Analysis and Visualization of In-patients' Nursing Record Using Text Mining Technique

Muneo Kushima, Kenji Araki, Muneou Suzuki, Sanae Araki, and Terue Nikama

***Abstract*—It is an important research assignment in the field of medical information to make effective use of the treatment information and administrative information that the care card manages. One example of this is the use of the text mining technique shown in this text. In this research, the in-patients' nursing record in the internal medicine, managed by the electronic medical record system of the University of Miyazaki Hospital, was first visualized by KeyGraph. As a result of this, a quantitative nursing record that used the text mining technique was able to be analyzed and an initial purpose of visualization was able to be achieved. The result of this research contributes to the work evaluation and the education of the nursing master.**

***Index Terms*—electronic medical record, text mining, nursing record, KeyGraph, visualization.**

I. Introduction

WHEN the medical information system [1] was updated on May, 2006, University of Miyazaki Hospital introduced the package version of the Electronic Medical Record (EMR) system, which was developed in collaboration with a local IT company. The EMR records information on patients not by paper cards, but by computers. The expected effect is to make the process of the entire hospital management easy to understand and to raise the quality of medical care. The recorded main data includes patients' symptoms, laboratory results, kinds of medicine prescribed, and the tracking of the data changed [2]-[3]. Doctors and nurses are not only able to share patients' information, but also coordinate it with arrangements of various inspections and the medical accounting system, etc.

Data mining [4] involves searching for correlation between items by analyzing a great deal of accumulated data, such as sales data and telephone call history. Text mining [5] is the same in that it aims to extract useful information by analysis from diversified viewpoints of written data. Recently, interest has risen for text mining that reveals useful knowledge buried in a large amount of accumulated documents. Research has started which applies text mining to medicine and the realm of healing [6]-[9]. In addition, electronic medical treatment data is increasing with an accelerated speed because of rapid informationization of the medical system, including the EMR. Recently, research on "Data mining in medical treatment", aiming at information and pattern extraction from a huge accumulated database, has been actively carried out. However, many medical documents, including the EMRs where the treatment information of patients is described, are text information. There is a problem with the complication of mining as well. The arrangement and the retrieval of data becomes difficult concerning these text parts because they are often described in a free format; words and phrases and the expressions used are different depending on the writers. It is possible that text mining of documents will be used for a lateral retrieval in the world of medical treatment in the future by not only the numerical value of the inspection data, but also the computerization of documents.

In the present study, in-patients' nursing records are chosen from among the nursing records preserved by the EMR system of University of Miyazaki Hospital. Sentences are analyzed in morphemes, relations between feature vocabularies are analyzed by using KeyGraph, and visualization of this information has been attempted.

Manuscript received December 8, 2010; revised January 10, 2011. M. Kushima is a doctral candidate in the faculty of medicine at the University of Miyazaki Hospital. e-mail: kushima@fc.miyazaki-u.ac.jp, (see http://mit.med.miyazaki-u.ac.jp/).

K.Araki, M.Suzuki S.Araki and T.Nikama are with the Faculty of Medicine, University of Miyazaki Hospital. 5200, Kihara, Kiyotake-cho, Miyazaki-shi 889-1692 Japan. Tel: +81-985-85-9057, Fax: +81-985-84-2549.

II. Nursing record

The text data in the EMR consists of paper passages concerning the inspection report, the in-patient care plan, the nutrition management plan, bedsore plan, the fall check, the operative note, and summaries. The doctor fills in the passage record and the nurses fill in the nursing record. In the nursing record, life history and inspection history of a patient are written. In addition, the nursing record has small notes about reservations, etc. There are no rules about text recording and, as a result, ambiguous feelings or impressions are sometimes written down. Nurses remember or take notes about what patients say during their busy schedules and later input it collectively to the EMR. There are 4 forms of recording: S (Subjective data) writes about subjective information, O (Objective data) about objective information, A (Assessment) about assessment, P (Plan) about future planning.

- S (Subjective data) = Information a patient directly offers or words a patient says.
- O (Objective data) = Objective facts, appearance, or state of a patient through the eyes of co-medicals.
- A (Assessment) = Evaluation and judgment derived from this information.
- P (Plan) = Plan in the future and care actually taken.

III. Text mining application to medicine

Text mining is a technique often used to analyze information hidden in the text of a document, to extract key words, phrases and even concepts from a written document. Text mining, sometimes alternately referred to as text mining, is roughly equivalent to text analytics, refering to the process of deriving high-quality information from text. Text mining usually involves the process of structuring the input text (usually parsing, along with the addition of some derived linguistic features and the removal of others, and subsequent insertion

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

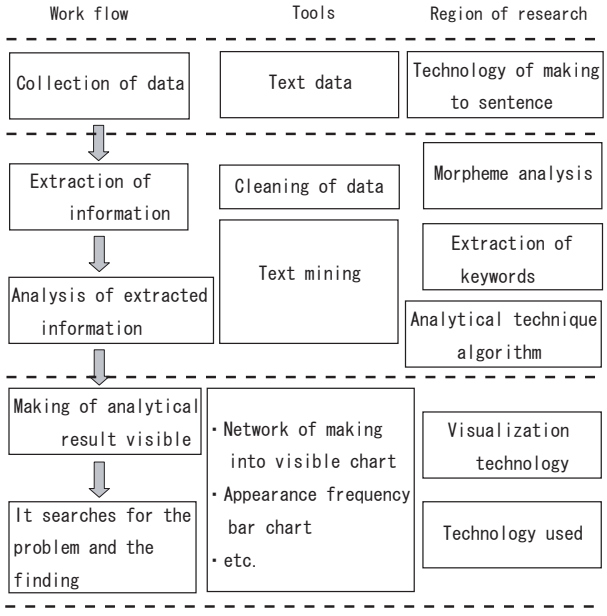

Fig. 1. Process of text mining.

into a database), deriving patterns within the structured data, and, finally evaluation and interpretation of the output. Figure 1 shows the process of text mining.

The following points are paid attention to when text mining is applied to medicine.

- In medical treatment, it is to be used in the understanding of the importance of rare events such as side effects from medication and arrhythmia.
- To obtain final decisions about the courses of treatment.

The entire process of identifying symptoms or understanding the associated risk, while taking appropriate action, can be a problem.

IV. Supporting tools

Tools were selected in order to apply text mining techniques and are as follows:

- Chasen is applied for morphologic analysis [10]-[11]. Chasen is a morphological parser for the Japanese language. This tool for analyzing morphemes was developed at Matsumoto laboratory, NAIST (NAra Institute of Science and Technology).
- KeyGraph in Polaris is applied for extracting key words [12]-[14].

V. KeyGraph

Figure 2 shows an image graph of KeyGraph. Figure 3 shows an example of KeyGraph when it is applied to the text data. KeyGraph is a graph-making method that uses the frequency in the use of a word and the co-occurrence among words, clarifies the important relationships among them, and extracts key factors from them.

A. The Algorithm of KeyGraph

KeyGraph, originally an algorithm for extracting terms (words or phrases), expresses assertions based on the co-occurrence graph of terms from text data. The strategy of KeyGraph comes from considering that a document is

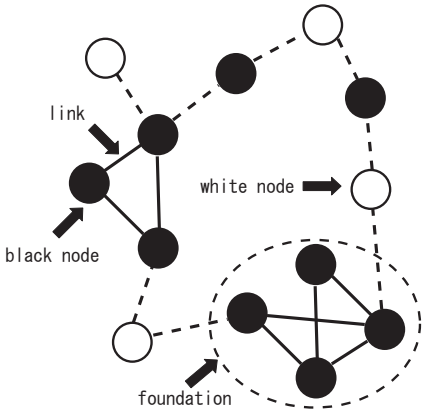

Fig. 2. Image graph of KeyGraph.

constructed like a building for expressing new ideas based on traditional concepts as follows:

A building has *foundations* (statements for preparing basic concepts), walls, doors and windows (ornamentation). But the *roofs* (main ideas in the document), without which the building's inhabitants cannot be protected against rain or sunshine, are the most important. These *roofs* are supported by *columns*. Simply put, KeyGraph finds the *roofs*.

The process of KeyGraph consists of four phases:

1) Document preparation: Prior to processing a document *D*, stop words that have little meaning and are discarded from *D*, words in *D* are stemmed, and phrases in *D* are identified. Hereafter, a *term* means a word or a phrase in processed *D*.

2) Extracting foundations: Graph *G* for document *D* is made of nodes representing terms and links representing their *co-occurrence* (term-pairs which frequently occur in same sentences throughout *D*). Nodes and links in *G* are defined as follows:

a)Nodes: Nodes in *G* represent high-frequency terms in *D* because terms might appear frequently for expressing typical basic concepts in the domain. High frequency terms are the set of terms above the highest frequency (black nodes). We denote this set by *HF*.

b)Links: Nodes in *HF* are linked if the association between the corresponding terms is strong. The association of terms w_i and w_j in *D* is defined as:

$$assoc(w_i, w_j) = \sum_{s \in D} min(|w_i|_s, |w_j|_s), \qquad (1)$$

where $|w|_s$ denotes the count of w in sentence s. Pairs of high-frequency terms in *HF* are sorted by *assoc* and the pairs above the (*number of nodes in G*) -1 th tightest association are represented in *G* by links between nodes (solid lines). Then, each cluster (called a foundation) is obtained as a set of nodes and links forming a connected graph.

c)Extracting columns: The probability of term w to appear near clusters is defined as *key (w)*, and the *key (w)* is defined by

$$key(w) = 1 - \prod_{g \subset G} \left[1 - \frac{based(w, g)}{neighbors(g)} \right], \qquad (2)$$

$$based(w, g) = \sum_{s \in D} |w|_s |g - w|_s, \qquad (3)$$

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$$neighbors(g) = \sum_{s \in D} \sum_{w \in s} |w|_s |g - w|_s. \quad (4)$$

$$|g - w|_s = \begin{cases} |g|_s - |w|_s, & w \in g \\ |g|_s, & w \notin g \end{cases} \quad (5)$$

where *g* represents each cluster in *G*. Sorting terms in *D* by *key* produces a list of terms ranked by their association with clusters and the several top *key* terms are taken for *high key terms*.

d)Extracting roofs: The strength of the column between a *high key term* w_i and a high frequency term $w_j \subset HF$ is expressed as:

$$column(w_i, w_j) = \sum_{s \in D} min(|w_i|_s, |w_j|_s). \quad (6)$$

Columns touching w_i are sorted by *column*(w_i, w_j) for *each high key term* w_j. Columns with the highest *column* values are selected to create new links in *G*. We depict such links, representing columns, by dotted lines. Then, each term w_i is connected by these attached columns to terms in two or more clusters. Finally, nodes in G are sorted by the sum of *column* values of its touching columns. Terms represented by nodes of higher values of these sums than a certain threshold are extracted as the keywords for document D, as depicted by node of term (*Roof*).

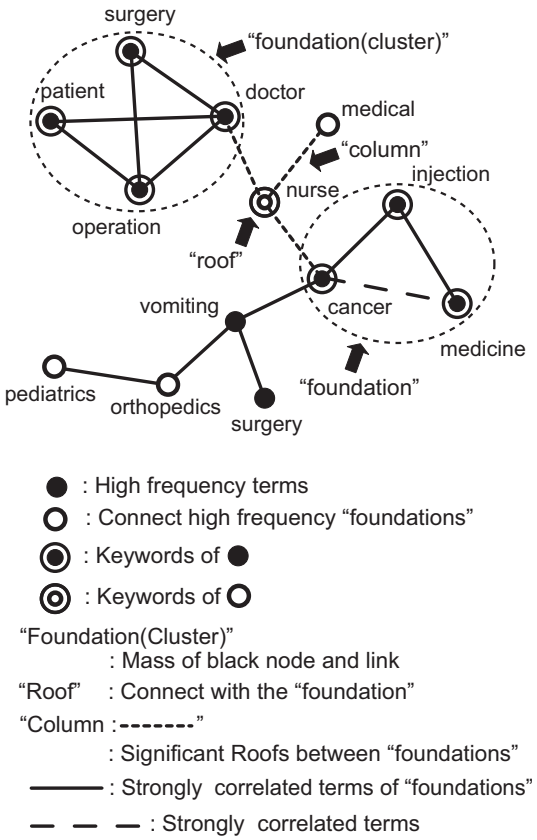

Fig. 3. An example of KeyGraph when it is applied to the text data.

VI. Result of analysis

In this paper, the in-patients' nursing record of University of Miyazaki Hospital was assumed to be data and the nursing record of June 2007 was used. Here we made an analysis in 3 different cases (Internal Medicine 1-2-3). The analysis results were shown as follows:

- Internal Medicine 1 : Firstly, Figure 4 shows a result of KeyGraph co-occurrence network chart when it is applied to the text data of the internal medicine 1. "Foundations" are obtained from text data, each including an event-set { Patient, Consent, Setup, Wearing, Running fire, Above, Word, Electrocardiogram }. The internal medicine 1 has the most patients with circulatory organ problems; therefore, doctors frequently ask patients for "consent" to administer specific tests because there are many inspections such as, the cardiac catheter tests and endoscopy, etc. Moreover, there is an emphasis on whether to transmit the instruction from the doctor to the nurse or not.
- Internal Medicine 2 : Secondly, Figure 5 shows a result of KeyGraph co-occurrence network chart when it is applied to the text data of the internal medicine 2. "Foundations" are obtained from text data, each including an event-set { Start, Today, Eye, Internal use }, { Progress, Connection, Medication, Feeling, Urine, Defect }. The internal medicine 2 has a lot of serious cases such as blood disease and liver cancers, and is a department where anti-cancer drug administration or chemotherapy is frequently performed. There are many terms related to the confirmation of medicine being given or taken and there are many terms related to the medicine's side effects.
- Internal Medicine 3 : Thirdly, Figure 6 shows a result of KeyGraph of co-occurrence network chart when it is applied to the text data of the internal medicine 3. "Foundations" are obtained from text data, each including an event-set { Charge, Sickbed, Skin, Perspiration, Anesthsia, Plan, Whole body, Continuation, Scab }. There are a lot of cases with respiratory illness, diseases of the nervous system, and there are a lot of bedridden patients; therefore, there are a lot of terms such as "anesthesia", etc.

VII. Consideration

This paper shows a greater possibility of a disease being automatically specified and classified from documents used at the medical treatment site. In the future, the use of the text mining approach and processing of medical documents laterally will support disease classification, retrieving examples of similar syndromes, etc. It can also be applied to the discovery of new medical knowledge which enables new syndrome detection. Text mining is expected to become a valuable technique in the analysis of medical documents in the future. In this way, text mining techniques are expected to be applied to all medical documents in the future and help each field of research make favorable progress.

VIII. Conclusion

In this paper, the in-patients' nursing record at the EMR in the University of Miyazaki Hospital was assumed to be a theme and the analysis that used the text mining technique was done. Sentences were analyzed in morphemes and the relations between characteristic vocabularies were analyzed by using KeyGraph. Visualization of this information has been attempted. As a result, a history for the diagnosis and

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

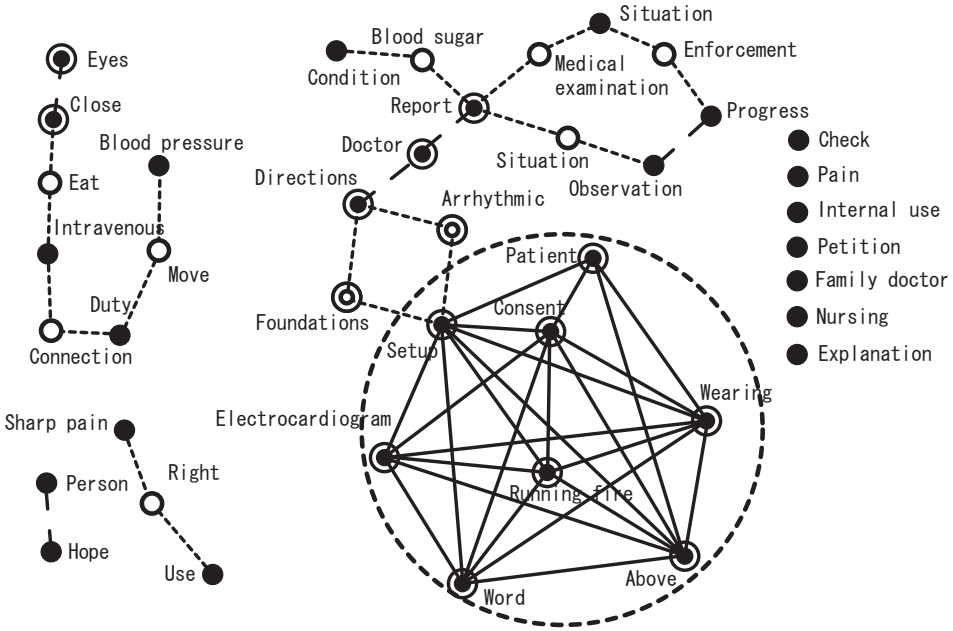

Fig. 4. A result of KeyGraph, for internal medicine 1 of University of Miyazaki Hospital, (nursing record of June 2007 was used). Parameter of KeyGraph : JaJa[30-30-10-15]

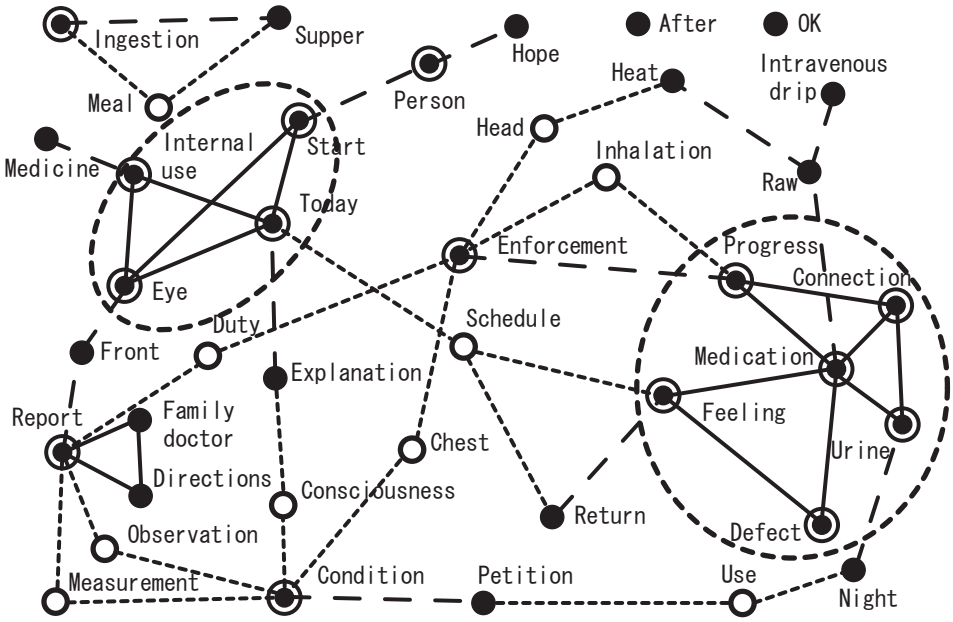

Fig. 5. A result of KeyGraph, for internal medicine 2 of University of Miyazaki Hospital, (nursing record of June 2007 was used). Parameter of KeyGraph : JaJa[30-30-10-15]

treatment department was visualized. Along with years of experience so far, the resultant visualization based on the relevant jargon, was obtained and compared with the result of the nursing record that had been reported. The result was an analysis of qualitative in-patients' nursing records using a text mining technique and the initial goal was achieved: a visual record of this information. In addition, this enabled the discovery of vocabularies relating to proper methods of treatment, resulting in a concise summary of the vocabularies extracted from the in-patients' nursing record. Important vocabularies characterizing each nursing record were also revealed. We intend to accumulate the clinical research data that evaluates the safety, etc., of the prognosis, the prognostic factor, treatment results, and the safety of Medical Technologies. This information will then be related to cost reduction, efficiency improvement, and the quality improvement of clinical research in the future.

IX. Future work

In the future, we will pay attention to the passage records that the doctors file in order to improve the analytical result. Moreover, we think that we can obtain further results by analyzing other diagnoses in other treatment departments, and comparing them all.

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

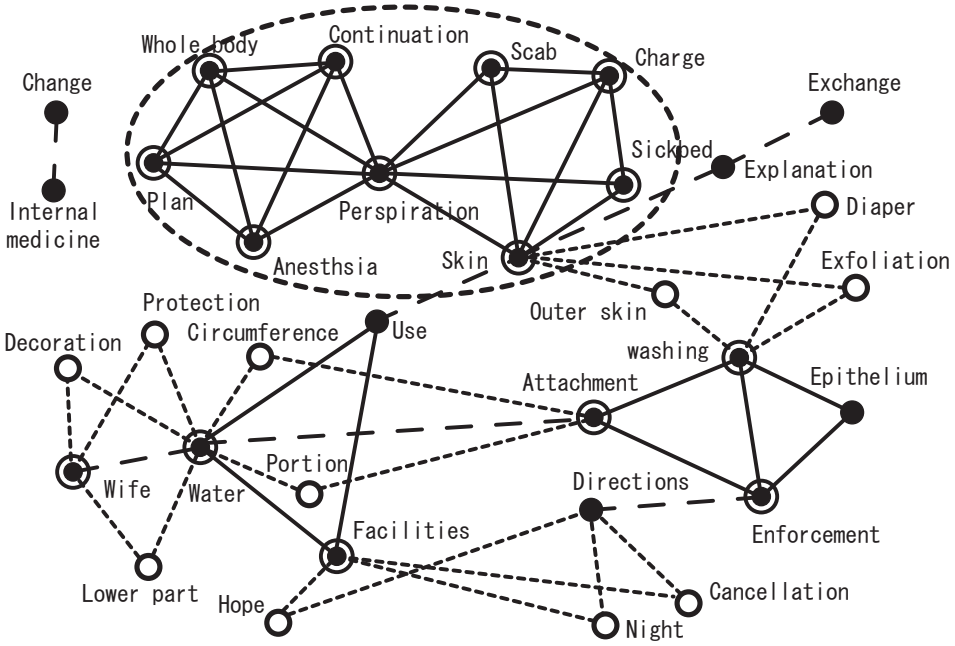

Fig. 6. A result of KeyGraph, for internal medicine 3 of University of Miyazaki Hospital, (nursing record of June 2007 was used). Parameter of KeyGraph : JaJa[30-30-10-15]

TABLE I
SUBJECT WORD

Fig.	Subject word
4	Eyes, Close, Report Patient, Setup, Consent Running fire, Above Directions, Wearing Electrocardiogram Doctor, Word
5	Ingestion, Person, Start Enforcement, Progress Eye, Medication, Feeling Report, Defect, Condition Internal use, Connection Today, Urine
6	Whole body, Continuation Anesthsia, Perspiration Washing, Attachment, Wife Enforcement, Facilities Water, Scab, Plan Charge, Sickbed, Skin

TABLE II
FREQUENT OCCURRENCE KEY WORD

Order of Frequent	Fig.4	Fig.5	Fig.6
1	Doctor	Internal use	Expla -nation
2	Direc -tions	Medication	Charge
3	Person	Connection	Sickbed
4	Patient	Urine	Scab
5	Internal use	Feeling	Skin
6	Report	Today	Today
7	Hope	Person	Use
8	Consent	Family doctor	Whole body
9	Obser -vation	Progress	Plan
10	Use	Night	Exchange
11	Petition	Hope	Change
12	Condition	Eye	Internal medicine
13	Blood pressure	Report	Perspir -ration
14	Pain	Condition	Facilities
15	Explanation	OK	Epithelium
16	Electro -cardiogram	Heat	Washing
17	Progress	Front	Construc -tion
18	Family doctor	Intrave -nouse	Continu -ation
19	Nursing	Start	Wife
20	Intrave -nouse	Petition	Water

ACKNOWLEDGMENT

The authors would like to thank members of the Medical Information Department, Faculty of Medicine, University of Miyazaki Hospital.

REFERENCES

[1] M, Kushima., K, Araki., M, Suzuki., S, Araki., and T, Nikama., “Graphic Visualization of the Co-occurrence Analysis Network of Lung Cancer in-patient nursing record,” *The International Conference on Information Science and Applications (ICISA 2010)*, Seoul, Korea, pp. 686–693, April 2010.

[2] A, Abe., H.I, Ozaku., N, Kuwahara., and K, Kogure., “Scenario Violation in Nursing Activities - Nursing Risk Management from the viewpoint of Chance Discovery - ,” *Soft Computing Journal*, vol.11, no.8, pp.799–809, 2007.

[3] A, Abe., H.I, Ozaku., K, Sagara., N, Kuwahara., and K, Kogure., “Nursing Risk Management by Forcusing on Critical Words or Phrases in Nurses’ Conversations,” *Int’l J. of Knowledge-based and Intelligent Engineering Systems*, vol.11, no.5, pp.281-289, 2007.

[4] A, Abe., N, Hagita., M, Furutani., Y, Furutani., and R, Matsuoka., “Possibility of Integrated Data Mining of Clinical Data,” *Data Science Journal*, Vol.6, 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, Kushima., K, Araki., M, Suzuki., S, Araki., and T, Nikama., “Graphic Visualization of the Co-occurrence Analysis Network of Lung Cancer in-patient nursing record,” *The International Conference on*

Information Science and Applications (ICISA 2010), Seoul, Korea, pp. 686–693, 4/2010.

[7] M, Kushima., K, Araki., M, Suzuki., S, Araki, H.Tamura, K.Tanno, T.Toyama, O.Ishizuka and M.Ikeda, "Analysis of Nurse's Medical Record in The IZANAMI Using Text Mining Method," *18th International Workshop on Post-Binary ULSI Systems*, Okinawa, Japan, 5/2009.

[8] M, Kushima., K, Araki., M, Suzuki., S, Araki., H, Tamura., K, Tanno., T, Toyama., O, Ishizuka., and M, Ikeda., "Visualization of EMR by Using Mining Technology," *Proceedings of the 2009 IEICE General Conference*, D-4-6, 3/2009.

[9] M, Kushima., K, Araki., M, Suzuki., S, Araki., H, Tamura., K, Tanno., T, Toyama., O, Ishizuka., and M, Ikeda., "Network Visualization of Electronical Medical Record(IZANAMI) by Using Mining Technology," *Technical Report Of IEICE*, vol. 108, no. 388, CAS2008-97, pp. 187–192, 2009.

[10] Chasen web site. Availaible on: http://chasen-legacy.sourceforge.jp/.

[11] Y, Matsumoto., A, Kitauchi., T, Yamashita., Y, Hirano., H, Matsuda., K, Takaoka, and M, Asahara., "Morphological Analysis System Chasen Version 2.2.1 Manual," *Nara Institute of Science and Technology*, 2000.

[12] Y, Ohsawa., Nels E, Benson., and M, Yachida., "KeyGraph: Automatic Indexing by Segmenting and Unifing Co-occurrence Graphs," *The Institute of Electronics, Information and Communication Engineers*, vol.J82–D–I, no.2, pp.391-400, 2/1998.

[13] N, Okazaki., and Y, Ohsawa., "Polaris: An Integrated Data Miner for Chance Discovery," *In Proc. of Workshop of Chance Discovery and Its Management (in conjunction with International Human Computer Interaction Conference (HCI2003)*, Crete, Greece, 1/2003.

[14] A, Abe., Y, Ohsawa., H, Kosaku., K, Aira., N,Kuwahara., and K, Kogure., "Discovery of Hidden Factors for Incident by Kamishibai KeyGraph," *The 22nd Annual Conference of the Japanese Society for Artificial Intelligence*, vol.2B2-3, pp.1-4, 2008.

ISBN: 978-988-18210-3-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)