# Text Data Mining of In-patient Nursing Records Within Electronic Medical Records Using KeyGraph

Muneo Kushima, *Member, IAENG*, Kenji Araki, Muneou Suzuki, Sanae Araki and Terue Nikama

***Abstract*—This research used a text data mining technique to extract useful information from nursing records within Electronic Medical Records. Although nursing records provide a complete account of a patient's information, they are not being fully utilized. Such relevant information as laboratory results and remarks made by doctors and nurses is not always considered. Knowledge concerning the condition and treatment of patients has been determined in a twofold manner: a text data mining technique identified the relations between feature vocabularies seen in past in-patient records accumulated on the University of Miyazaki Hospital's Electronic Medical Record, and extractions were made. The qualitative analysis result of in-patient nursing records used a text data mining technique to achieve the initial goal: a visual record of such information. The analysis discovered vocabularies relating to proper treatment methods and concisely summarized their extracts from in-patient nursing records. Important vocabularies that characterize each nursing record were also revealed. The results of this research will contribute to nursing work evaluation and education.**

***Index Terms*—text data mining, nursing records, electronic medical records, visualization**

## I. Introduction

AN Electronic Medical Record (EMR) records information on patients by computers instead of by paper. 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 and compared with medical records on paper [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, interest has risen in text data mining because it uncovers useful knowledge buried in a large amount of accumulated documents [7],[8]. Research has started to apply text data mining to medicine and healing [9],[10],[11],[12],[13]. 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 [14]. Perhaps 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 [15],[16]. Nursing records, which are listed as course text information, account for almost the entire record. In this study, we chose in-patient nursing records from among nursing records preserved by the EMR system at the University of Miyazaki Hospital. Sentences were analyzed into morphemes, and the relations among feature vocabularies were analyzed using KeyGraph [17],[18],[19],[20]. Then we visualized this information.

Manuscript received February 12, 2011; revised August 10, 2011.

M. Kushima is a doctoral candidate in the Faculty of Medicine at the University of Miyazaki Hospital.

e-mail: kushima@fc.miyazaki-u.ac.jp, (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. IZANAMI (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 (IZANAMI), which was developed in collaboration with a local IT company. The recorded main data include a patient's symptoms, laboratory results, prescribed medicines, and the tracking of the changed data. Cases that make both the images of X-rays and the appended material electronic 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.

IZANAMI has a unique feature that is different from those being operated at many other university hospitals [21],[22],[23].

First, the electronic card systems used so far in university hospitals were all developed by major medical system venders, but IZANAMI 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

(Advance online publication: 24 August 2011)

Hospital after it was incorporated. We made the medical staff concretely 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. In the several years since IZANAMI was introduced, there have been no big problems or confusion involved in its operation. IZANAMI has received high praise from doctors and other medical personnel and has also attracted many visitors from outside the hospital.

Figure 1 shows an organizational chart of the 18 clinical departments of the University of Miyazaki Hospital.

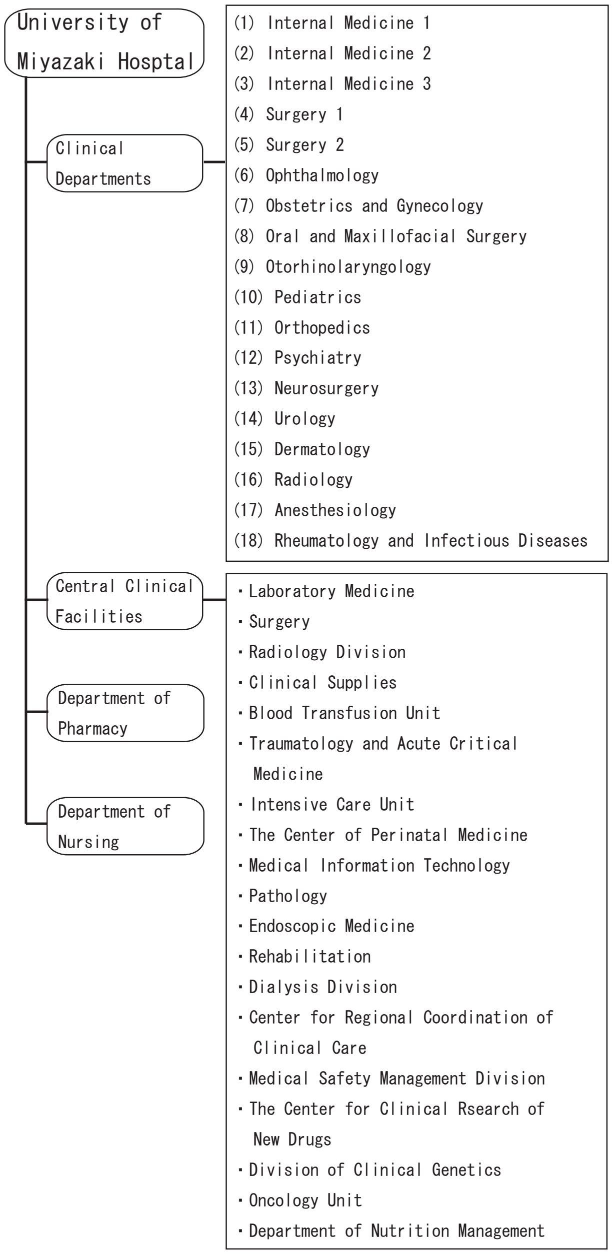


Fig. 1. Organization chart of University of Miyazaki Hospital

## III. Nursing records

The text data in EMR consist of paper notations about inspection reports, in-patient care plans, nutrition management plans, bedsore-prevention plans, fall checks, operation notes, and summaries. The doctor fills in the passage record and the nurses fill in the nursing records, which include the life and inspection history of a patient [24]. The nursing record also has small notes about reservations etc. Since no guidelines exist about recording text, ambiguous feelings or impressions are sometimes included. Nurses remember or take notes about what their patients say while working and later input them into the EMR. There are four recording categories: subjective data (S), objective data (O), assessment (A), and plans (P).

- S = information directly gleaned from patients
- O = objective facts and observations about the patient appearance or state by co-medicals
- A = evaluations and judgments derived from this information
- P = future plans and care actually taken

The nursing record, which records the care activities practiced by nurses, contains many notes about nursing processes.

It helps ensure high quality nursing and evaluates nursing practices. It is also used to calculate treatment fees. By law, nursing records must be retained for at least two years. The following are some features of nursing records:

- Must be suitable for offering treatment information and the release of diagnosis and treatment records.
- Must show awareness of risk management.
- Their quality and efficiency must be standardized.
- Must maximize time of patient care and minimize recording time.
- Must be available for future treatment.

## IV. Text data mining applications 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 insertion into a database), deriving patterns within the structured data, and finally evaluating and interpreting 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.

First, such rare events as side effects from medication and arrhythmia can be highlighted to improve understanding of their occurrences.

Second, final decisions can be obtained regarding courses of treatment. One difficulty with applying text data mining to medicine is the entire process of identifying symptoms for understanding the associated risks while taking appropriate action.

## V. KeyGraph

We applied KeyGraph to the text data mining technique [25],[26],[27],[28]. We also applied it for extracting key words.

### A. Example of KeyGraph performance

Figure 3 shows an image of KeyGraph. Figure 4 shows an example when it is applied to text data.

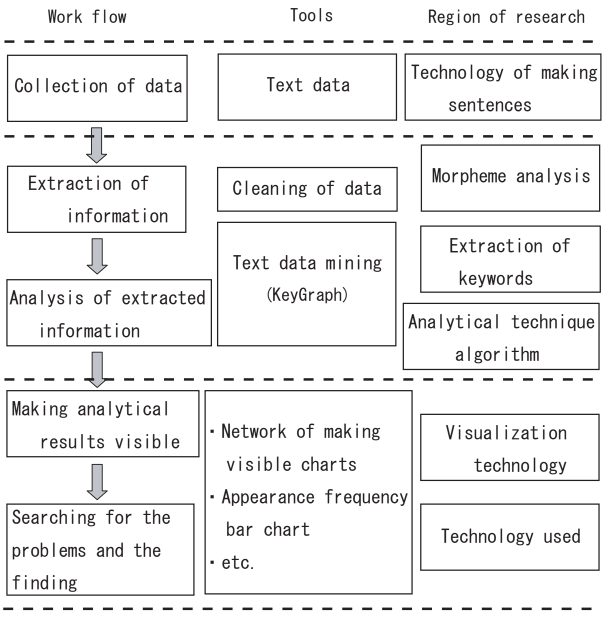


Fig. 2. Process of text data mining

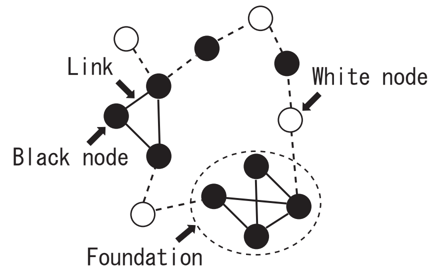


Fig. 3. Image graph of KeyGraph

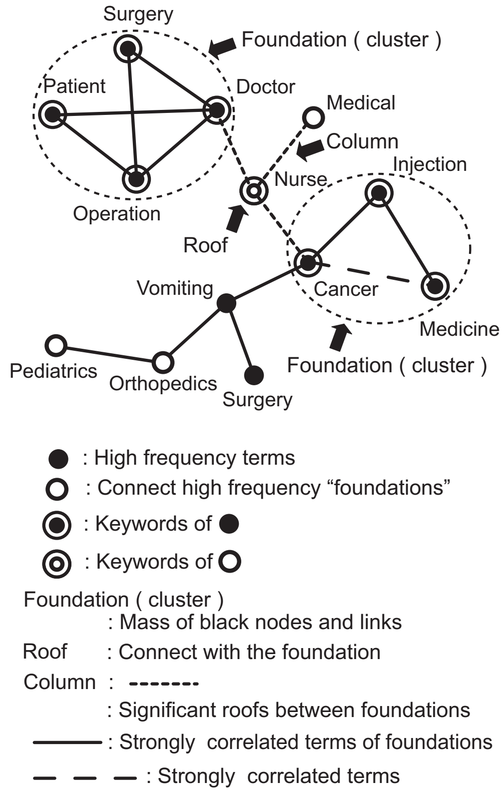


Fig. 4. KeyGraph example when applied to text data

- Black nodes indicate items that frequently occur in a data set.
- White nodes indicate the items that occur less frequently overall but frequently occur with black nodes in a data set.
- Double-circled nodes indicate items whose co-occurrence frequency with black nodes is especially high. Double-circled nodes are considered keywords.
- Links indicate that the connected item pair frequently co-occurs in a data set. Solid lines form a foundation, which dotted lines connect.

Foundations, which are circles of dotted lines, are obtained from the text data. In Fig. 4, two foundations have strong linkages with event-sets:
{doctor, surgery, patient, operation}, and {cancer, medicine, injection}.

### B. Outline of KeyGraph

Instead of giving a detailed explanation of KeyGraph, we briefly outline it here. KeyGraph consists of three major components derived from building construction metaphors. Each component is described as follows:

1) Foundations: sub-graphs of highly associated and frequent terms that represent basic concepts in the data. A foundation is defined as a cluster that consists of black nodes linked by solid lines. The foundations are underlying common contexts because they are formed by a set of items that frequently co-occur in the data set.
2) Roofs: terms that are highly associated with foundations.
3) Columns: associations between foundations and roofs that are used for extracting keywords, i.e., the main concepts in the data. A column is a dotted line that connects foundations. Although the common context represented by a foundation is widely known, the context represented by a column is not. Columns are important because they connect two common contexts in items that do not frequently occur.

### C. KeyGraph Algorithm

KeyGraph was originally an algorithm for extracting assertions based on the co-occurrence graph of terms from text data.

Its process consists of four phases:

*1) Document preparation:* Before processing document *D*, stop words [29] with little meaning are discarded from *D*, the words in *D* are stemmed [30], and the phrases in *D* are specified [31]. 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 the *co-occurrence* (term-pairs that frequently occur in the same sentences throughout *D*). Nodes and links in *G* are defined as follows:

- Nodes: nodes in *G* represent high-frequency terms in *D* because they might appear frequently for expressing typical, basic concepts in the domain. High-frequency terms are a set of terms above the 30th highest frequency. We denote this set by *HF*.

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

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

where $|x|_s$ denotes the count of $x$ in sentence $s$. Pairs of high-frequency terms in *HF* are sorted by *assoc* and the pairs above the (*number of nodes in G*) -1th tightest association are represented in *G* by links between nodes.

*3) Extracting columns:* The probability that term $w$ appears is defined as *key (w)*, and *key (w)* is defined by

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

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

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

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

where *g* represents each cluster in *G*. Sorting the terms in *D* by *keys* produces a list of terms ranked by their association with *clusters*, and the 12 top *key* terms are considered *high key terms*.

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

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

Columns touching $w_i$ are sorted by *column(*$w_i$*,* $w_j$*)* for each *high key term* $w_i$. Columns with the highest *column* values connecting term $w_i$ to two or more clusters are selected to create new links in *G*. Finally, the nodes in $G$ are sorted by the sum of the *column* of the touched columns. Terms represented by the nodes of higher values than a certain threshold are extracted as the keywords for document $D$.

## VI. Analysis results

In this paper, EMR data were collected from all the departments of the Faculty of Medicine at the University of Miyazaki Hospital, and the nursing records from June 2007 were used. The following analysis results are shown:

1) Internal Medicine 1 (Fig. 5) [Cardiovascular, Kidney, Gastrointestinal, General Medicine]: Most of the patients in Internal Medicine suffer from circulatory organ problems. Doctors frequently seek consent from patients before administering specific tests such as cardiac catheter tests, endoscopy, and so forth. In addition, the efficient transmission of instructions from doctors to nurses is emphasized. The foundation is obtained from the text data with event-set {consent, patient, mounting, above, running continuously, ward, electrocardiogram, setting}.
2) Internal Medicine 2 (Fig. 6) [Gastroenterology, Hepatology, Hematology, Oncology, General Medicine]: In this department, most patients suffer from serious blood disease or liver cancer, and anti-cancer drug administration or chemotherapy is frequently performed. Many terms are related to the confirmation of medicine being given/taken or side effects. The foundations are obtained from the text data with event-sets {start, today, sleep, internal use} and {progress, connection, urine, bad, medication, feeling}.
3) Internal Medicine 3 (Fig. 7) [Neurology, Respirology, Endocrinology, Metabolism, General Medicine]: Here many cases involve respiratory illness and diseases of the nervous system. Many patients are bedridden. Such terms as reddening and anesthesia are common. The foundation is obtained from the text data with event-set {reddening, anesthesia, plan, entire body, continue, acne, charge, ulcer, skin}.
4) Surgery 1 (Fig. 8) [Surgery of Digestive Organs, General Surgery (liver, bile duct, pancreas, esophagus, stomach, colorectum, anus, breast, thyroid)]: Since the main subject is abdominal surgery, many words and phrases address post-operative management. The foundation is obtained from the text data with event-set {reddening, anesthesia, continuation, ulcer, skin, charge, acne, plan, entire body}.
5) Surgery 2 (Fig. 9) [Cardiovascular Surgery, Thoracic Surgery, Gastroenterological Surgery, Endocrinological Surgery, General Surgery]: The main object of concern is surgery of the circulatory organs. This department ’ s procedures sometimes last a long time, and there are many words and phrases about informed consent. The foundation is obtained from the text data with event-set {patient, nursing, description, principal, hope, internal use, sleep}.
6) Ophthalmology (Fig. 10): There are many words and phrases about the conditions of patients after ophthalmic surgery. The foundation is obtained from the text data with event-set {surgery, eye, trial, return, left, ingestion, rest, terminology}.
7) Obstetrics and Gynecology (Fig. 11): Many words and phrases concerning uterus and pregnancy are particular to obstetrics and gynecology, especially terms concerning cooperation with doctors. The foundations are obtained from the text data with event-sets {uterus, contraction, report, installing, doctor} and {sore, above, internal use, complaint, stop}.
8) Oral and Maxillofacial Surgery (Fig. 12): Many words and phrases address post-operative care and conferences. The foundations are obtained from the text data with event-sets {moisture, treasurer, attachment, reference, leakage, blood vessel, punctures, admission, connection} and {time, attendance, evaluation, plan, content}.
9) Otorhinolaryngology (Fig. 13): Otorhinolaryngology is a particularly difficult medical field. The diagnoses of diseases in the ears, nose, or throat are generally obtained through videography and photography. Therefore, the nursing record is primarily composed of images rather than text. Many records indicate only bed soreness instead of symptoms that are influenced by a

particular disease. The foundations are obtained from the text data with event-sets {neck, right, destruction, pressure, shade} and {reddening, continue, ulcer, acne, charge}.

10) Pediatrics (Fig. 14): Many words and phrases are related to family because children are naturally visited by their families in the pediatrics department. Many words and phrases like understanding or description are found because the patients are so young. Many references are related to cancers and tumors because the pediatrics department deals with many malignant tumor cases, especially cancer and leukemia. The foundation is obtained from the text data with event-set {understanding, description, medical treatment, bone, leukemia, bleeding, cells, family}.
11) Orthopedics (Fig. 15): The foundation regarding inspections in the orthopedic department can be seen. The terms in double circles are concerned with explanations about operations and post-operative care. The foundation is obtained from the text data with event-set {description, necessary, plan, reddening, ulcer, acne, entire body, charge}.
12) Psychiatry (Fig. 16): Patients sometimes suffer from psychiatric problems, and most of the terms confirm the patients' mental state in the room. The foundation is obtained from the text data with event-set {internal use, blood sugar, check, night, sleep, then, inspection tour, state, breathing, my room}.
13) Neurosurgery (Fig. 17): A doctor's analysis, especially opinions about the nervous system of a patient, comprises the main words and phrases in neurosurgery. The foundation is obtained from the text data with event-set {progress, clear, paralysis, limbs, pupils, urine, reference, consciousness, direction}.
14) Urology (Fig. 18): Many records regard treatment plans after being discharged (where to continue treatment) because patients are usually hospitalized for a short period. The foundation is obtained from the text data with event-set {leaving hospital, caution, visits, clinic, urinary organs, contact, hospital}.
15) Dermatology (Fig. 19): Many terms describe pain or soreness because dermatology deals with such painful conditions as burns. The foundation is obtained from the text data with event-set {pain, soreness, internal use, hope, wounds, enhancement}.
16) Radiology (Fig. 20): Many words and phrases concern intravenous drip infusion and blood vessel preservation. The foundations are obtained from the text data with event-sets {connection, rest, blood vessel, tip, punctures, left hand, leakage} and {attendance, time, content, evaluation}.
17) Anesthesiology (Fig. 21): The pain clinic is the main subject in anesthesiology. Many words and phrases regard catheter insertion, fixation, and the management of infections. The foundations are obtained from the text data with event-sets {rest, infusion, fixation, punctures, contamination, connection} and {content, attendance, time, evaluation, plan, nursing}.
18) Rheumatology and Infectious Diseases (Fig. 22): Many words and phrases are concerned with the respiratory organs because collagen diseases arise from troubles with respiratory organs. The foundation is obtained from the text data with event-set {progress, mounting, urine, breathing, inhalation, oxygen, decline}.

Table I shows the relationship between the top ten words and their frequency in the University of Miyazaki Hospital in-patient nursing records. For example, in Internal Medicine 1, doctor, the highest occurring word, appeared 167 times. Direction, principal, patient, internal use, report, hope, consent, observation, and use also appear often.

## VII. Consideration

The following is an overall evaluation of this research:

- Due to KeyGraph analysis, characteristic keywords were extracted from nursing records and the formation of fundations was confirmed.
- Grasping the characteristics of fundations makes it possible to classify keywords and medical practices in each department.
- Note that many keywords are ambiguous and difficult to associate with a certain word.
- It may be possible to visualize the work practiced by nurses.
- No words or terms considered discriminative or derogatory were found.
- The information selected from the notes written by experienced staff could be made available for education.

Text data mining in general and data analysis of EMRs in particular 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.

The present study offers two possibilities.

One is to refine the analysis of electronic medical text records. Using text data mining techniques, EMR can be quantified and visualized to acquire new knowledge. In addition, since this technique makes it possible to change the number of words displayed as a parameter, further analysis can be done by exploiting such flexibility.

The other possibility guarantees that crucial medical knowledge or information is adequately shared by all medical staff. It can work as a surveillance technique. We only analyzed the nursing records for each medical department and described just one aspect of the features. In the future we might analyze the nursing records for a specific disease in the same way. Such analysis will provide crucial information not only to the medical staff around patients but also to the patients themselves and their families.

## VIII. Conclusion

In the present study, the in-patient nursing records were chosen from among the nursing records preserved by the EMR of the University of Miyazaki Hospital. Sentences were analyzed into morphemes, and the relations between feature vocabularies were analyzed by a text data mining technique to visualize this information. The result analyzed the qualitative in-patient nursing records using the text data mining technique and achieved our initial goal: a visual record of this information. In addition, this result identified vocabularies relating to the proper methods of treatment, resulting in a concise summary of the vocabularies extracted

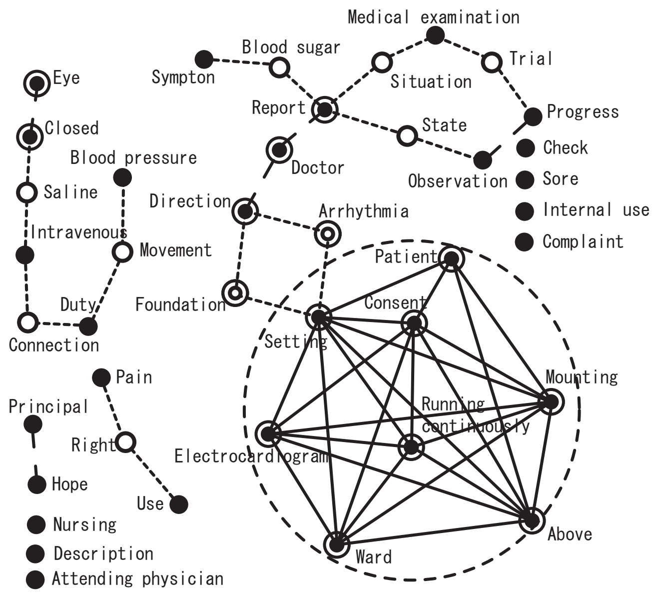


Fig. 5. Internal Medicine 1

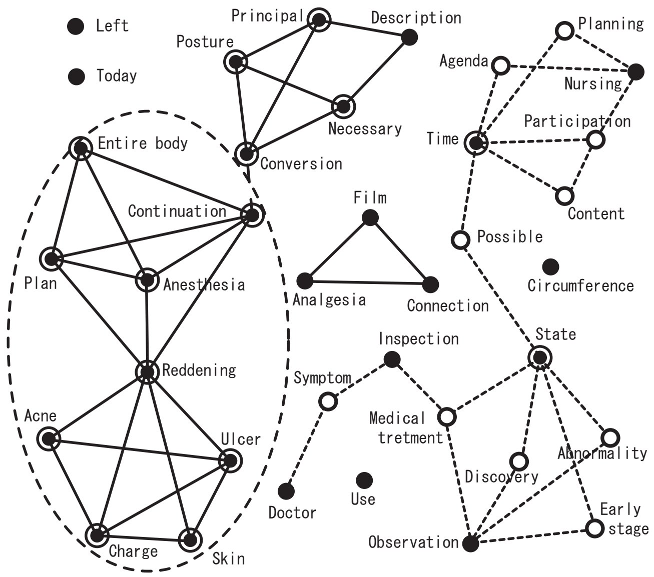


Fig. 8. Surgery 1

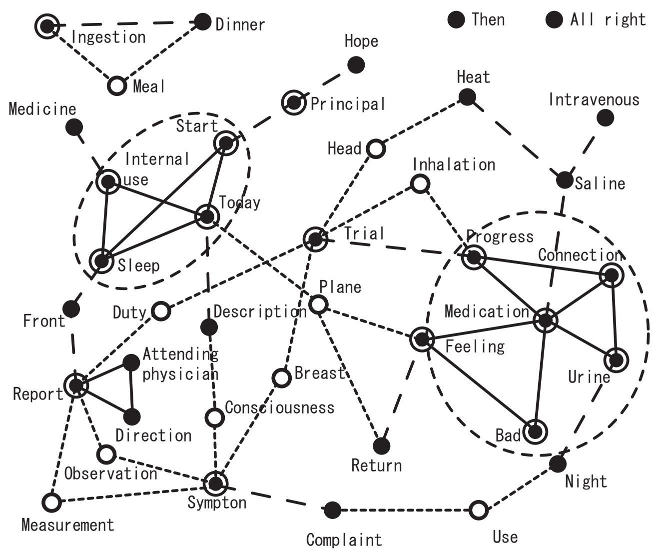


Fig. 6. Internal Medicine 2

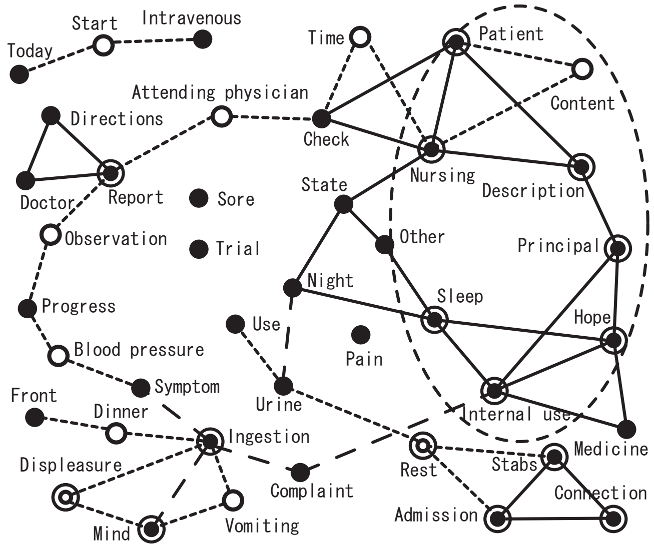


Fig. 9. Surgery 2

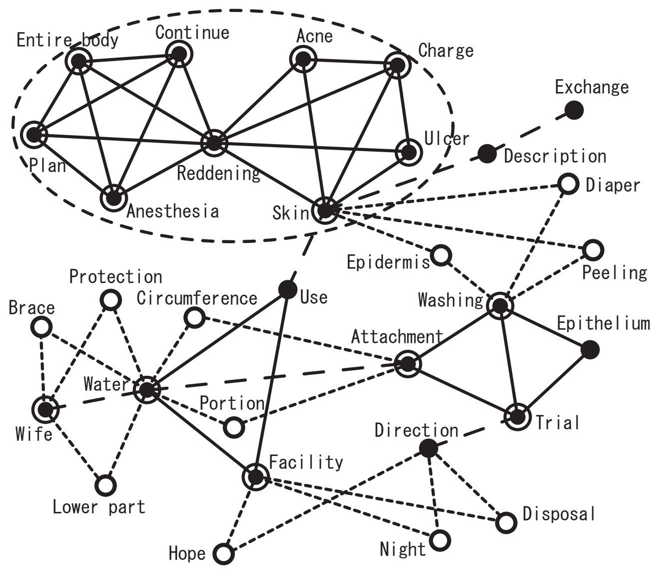


Fig. 7. Internal Medicine 3

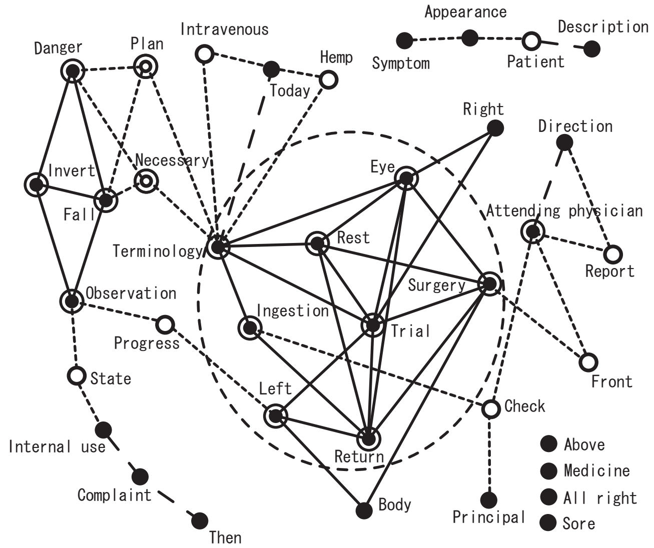


Fig. 10. Ophthalmology

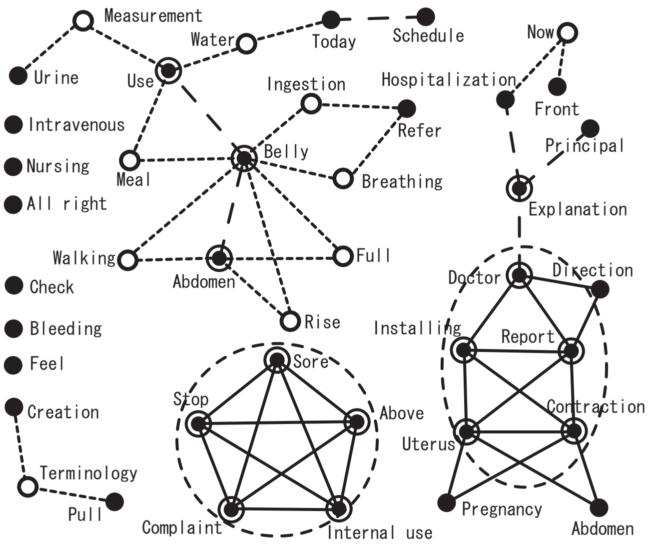


Fig. 11. Obstetrics and Gynecology

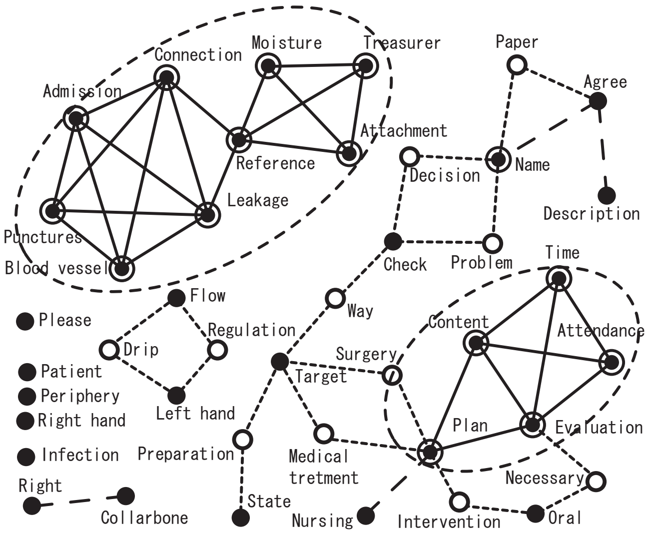


Fig. 12. Oral and Maxillofacial Surgery

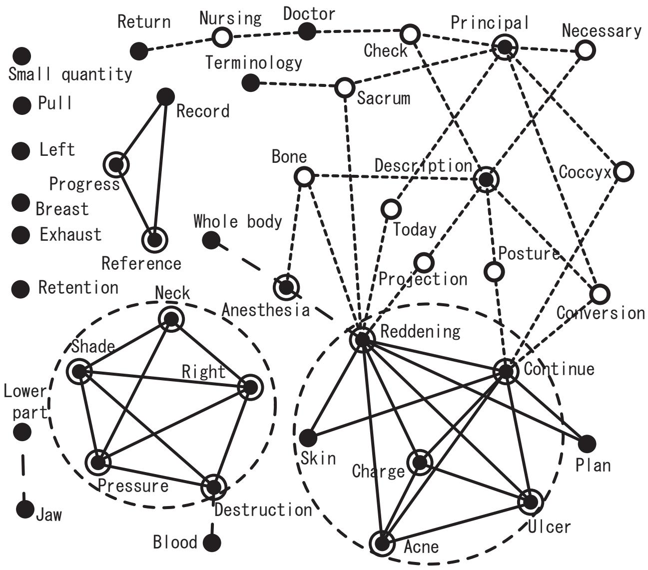


Fig. 13. Otorhinolaryngology

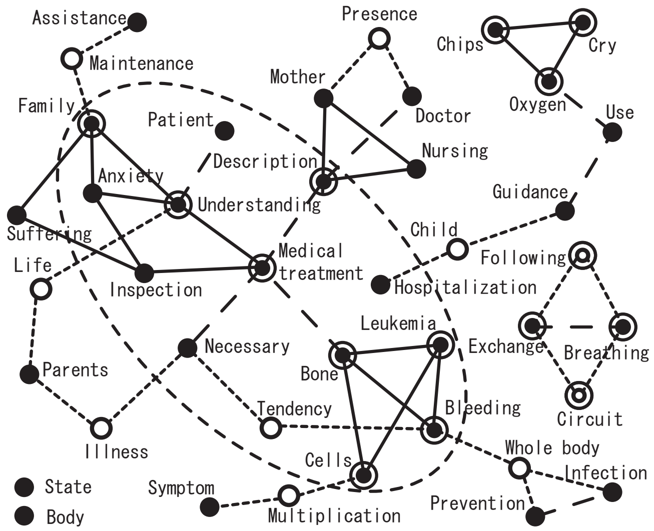


Fig. 14. Pediatrics

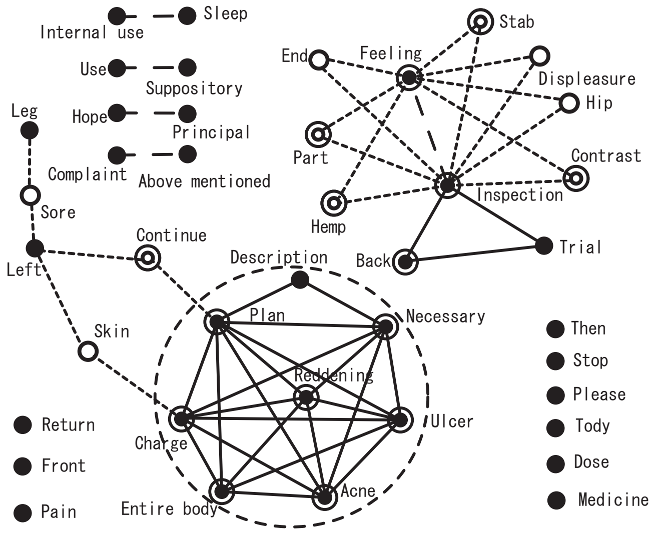


Fig. 15. Orthopedics

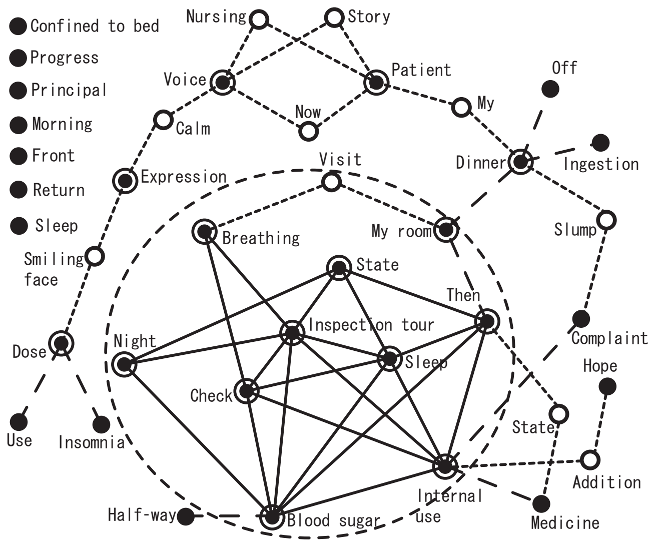


Fig. 16. Psychiatry

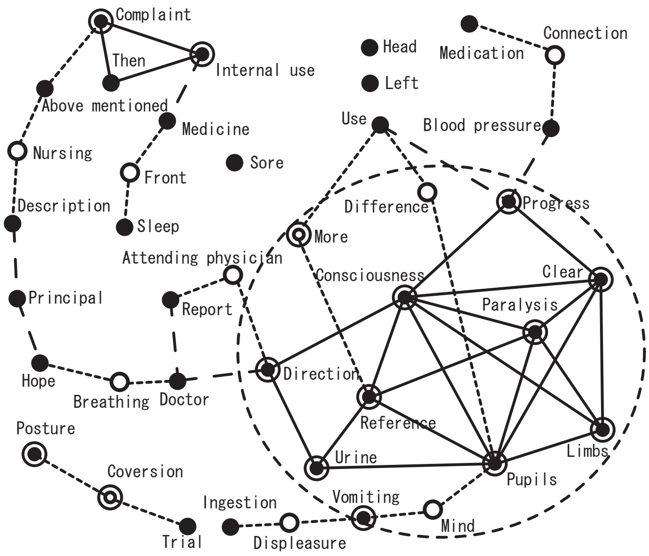


Fig. 17. Neurosurgery

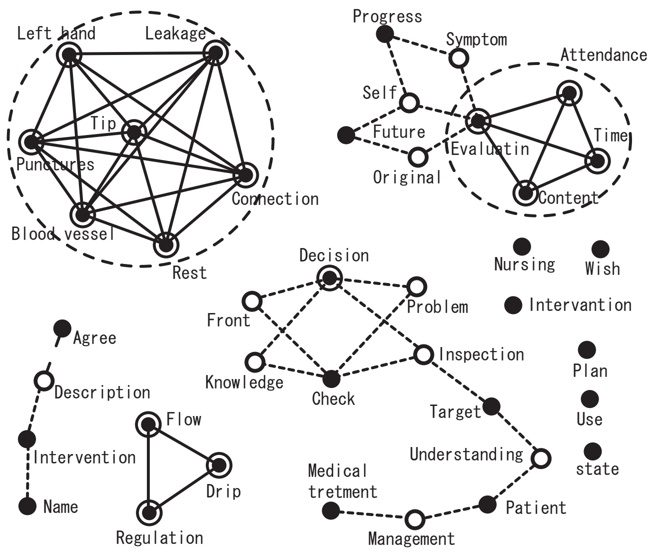


Fig. 20. Radiology

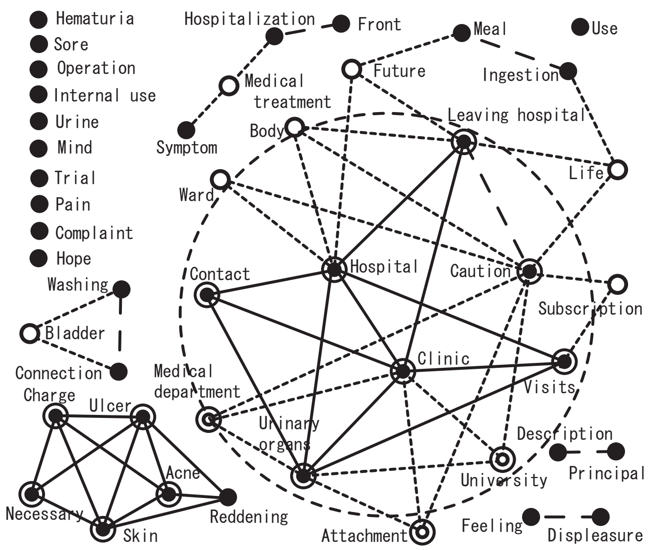


Fig. 18. Urology

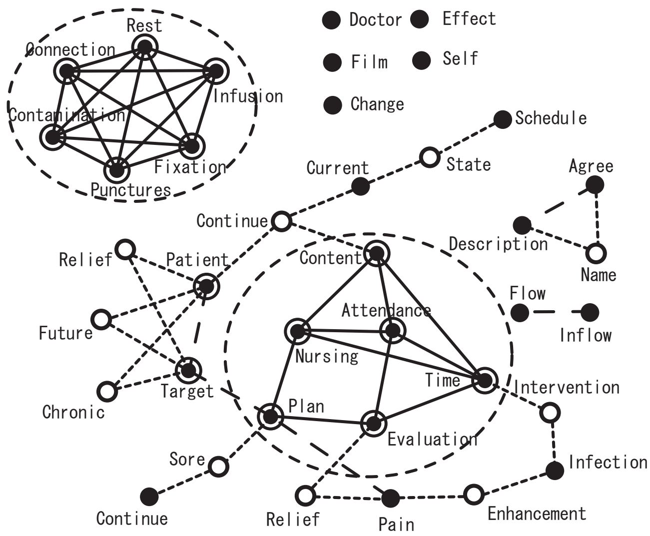


Fig. 21. Anesthesiology

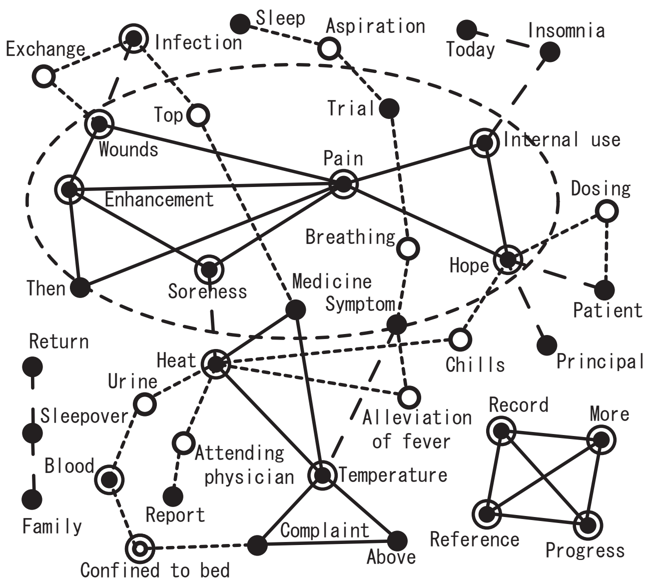


Fig. 19. Dermatology

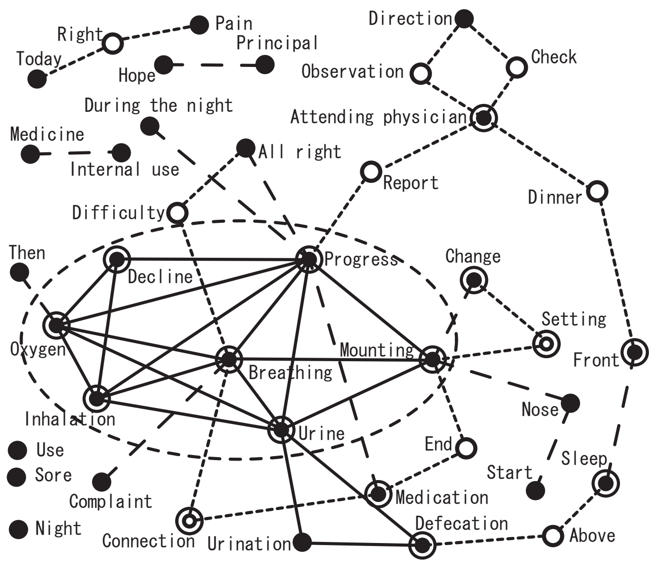


Fig. 22. Rheumatology and Infectious Diseases

TABLE I
RELATIONSHIP BETWEEN TOP TEN WORDS AND THEIR FREQUENCY

| Rank | Internal Medicine 1 | Frequency | Internal Medicine 2 | Frequency | Internal Medicine 3 | Frequency |
|---|---|---|---|---|---|---|
| 1 | Doctor | 167 | Internal use | 131 | Description | 8 |
| 2 | Direction | 135 | Medication | 79 | Charge | 8 |
| 3 | Principal | 134 | Connection | 67 | Ulcer | 8 |
| 4 | Patient | 134 | Urine | 65 | Acne | 8 |
| 5 | Internal use | 126 | Feeling | 55 | Skin | 7 |
| 6 | Report | 125 | Today | 54 | Today | 6 |
| 7 | Hope | 97 | Principle | 49 | Use | 5 |
| 8 | Consent | 87 | Attending physician | 49 | Entire body | 4 |
| 9 | Observation | 86 | Progress | 47 | Plan | 4 |
| 10 | Use | 84 | Hope | 46 | Exchange | 4 |
| Rank | Surgery 1 | Frequency | Surgery 2 | Frequency | Ophthalmology | Frequency |
| 1 | Acne | 49 | Internal use | 509 | Eye | 66 |
| 2 | Ulcer | 44 | Principal | 260 | Internal use | 59 |
| 3 | Change | 44 | Sleep | 240 | Complaint | 50 |
| 4 | Skin | 43 | Use | 238 | Trial | 42 |
| 5 | Reddening | 39 | Complaint | 195 | Terminology | 38 |
| 6 | Description | 38 | Description | 190 | Today | 35 |
| 7 | Entire body | 30 | Hope | 181 | Direction | 35 |
| 8 | Plan | 30 | Progress | 178 | Rest | 33 |
| 9 | Necessary | 29 | Urine | 165 | Sore | 32 |
| 10 | Continuation | 27 | Then | 164 | Surgery | 31 |
| Rank | Obstetrics and Gynecology | Frequency | Oral and Maxillofacial Surgery | Frequency | Otorhinolaryngology | Frequency |
| 1 | Sore | 97 | Admission | 94 | Pressure | 97 |
| 2 | Internal use | 77 | Blood vessel | 93 | Shade | 96 |
| 3 | Complaint | 75 | Punctures | 93 | Destruction | 93 |
| 4 | Direction | 69 | Connection | 93 | Right | 83 |
| 5 | Today | 63 | Leakage | 92 | Neck | 80 |
| 6 | Bleeding | 63 | Nursing | 92 | Left | 67 |
| 7 | Doctor | 63 | Reference | 87 | Blood | 63 |
| 8 | Report | 57 | Infection | 69 | Chest | 40 |
| 9 | Above mentioned | 56 | Plan | 66 | Progress | 36 |
| 10 | Uterus | 51 | Attachment | 66 | Reference | 35 |
| Rank | Pediatrics | Frequency | Orthopedics | Frequency | Psychiatry | Frequency |
| 1 | Medical treatment | 18 | Internal use | 426 | Internal use | 234 |
| 2 | Description | 17 | Sleep | 181 | State | 203 |
| 3 | Nursing | 15 | Hope | 114 | Sleep | 188 |
| 4 | Doctor | 15 | Principal | 112 | Arousal | 168 |
| 5 | Suffer | 14 | Medicine | 105 | Then | 162 |
| 6 | State | 14 | Complaint | 94 | Check | 151 |
| 7 | Necessary | 12 | Above mentioned | 66 | Inspection tour | 131 |
| 8 | Mother | 12 | Use | 69 | Night | 102 |
| 9 | Breathing | 12 | Trial | 62 | Complaint | 101 |
| 10 | Family | 11 | Description | 60 | Principal | 92 |
| Rank | Neurosurgery | Frequency | Urology | Frequency | Dermatology | Frequency |
| 1 | Internal use | 132 | Internal use | 29 | Internal use | 125 |
| 2 | Then | 110 | Principal | 19 | Pain | 71 |
| 3 | Complaint | 104 | Bloody urine | 15 | Hope | 49 |
| 4 | Direction | 90 | Complaint | 15 | Progress | 44 |
| 5 | Consciousness | 88 | Urine | 14 | Reference | 40 |
| 6 | Doctor | 79 | Use | 13 | Principal | 31 |
| 7 | Sore | 78 | Description | 13 | Record | 31 |
| 8 | Progress | 78 | Hope | 13 | More | 30 |
| 9 | Description | 72 | Skin | 12 | Patient | 29 |
| 10 | Above mentioned | 70 | Operation | 12 | Sore | 29 |
| Rank | Radiology | Frequency | Anesthesiology | Frequency | Rheumatology and Infections Diseases | Frequency |
| 1 | Blood vessel | 42 | Infusion | 111 | Internal use | 63 |
| 2 | Leakage | 41 | Punctures | 110 | Urine | 60 |
| 3 | Admission | 41 | Admission | 100 | Progress | 56 |
| 4 | Punctures | 41 | Rest | 108 | Breathing | 47 |
| 5 | Connection | 41 | Connection | 106 | Oxygen | 44 |
| 6 | Tip | 40 | Fixation | 105 | All right | 32 |
| 7 | Rest | 39 | Pollution | 105 | Use | 32 |
| 8 | Left hand | 38 | Nursing | 27 | Inhalation | 30 |
| 9 | Nursing | 33 | Content | 23 | Night | 29 |
| 10 | Plan | 24 | Attendance | 22 | Principal | 28 |

from the in-patient nursing records. We constructed important vocabularies characterizing each nursing record.

This research suggests the fruitful possibility of automatically detecting a disease and classifying it from documents used at medical treatment sites. In the future, using a text data mining approach and laterally processing medical documents will support disease classification by retrieving the examples of similar syndromes. Our approach can also be applied to the discovery of new medical knowledge for new syndrome extraction. Text data mining is expected to become a valuable technique in the analysis of medical documents in the future. We intend to accumulate clinical research data from care cards that evaluate the prognosis, the prognostic factors, the treatment results, and the safety of Medical Treatment Technologies. In the future, this information will be related to cost reduction and improvements in the efficiency and quality of clinical research.

## IX. Future work

Future work will concentrate on passage records filed by doctors to improve analytical results. In addition, we expect to obtain further results by analyzing other diagnoses in other treatment departments and comparing them.

## Acknowledgments

The authors thank the members of the Medical Information Department at the University of Miyazaki Hospital. We also thank Noah Morris for his help with editing this research paper.

## References

[1] D.A. Ludwick and J. Doucette, “Adopting electronic medical records in primary care: Lessons learned from health information systems implementation experience in seven countries,” *International Journal of Medical Information*, no. 78, pp. 22-31, 2009.

[2] Y. Kinosada, “Bio Medical Informatics in the Era of Electronic Patient Record System,” *Japan Association for medical Informatics*, vol. 23, no. 5, pp. 397-405, 2003.

[3] M. Suzuki, K. Araki and H. Yoshihara, “Audit Report of the Medical Information System,” *Japan Association for medical Informatics*, vol. 22, no. 4, pp. 347-353, 2002.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] H. Abe, S. Hirano and S. Tsumoto, “Evaluation Temporal Models Based on Association Mining From Medical Documents,” *Japan Association for Medical Informatics*, vol. 27, no. 1, pp. 33-38, 2007.

[14] K. Izumi, K. Goto and M. Matsui, “Analysis of Financial Markets’ Fluctuation by Textual Information,” *The 23rd Annual Conference of the Japanese Society for Artificial Intelligence*, 2009.

[15] 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)*, pp. 686-693, Seoul, Korea, April. 2010.

[16] 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.

[17] Y. Ohsawa, Benson, Nels E. 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(2), pp. 391-400, 1998.

[18] K. Horie, Y. Ohsawa and N. Okazaki, “Products Designed on Scenario Maps using Pictorial KeyGraph,” *Proceedings of the 5th WSEAS International Confernce on Applied Computer Science*, pp. 801-808, Hangzhou, China, April, 2006.

[19] Kyung-Joong. Kim, Myung-Chul. Jung and Sung-Bae. Cho, “KeyGraph-based chance discovery for mobil contents management system,” *International Journal of Knowledge and Intelligent Engineering System*, Vol.11, pp. 313-320, 2007.

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

[21] Y. Ohsawa, Benson, Nels E. and M. Yachida, “KeyGraph: Automatic Indexing by Co-occurrence Graph based on Building Construction Metaphor,” *Fifth International Forum on Research and Technology Advances in Digital Libraries (ADL’98)*, 1998.

[22] 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, 2009.

[23] M. Kushima, K. Araki, S. Suzuki and S. Araki, “Challenge to Qualitative Analysis for Electronic Medical Record (IZANAMI) Using KeyGraph,” *Asia Pacific Association for Medical Informatics (APAMI)*, Hiroshima, Japan, 2009.

[24] 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.

[25] K. Tsuda and R. Thawonmas, “Visualization of Discussions in Comments of a Blog Entry Using KeyGraph and Comment Scores,” *Proc. Of 4th WSEAS International Conference on E-ACTIVITIES*, Florida, USA, vol. 5, pp. 21-26, 2005.

[26] R. Thawonmas and K. Hata, “Aggregation of Action Symbol Subsequences for Discovery of Online-game Player Characteristics Using KeyGraph,” *Proc. of IFIP 4th International Conference on Entertainment Computing (ICEC 2005)*, Sanda, Japan, vol. 3711, pp. 126-135, 2005.

[27] Y. Seo, Y. Iwase and Y. Takama, “KeyGraph-based BBS for Supporting Online Chance Discovery Process,” *SCIS ISIS2006*, vol. FR-I3-2, pp. 1210-1214, 2006.

[28] K. Tsuda and R. Thawonmas, “KeyGraph for Visualization of Discussions in Comments of a Blog Entry with Comment Scores,” *World Scientific and Engineering Academy and Society(WSEAS) Trans, Computers*, vol. 12, no. 4, pp. 1794-1801, 2005.

[29] G. Salton and M. J. McGill, “Introduction to Modern Information Retrieval,” *McGraw-Hill*, 1983.

[30] M. F. Porter, “An Algorithm for Suffix Stripping,” *Automated Library and Informations Systems*, vol. 14, no. 3, pp. 130-137, 1980.

[31] J. Cohen, “Language and Document, Automatic Indexing Terms for Abstracting,” *Journal of American Society for Information Science*, vol. 46, pp. 162-174, 1995.