Proceedings of the International MultiConference of Engineers and Computer Scientists 2019
IMECS 2019, March 13-15, 2019, Hong Kong

Text Data Mining of the Nursing Care Life Log from Electronic Medical Record

Muneo Kushima, Tomoyoshi Yamazaki and Kenji Araki

Abstract—**In this research, we analyze nursing care records by applying an analysis tool KH Coder to nursing care life logs, in order to develop a method for visualizing and verifying nursing care actions. We use 161 nursing care life logs recorded in Long-term Care Health Facility S in M City. We identify the descriptions related to work contents of nursing care workers from the nursing care records including various intentions and contexts. The analysis results using KH Coder showed that central issues in nursing care were extracted, and the role of nursing care such as overall structure of various subjects was clarified. We found that the analysis results have potential to clarify the work content of care workers. As the nursing field requires efficiency in health care services, improvement and continuous data collection are important for the long-term building of health care services as well as large-scale data collection. In the future, we aim to develop an Electronic Medical Record that can be created semi-automatically in accordance with the level of care required.**

Index Terms— **Medical information, Electronic Medical Record, Text data mining, Nursing care life log, KH Coder.**

I. Introduction

An Electronic Medical Record (EMR) records patient information by computers instead of by papers. Not only the data but also the entire management system may be called EMR. The expected effect simplifies the entire process of hospital management and improves medical care [1, 2]. Because data are managed electronically, input data can be easily managed in comparison with paper-based medical records [3]. Information can be easily shared electronically [4, 5]. On the other hand, falsification must be prevented and the originality of the data must be guaranteed.

Data mining searches for correlations among items by analyzing a great deal of such accumulated data as sales data and telephone call histories. Text data mining resembles data mining because it extracts useful knowledge and information by analyzing the diversified viewpoints of written data [6].

Recently, the interest has risen in text data mining because it uncovers useful knowledge buried in a large amount of accumulated documents [7, 8].

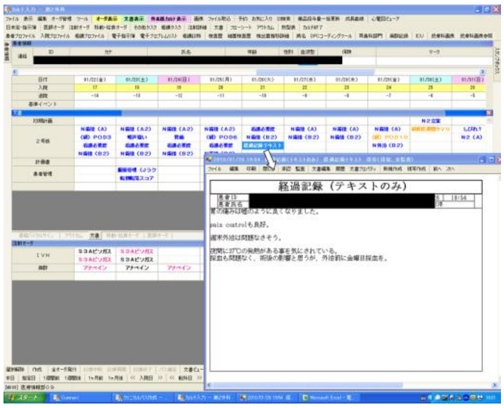


Fig. 1 Screen shot of an EMR

Research has started to apply text data mining to medicine and healing [9, 10]. In addition, the speed of electronic medical treatment data is accelerating because of the rapid informationization of medical systems, including EMRs. Recently, research on data mining in medical treatment that aims for knowledge and pattern extraction from a huge accumulated database is increasing. However, many medical documents, including EMRs that describe the treatment information of patients, are text information. Moreover, mining such information is complicated. The data arrangement and retrieval of such text parts become difficult because they are often described in a free format; the words, phrases, and expressions are too subjective and reflect each writer [11].

In the future, the text data mining of documents will be used for lateral retrieval, even in the medical treatment world, not only by the numerical values of the inspection data but also by computerizing documents.

In this research, we analyze the care life logs using an analysis tool KH Coder for visualizing and verifying nursing care actions.

II. EMR at University of Miyazaki Hospital

Fig. 1 shows a screen shot of an EMR. When the medical information system was updated on May, 2006, the University of Miyazaki Hospital introduced a package version of the EMR system called Integrated Zero-Aborting NAvigation system for Medical Information, which was developed in collaboration with a local IT company. The recorded main data include patient's symptoms, laboratory results, prescribed medicines, and the tracking of the changed data. The cases of making both the images of

Manuscript received 10. Nov 2018, revised 5. Jan, 2019.
This work was supported by JSPS KAKENHI Grant Number JP18K11530.
M. Kushima, T. Yamazaki and K. Araki are with the Faculty of Medicine at the University of Miyazaki Hospital.
e-mail: muneo_kushima@med.miyazaki-u.ac.jp,
e-mail: yama-cp@med.miyazaki-u.ac.jp,
e-mail: taichan@med.miyazaki-u.ac.jp.
http://mit.med.miyazaki-u.ac.jp/.
5200, Kihara, Kiyotake-cho, Miyazaki-shi, Miyazaki 889-1692 Japan.
Tel: +81-985-85-9057, Fax: +81-985-84-2549.

X-rays and the appended electronic materials are not infrequent either. If a network is used, EMR can be shared not only in one hospital but also among two or more hospitals.

EMR has a unique feature that is different from those being operated at many other university hospitals.

First, the electronic card systems used so far in university hospitals were all developed by major medical system venders, but EMR was developed in collaboration with local companies. The advantages of collaboration with local companies included prompt communication and lower costs.

Second, we focused on performance, especially the speed at which the screen opens.

Third, we aimed for a useful system to improve management, reflecting a request by the University of Miyazaki Hospital after it was incorporated.

We made the medical staff concretely are aware of the cost and made the management analysis system work closely with the EMR system and showed its cost when the system was ordered.

III. TEXT DATAMINING APPLICATION TO MEDICINE

Text data mining is often used to analyze information hidden in the text of a document and to extract key words, phrases, and even concepts from written documents. Text data mining or data mining, which is roughly equivalent to text analytics, refers to the process of deriving high-quality information from texts.

Text data mining usually structures the input text (often by parsing, adding derived linguistic features, removing others and inserting into a database), derives patterns within the structured data, and finally evaluates and interprets the output.

Fig. 2 shows the process of text data mining. Two particular aspects should be considered when applying text data mining to a medical context. Second, final decisions regarding courses of treatment can be obtained.

One difficulty in applying text data mining to medicine is the entire process of identifying symptoms for understanding the associated risks while taking appropriate action.

IV. NURSING CARE LIFE LOG

A Nursing Care Life log records a 24-hour period of the caregiver's activity. It is also utilized as a long-term service content record. The recording itself is not the main purpose, but it transmits information to others, accumulates and analyzes data, and aims to lead the service to better care.

The text data of the nursing care record is a text record integrating the facility service usage record of the care receiver and the observation record of the care giver. It is used for the cooperation and transmission to other occupations and grasp of the state of the care receiver among the care givers.

Also, due to effective operation and improvement of nursing care work, education / training of nursing care workers, development of secondary use of nursing care records is strongly desired from nursing staff in the field.

In the development of secondary usage of data accumulated in nursing care records at nursing care facilities, the amount of text data is enormous, and it was a major obstacle to organize data systematically. As a method to overcome this obstacle and to acquire knowledge that can be used to solve the above problem from enormous text data, text mining technology has attracted attention.

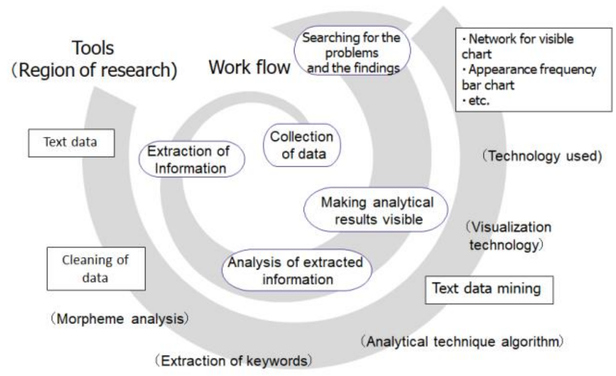


Fig.2 Process of text data mining

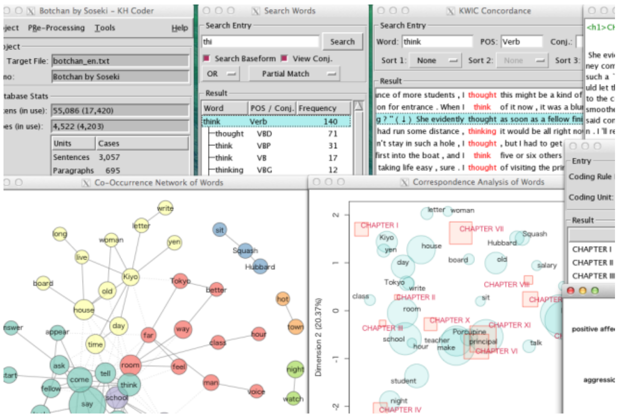


Fig. 3 Example of a screen shot of KH Coder

Generally, a life log is a technique of recording human life, work, experience as digital data such as video / audio / position information, or the record itself. In this research, we use text data recorded at nursing care site.

V. KH CODER

KH Coder is an open source software for computer assisted qualitative data analysis, particularly quantitative content analysis and text mining. It can be also used for computational linguistics. It supports processing and etymological information of text in several languages, such as Japanese, English, French, German, Italian, Portuguese and Spanish. Specifically, it can contribute factual examination co-event system hub structure, computerized arranging guide, multidimensional scaling and comparative calculations.

It is well received by researchers worldwide and used in a large number of disciplines, including neuroscience, sociology, psychology, public health, media studies, education research and computer science.

KH Coder has been reviewed as a user friendly tool "for identifying themes in large unstructured data sets, such as

ISBN: 978-988-14048-5-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

online reviews or open-ended customer feedback" and has been reviewed in comparison to WordStat.

KH Coder supports various kinds of searches with frequency tables indicating what kind of words appeared frequently. Furthermore, the concepts contained in the data can be investigated by looking at groups of words appearing together or groups of documents containing the same words, based on multivariate analysis [12]. Moreover, the characteristics of the document group can be identified by listing words which appear particularly frequently in the document group. It is possible to automatically classify documents according to criteria designated by analysts. Fig. 3 is an example of a screen shot of KH Coder.

VI. Analysis Results

In order to identify the care worker's work contents and the related descriptions, we analyze 161 nursing care life logs including various intentions and contexts, recorded in geriatric health facility S in M city. In this research, we evaluate the visualization result to judge whether or not it is important for the care worker desires.

Text analysis results where the input is nursing care record text data are shown as below. It is an environment that can interactively acquire output results according to the interests of care workers using multiple result diagram panels.

Table 1 shows frequently occurring words. In Table 1, the most frequent word was "toilet". Frequent keywords related to the work content in nursing care facilities were "toilets", "urination", "doing", "calling", "saying", "morning", "sleeping", "wheelchair", "voice", "induction", "putting", "assistance".

Fig. 4 (a)(b) show network diagrams of words and word connections.

In Fig. 4 (a), "toilet" is a keyword because it is the center of strong co-occurrence. As for "toilet", the connection of nursing care was seen mainly from "urination", "induction" that co-occurred with "toilet".

In Fig. 4 (b), looking at the word arrangement, "wheelchair", "hole", "night time", and "appearance" were almost in the center.

In the co-occurrence network, knowledge extraction was classified into five groups.

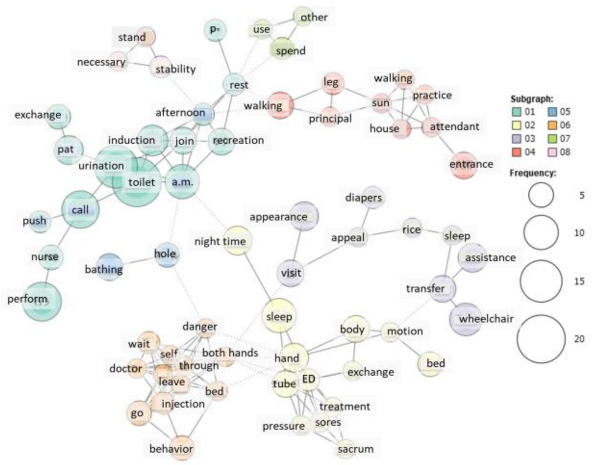
From the set of extracted words, group 1 was interpreted as "toilet", group 2 as "family", group 3 as "assistance", group 4 as "procedure" and group 5 as "motion".

In Fig. 5, focusing on the toilet in the cluster represented in the self-organizing map, "toilets", "guidance", "afternoon", "morning", "participation", "walking", "recreation", "rest", "pat", "exchange", "hall", "exercise", "walking", "call", "nurse" were formed in the same cluster.

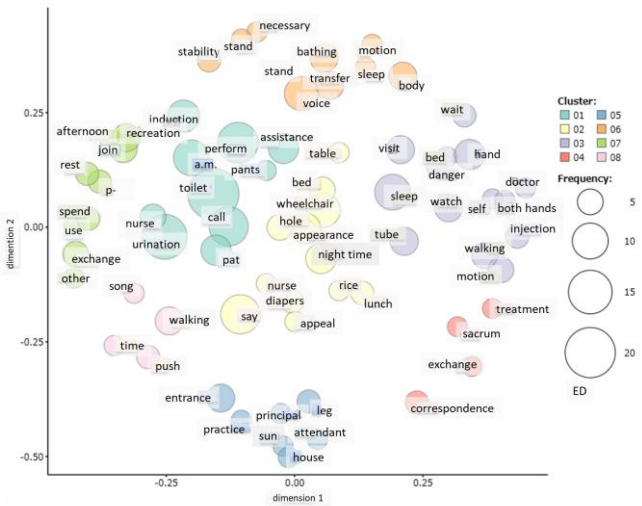
Fig. 6 shows a diagram in which similar words for executing cluster analysis are classified in a hierarchical structure, in order to hierarchically capture combinations of words having similar appearance patterns from the extracted words.

Table 1 Frequent occurring words

No	word	frequency	No	word	frequency
1	toilet	21	24	walking	6
2	urination	16	25	visit	6
3	perform	13	26	nurse	5
4	call	12	27	bed	5
5	say	12	28	hole	5
6	a.m.	10	29	exchange	5
7	sleep	10	30	go	5
8	wheelchair	9	31	behavior	5
9	voice	9	32	ED	4
10	induction	8	33	p-	4
11	pat	7	34	diapers	4
12	assistance	7	35	stability	4
13	hand	7	36	push	4
14	night time	7	37	spend	4
15	appearance	7	38	rest	4
16	tube	6	39	afternoon	4
17	recreation	6	40	wait	4
18	transfer	6	41	leg	4
19	watch	6	42	correspondence	4
20	join	6	43	lunch	4
21	body	6	44	injection	4
22	entrance	6	45	doctor	3
23	bathing	6			



(a)



(b)

Fig.4 Co-occurrence network

ISBN: 978-988-14048-5-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Hierarchical cluster analysis was performed with the minimum number of occurrences limited to 4 or more. "Induction", "toilet", and "urination" were included in the same cluster, and other things concerning the elderly home facilities regarding "wheelchair", "assistance" and "transfer" were also convincing. This indicates that the numerical value of the degree of care required is high. In the cluster analysis, knowledge extraction was classified into six clusters. From the clustering of knowledge extraction, cluster 1 was interpreted as "toilet", cluster 2 as "night time", cluster 3 as "bed", cluster 4 as "wheelchair", cluster 5 as "sleep", and cluster 6 as "motion".

VII. Consideration

The following is an overall evaluation.

Text data mining in general or data analysis of EMRs remains a relatively unexplored field. Greater collaboration between medical and information sectors will improve the technology so that it can be applied in clinical practice.

As a result of this research, extracted frequent words are the theme of this research.

Care records are mainly focused on basic vocabulary in nursing care. Although it is mere records and memorandums, it can be shared with other care givers, because it is described as a general natural language.

It is possible to interpret the state of nursing care by visualization, and the vocabulary extracted this time is valid for creating a nursing care dictionary.

By visualizing the relation of the extracted vocabulary, it shows the possibility of standardizing the nursing care recording method while characterizing the nursing care point.

Furthermore, from the present study, it was possible to suggest a direction to construct an electronic nursing care system of care records.

VIII. Conclusion

In this research, we attempted to examine the extraction of knowledge that the care worker recognizes, from the care life logs by using the text mining method.

The analysis results could contribute to the clarification of knowledge content of a wide range of care workers.

In future work, we further build up research on nursing care record analysis based on the analysis results clarified in this research, and build a record database.

Acknowledgment

The authors thank the members of the Medical Information Department at the University of Miyazaki Hospital.

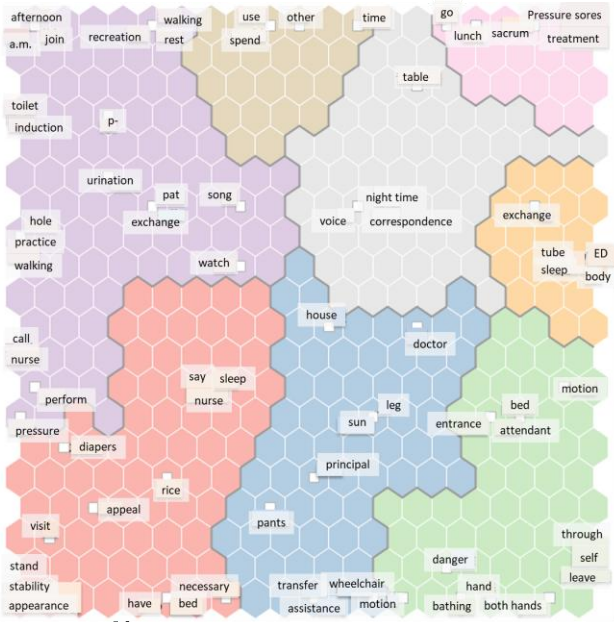


Fig. 5 Self-organizing map

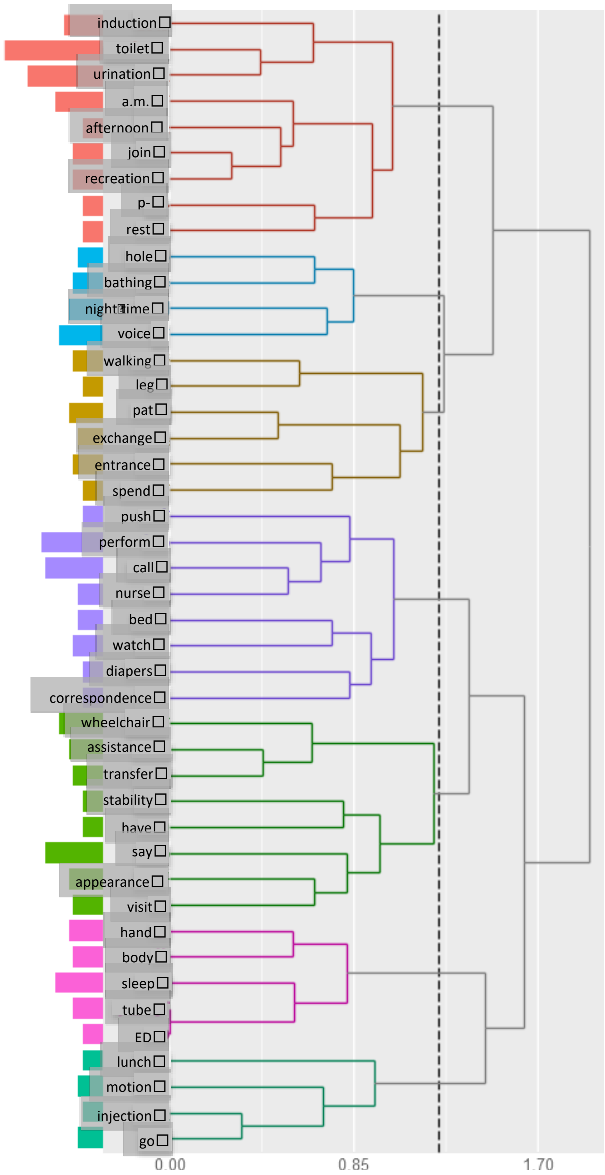


Fig. 6 Hierarchical cluster analysis

ISBN: 978-988-14048-5-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

References

[1] Information System," *Japan Association for medical Informatics*, vol. 22, no. 4, pp. 347-353, 2002.

[2] K. Yamamoto, S. Matsumoto, H. Matsuda, H. Tada, A. Matsuyama, K. Yanagihara, S. Teramukai and M. Fukushima, "Development and Prospects of Data Capture System for Clinical Study by the Secondary Use of Electronic Medical Records," *Japan Association for medical Informatics*, vol. 27, no. 2, pp. 211-218, 2007.

[3] Y. Matsumura, H. Nakano, H. Kusuoka, K. Park, M. Matsuoka, H. Oshima, M. Hayakawa and H. Takeda, "Clinic Hospital Cooperation System Based on The Network Type Electronic Patient Record," *Japan Association for medical Informatics*, vol. 22, no. 1, pp. 19-26, 2002.

[4] S. Murayama, K. Okuhara and H. Ishii, "Innovation in Manufacturing Premise by New Finding Obtained from Accident Relapse Prevention Report," *Proceedings of The 13th Asia Pacific Management Conference*, Melbourne, Australia, vol. 13, pp. 1124-1129, 2007.

[5] Y. Takahashi, K. Miyaki, T. Shimbo and T. Nakayama, "Text-mining with Network Analysis of News About Asbestos Panic," *Japan Association for medical Informatics*, vol. 27, no. 1, pp. 83-89, 2007.

[6] M. Usui and Y. Ohsawa, "Chance Discovery for Decision Consensus in Company - Touchable, Visible, and Sharable for a Textile Manufacture Company -," *Japan Society for Fuzzy Theory and Intelligent Informatics*, vol. 15, no. 3, pp. 275-285, 2004.

[7] H. Abe, S. Hirano and S. Tsumoto, "Mining a Clinical Course Knowledge from Summary Text," *The 19th Annual Conference of the Japanese Society for Artificial Intelligence*, vol. 1F4-04, pp. 1-2, 2005.

[8] Y. Kinosada, T. Umemoto, A. Inokuchi, K. Takeda and N. Inaoka, "Challenge to Quantitative Analysis for Clinical Processes by Using Mining Technology," *Data Science Journal*, vol. 26, no. 3, pp. 191-199, 2006.

[9] H. Ono, K. Takabayashi, T. Suzuki, H. Yokoi, A. Imiya and Y. Satomura, "Classification of Discharge Summaries by Text Mining," *Japan Association for medical Informatics*, vol. 24, no. 1, pp. 35-44, 2004.

[10] Y. Sato, H. Takeuchi K. Hoshi, N. Uramoto, T. Satoh, N. Inaoka, K. Takeda and N. Yamaguchi, "The Effectiveness of the Text Mining and Similar Document Search System for Evidence-Based Guideline Development," *Japan Association for medical Informatics*, vol. 24, no. 2, pp. 315-322, 2004.

[11] M. Kushima, K. Araki, M. Suzuki, S. Araki and T. Nikama, "Analysis and Visualization of In-patients' Nursing Record Using Text Mining Technique," *International MultiConference of Engineers and Computer Scientists 2011 (IMECS2011)*, vol.2188, pp. 436-441, Hong Kong, March, 2011.

[12] Tsukada, S. and Morita, T. 2018 "A Study of Public Opinion on Green Spaces Using Data from Free-Text Descriptive Responses- a Case Study of Regional City in Japan" International Journal of GEOMATE 14: 21–26.

ISBN: 978-988-14048-5-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)