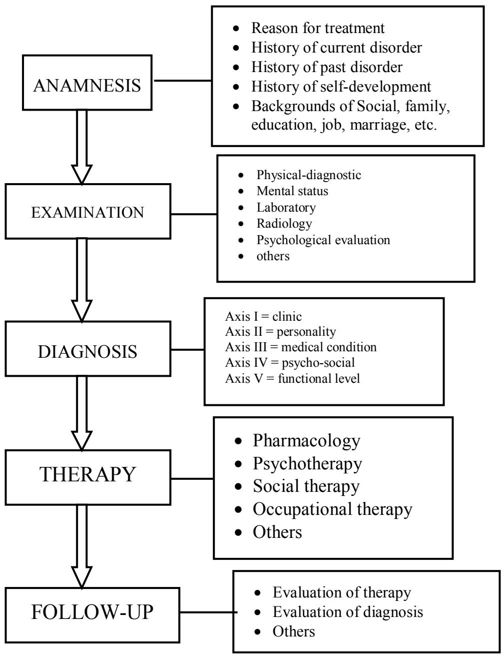

Fig. 1. Procedure of mental disorder diagnosis [1]

The examination can only be conducted by a psychologist or psychiatrists while the diagnosis, therapy, and follow-up processes undertaken by a specialist. Physicians and other paramedics may conduct anamnesis.

In some situations, the patients may not have the ability to express the symptoms being felt, therefore, the examiner would be required to make a narrative text concerning the condition. Further-more, the process of diagnosis, therapy, and follow-up could be conducted through the use of a CBR system and the procedure is shown in Fig-2.

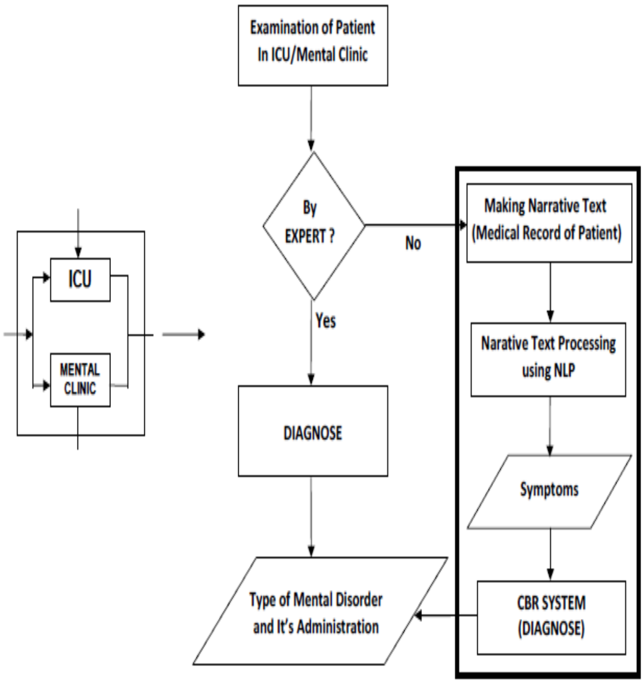

Fig-2 : Flowchart of examination for patients with CBR

However, the textual document processed directly in the CBR system is not structured. Therefore, a model of text processing with NLP is needed to determine different types of schizophrenia and mood disorders. This involves the processing of the narrative text obtained from the examination results to generate the symptoms.

Therefore, after the medical personnel has written texts describing the conditions of patients, the *Natural Language Processing* (NLP)-based system could be used to process the texts to get the symptoms.

Natural Language Processing (NLP) is one of the *Artificial Intelligence* (AI) applications developed in which a computer understands an inputted natural language and gives desirable responses after some processing. Natural language is used by humans for communications and the computer also needs to process the language in order to make it understandable and useful for the users.

II. RESEARCH METHOD

A. *Literature Review*

Text processing is part of text mining applied in different research areas. It involves changing textual documents into meaningful and vital information before the implementation of text-mining strategy and it is usually started with preprocessing to eliminate all undesirable and unnecessary data. *Natural Language Processing* (NLP) widely contributes to text processing because it involves preprocessing steps such as stemming, part-of-speech (POS) tagging, chinking, parsing, and information extraction. However, each of these has a different procedure, rules, approaches and methods [2].

In the medical field, *Clinical Decision Support Systems* (CDSS) software has been designed to help make a clinical decision where individual characteristics of patients are matched to the records stored in computers in order to generate specific assessment/recommendations needed by physicians to take appropriate actions [3]. This system has several types of entries including structured data like electronic medical records, semi-structured data like XML documents or results of laboratory examination, or unstructured data such as narrative texts, clinical observation for patients, radiologic report, surgery notes. However, most of the medical information forms for patients are classified under unstructured data because they are either narrative or free texts. Therefore, the use of *Natural Language Processing* techniques has been proposed for the implementation of *Clinical Decision Support Systems(CDSS)*, especially with entry text processing [4].

Natural Language Processing (NLP) is the study and application of the mechanism used in understanding natural language [5]. It is a scientific discipline supporting CDSS because it has the system to extract free texts with relevant vital information. Furthermore, the information extracted can be used by CDSS to obtain the results needed by a physician to make accurate decisions.

The patients are allowed to enter medical data concerning their health status into the NLP-based CDSS system while the physicians also have the ability to enter information in order to come up with conclusions and recommendations needed for the diagnosis, treatment, and monitoring of patients' health.

Most of the NLP-based CDSS applications developed make use of English domain as observed in [6] [7] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19] [20] [21] [22] [23] [24]. Some applications other than English include Spanish [25] and German [26].

The input of clinical documents into CDSS either as free texts or structured data determines the methods to be used by NLP. This is important because there is a significant difference between NLP-based CDSS processing for structured data like electronic medical records of database or XML [21] and natural language with free texts [17]. Such that the processing of documents with free texts requires more processes than structured data.

B. *Development of Text Processing Model using NLP*

These text-processing models with NLP are part of the *Case-Based Reasoning* Syatem as shown in Fig. 3.

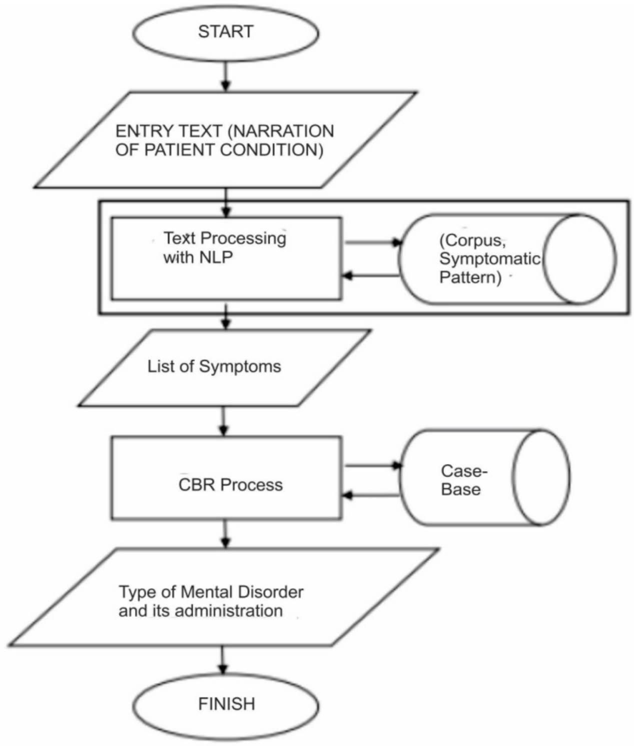

Fig. 3. General Picture of Research

The initial step to support the development of text processing with NLP is the structuringn of database and case basis. Structuring of database and case-base is based on medical records of patients and the stages are as shown in Fig.4.

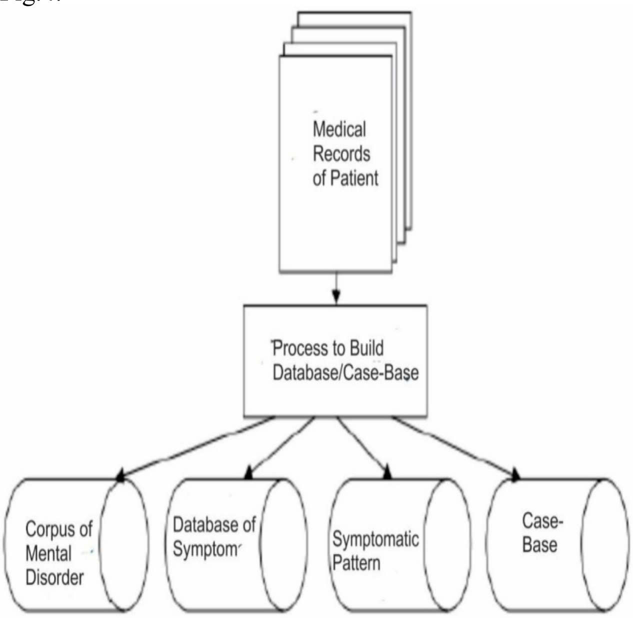

Fig.4. structuring of Database and Case-Base

To support text-processing models with NLP, it is necessary to build the following:

1. Case-Base

Case-Base contains a collection of cases that has ever occurred based on the collected medical records. An example is shown in Table 1.

Table 1. Example of Case-Base

No	**Code**	**Type of disorder**	**Symptom**
1	F32.0	Mild depression	G09: decreasing energy and activities G10: loss of interest and gladness G15: feeling false and useless G18: insomnia & easily fatigued

2. Database of mental disorder symptoms

This contains all symptoms facing patients based on medical records as illustrated in Table 2.

Table 2. Some Symptoms observed on Patients

No.	**Code**	**Name**
1	G01	Increasing mood
2	G02	Increasing energy and activities
3	G03	Confusion in working and activities
4	G04	Increasing competency to make more friendships
5	G05	Increasing sexual energy
6	G06	Decreasing sleep
7	G07	Talkative
8	G08	Easily offended
9	G09	Decreasing energy and activities
10	G10	Loss of interest and gladness
11	G11	Decreasing concentration and attention
12	G12	Easily distracted attention
13	G13	Great thought
14	G14	Confusing self-esteem
15	G15	Feeling false and useless
16	G16	Easily Suspicious
17	G17	Easily Fatigued
18	G18	Insomnia

3. Pattern of symptoms

This is a structure based on the results of symptomatic pattern analysis to determine main words/terms characterizing a symptom. The words are followed by specific other specific words as shown in Table 3.

Table-3. Symptomatic Pattern

No.	**Symptom**	**Symptomatic pattern**	
		Characteristics of symptoms	**Words following**
G01	Increasing mood	Increasing	Mood
G02	Increasing energy and activities	Increasing	Energy
			Activities
G03	Confusion in working and activating	Confusion	Working
			Activating
G04	Increasing competency to make more friendships	Increasing	Competency to make friendships
			More friendships
G05	Increasing sexual energy	Increasing	Sexual
			Sexual energy
G06	Decreasing sleep	Decreasing	Sleep
G07	Talkative	Talk	Active

G08	Easily offended	Easily	Offended
G09	Decreasing energy and activities	Decreasing	Energy Activities
G10	Losing interest and gladness	Losing	Interest Gladness
G11	Decreasing concentration and attention	Decreasing	Concentration Attention
G12	Easily distracted attention	Easily	Distracted Attention
G13	Great though	Great	Thought
G14	Raising self-esteem	Raising	Self-esteem
G15	Feeling false and useless	Feeling	False Useless
G16	Easily suspicious	Easily	Suspicious
G17	Easily fatigued	Easily	Fatigued
G18	Disturbed sleep	Disturbed	Sleep

4. Corpus of mental disorder

This is a list of all words/terms used in processing texts with NLP. It is usually obtained from results of the symptomatic pattern as shown in Table 4.

Table 4. Corpus of Mental disorders

Increasing, confusion, decreasing, many, easy, loss, great, raising, feeling, mood, sense, mood, energy, activities, friendship, competency to make friendship, making more friendships, sexual, energy, sexual, working, activating, sleeping, need of sleeping, concentration, attention, talk, offended, distracted attention, distracted, fatigued, gladness, interest, though, self-esteem, false, useless.

III. RESULTS AND DISCUSSION

This is a process of narrating the texts concerning the condition of patients based on their symptoms. The components are shown in Fig.5.

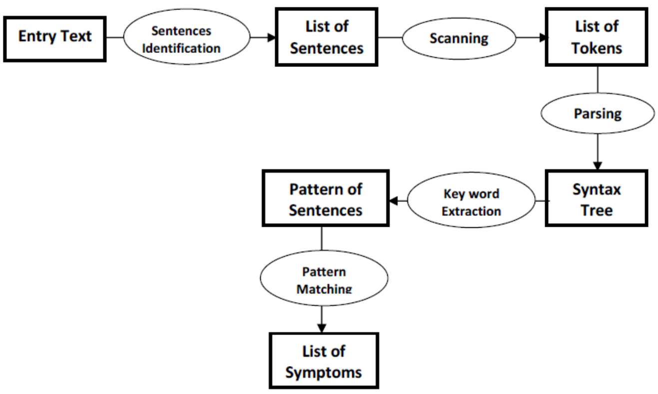

Fig.5. Stage of Text Processing with NLP

Based on the figure, it can be concluded that the overall text processing with NLP starts with sentence identification which involves dividing entry texts into sentences.

For example :

"Patients have lost interest and gladness. It seemed decreasing energy and activities, making the patients feel easily offended. It was also observed that there is a decreasing concentration proven by easily distracted attention. Patients also felt they were useless in implementing life."

Process of sentence identification divides the entry texts as shown in Fig. 6.

1. Patients have lost interest and gladness
2. It seemed decreasing energy and activities, making the patients feel easily offended
3. It was also observed that there is decreasing concentration proven by easily distracted
4. Patients felt that they have been useless in implementing their life.

Fig.6. Results of sentence identification

Each of the generated sentences was processed through the stage of scanning by determining words characterizing the symptoms found in them. Furthermore, parsing and extraction of keywords were conducted by matching each token with corpus specific to the characteristic of mental disorder symptoms as shown in Fig. 7.

1. Loss of gladness interest
2. Decreasing activity energy is easily fatigued
3. Decreasing attention concentration is easily distracted
4. Useless

Fig.7. Example of keyword extraction results

Furthermore, *pattern matching* of symptoms was conducted by matching tokens collected using the sentence pattern to select the symptoms. This process includes corpus, synonym, and *pattern matching* databases

It also considered synonyms in order to cater for the words excluded in the corpus such that words found to be synonymous were replaced by those in the corpus, otherwise, they were ignored because they were not associated with mental health. The scheme of *pattern matching* is as shown in Fig.8.

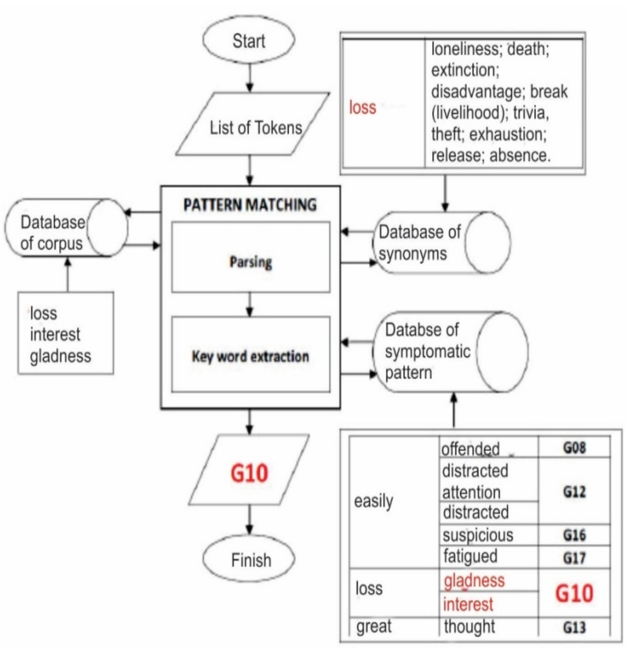

Fig.8. Scheme of *Pattern Matching*

Based on the list obtained, the rules were structured in selecting the symptoms based on *pattern matching* as shown in Table 5.

Table 5. Rules to determine symptoms based on *pattern matching*

Pattern of symptoms		Enforced symptoms
Increasing	Mood	G01
	Feeling	
	Sense	
	Energy	G02
	Activities	
	Friendship	G04
	Competency to make friendship	
	Making more friendship	
	Sexual	G05
	Sexual energy	
Confusion	Working	G03
	Activating	
Decreasing	Sleep	G06
	Need of sleep	
	Activities	G9
	Energy	
	Concentration	G11
	Attention	
Active	Talk	G07
Easily	Offended	G08
	Distracted attention	G12
	Distracted	
	Suspicious	G16
	Fatigued	G17
Loss	Gladness	G10
	Interest	
Great	Thought	G13
Raising	Self-esteem	G14
	False	G15
	Useless	
Disturbed	Sleep	G18

With respect to the list of tokens obtained, the following symptoms were generated by the *pattern matching*:

1. G09 : decreasing energy and activities
2. G10 : loss of interest and gladness
3. G11 : decreasing concentration and attention
4. G12 : easily distracted attention
5. G15 : feeling false and useless
6. G17 : easily fatigued

IV. CONCLUSION

The results showed the model of narrative text processing for mental disorder records with NLP to generate symptoms has been developed. It also revealed that the model built requires a *Case-Based Reasoning* system to help diagnose types of mental disorder and administration.

REFERENCES

[1] Maslim, R., 2013, Diagnosis Gangguan Jiwa, Rujukan Ringkas PPDGJ-III dan DSM, Bagian Ilmu Kedokteran Jiwa FK Unika Atma Jaya, Jakarta.

[2] Dudhabaware R. S. dan Madankar M. S., 2014, Review on Natural Language Processing Tasks for Text Documents, 978-1-4799-3975-6/14 IEEE

[3] Hunt D. L., Haynes R. B., Hanna S. E. dan Smith K., 1998, Effects of computerbased clinical decision support systems on physician performance and patient outcomes: a systematic review, Journal of the American Medical Association, vol. 280, no. 15, pp. 1339-1346.

[4] Ortiz J. A. R., Beltran B. A. G. Dan Lopez L. G., 2015, Clinical Decision Support Systems: A survey of NLP-based Approaches from Unstructures Data, 26th International Workshop on Databased ang Expert Systems Applications,pp. 163-167

[5] Allen J., 1995, Natural Language Understanding, Benjamin Cummings Publishing/ Addison Wesley, New York

[6] Mendona E. A., Haas J., Shagina L., Larson E. dan Friedman C. , 2005 Extracting information on pneumonia in infants using natural language processing of radiology reports, Journal of biomedical informatics, vol. 38, no. 4, pp. 314-321

[7] Hripcsak G., Bakken S., Stetson P. D. dan Patel V. L. , 2013, Mining complex clinical data for patient safety research: a framework for event discovery, Journal of biomedical informatics, vol. 36, no. 1, pp. 120-130

[8] Meystre S. dan Haug P. J., 2006, Natural language processing to extract medical problems from electronic clinical documents: performance evaluation, Journal of biomedical informatics, vol. 39, no. 6, pp. 589-599

[9] Anderson H. D., Pace W. D., Brandt E., Nielsen R. D., Allen R. R., Libby A. M. dan Valuck R. J., 2015, Monitoring suicidal patients in primary care using electronic health records, The Journal of the American Board of Family Medicine, vol. 28, no. 1, pp. 65-71

[10] Hazlehurst B., Frost H. R., Sittig D. F. dan Stevens V. J., 2005, MediClass: A system for detecting and classifying encounter-based clinical events in any electronic medical record, Journal of the American Medical Informatics Association, vol. 12, no. 5, pp. 517-529

[11] Friedlin F. J. dan McDonald C. J., 2008, A software tool for removing patient identifying information from clinical documents, Journal of the American Medical Informatics Association, vol. 15, no. 5, pp. 601-610

[12] Wagholikar K. B., MacLaughlin K. L., Henry M. R., Greenes R. A., Hankey R. A., Liu H. dan Chaudhry R., 2012, Clinical decision support with automated text processing for cervical cancer screening, Journal of the American Medical Informatics Association, vol. 19, no. 5, pp. 833-839

[13] Fiszman M., Chapman W. W., Aronsky D., Evans R. S. dan Haug P. J., 2000, Automatic detection of acute bacterial pneumonia from chest Xray reports, Journal of the American Medical Informatics Association, vol. 7, no. 6, pp. 593-604

[14] Jansen P. B., Jensen L. J. dan Brunak S., 2012, Mining electronic health records: towards better research applications and clinical care, Nature Reviews Genetics, vol. 13, no. 6, pp. 395-405

[15] Jain N. L., Knirsch C. A., Friedman C. dan Hripcsak G., 1996, Identification of suspected tuberculosis patients based on natural language processing of chest radiograph reports, in Proceedings of the AMIA Annual Fall Symposium, American Medical Informatics Association, Washington, USA, pp. 542546

[16] Raja A. S., Ip I. K., Prevedello L. M., Sodickson A. D., Farkas C., Zane R. D. dan Khorasani R., 2012, Effect of computerized clinical decision support on the use and yield of CT pulmonary angiography in the emergency department, Radiology, vol. 262, no. 2, pp. 468-474

[17] Roberts A., Gaizauskas R. and Hepple M., 2008, Extracting clinical relationships from patient narratives, in Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing, Association for Computational Linguistics, Ohio, USA, pp. 10-18

[18] Portet F., Reiter E., Gatt A., Hunter J., Sripada S., Freer Y. dan Sykes C. , 2009, Automatic generation of textual summaries from neonatal intensive care data, Artificial Intelligence, vol. 173, no. 7, pp. 789-816

[19] Matheny M. E., FitzHenry F., Speroff T., Green J. K., Griffith M. L., Vasilevskis E. E. dan Brown S. H., 2012, Detection of infectious symptoms from VA emergency department and primary care clinical documentation, International journal of medical informatics, vol. 81, no. 3, pp. 143-156

[20] Byrd R. J., Steinhubl S. R., Sun J., Ebadollahi S. dan Stewart W. F., 2014, Automatic identification of heart failure diagnostic criteria, using text analysis of clinical notes from electronic health records, International journal of medical informatics, vol. 83, no. 12, pp. 983-992

[21] Friedman C., Hripcsak G., Shagina L. H. dan Liu H., 1999, Representing information in patient reports using natural language processing and the extensible markup language, Journal of the American Medical Informatics Association, vol. 6, no. 1, pp. 76-87

[22] Fung K. W., Jao C. S. and Demner-Fushman D., 2013, Extracting drug indication information from structured product labels using natural language processing, Journal of the American Medical Informatics Association vol. 20, no. 3, pp. 482-488

[23] Hunter J., Gatt A., Portet F., Reiter E. dan Sripada S., 2008, Using natural language generation technology to improve information flows in intensive care units, in European Conference on Artificial Intelligence, Patras, Greece, pp. 678-682

[24] Imler T. D., Morea J., Kahi C. dan Imperiale T. F., 2013, Natural language processing accurately categorizes findings from colonoscopy and pathology reports, Clinical Gastroenterology and Hepatology, vol. 11, no. 6, pp. 689-694

[25] Watson A. J., O'Rourke J., Jethwani K., Cami A., Stern T. A., Kvedar J. C. dan Zai A. H., 2011, Linking electronic health record-extracted psychosocial data in real-time to risk of readmission for heart failure, Psychosomatics, vol. 52, no. 4, pp. 319-327

[26] Stocker C., Marzi L. M., Matula C., Schantl J., Prohaska G., Brabenetz A. dan Holzinger A., 2014, Enhancing patient safety through human-computer information retrieval on the example of German-speaking surgical reportsin 25th International Workshop on Database and Expert Systems Applications (DEXA), Munich, Germany, pp. 216-220