Abstract—This research used a text data mining technique to extract useful information from nursing records within Electronic Medical Records. Although nursing and passage 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 chronic hepatitis 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 and passage 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 and passage records. Important vocabularies that characterize each nursing and passage record were also revealed.

Index Terms—text data mining, electronic medical record, chronic hepatitis, nursing record, passage record, KeyGraph

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 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]. In this study, we chose in-patient nursing and passage records from among nursing and passage 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.

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 vendors, 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 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.

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 insertion into a database), deriving patterns within the structured data, and finally evaluating and interpreting the output. Fig. 1 shows the process of text data mining. Two particular aspects should be considered when applying text data mining to a medical context. Two particular aspects should be considered when applying text data mining to a medical context. 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.

IV. 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 2 shows an image of KeyGraph. Figure 3 shows an example when it is 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. 3, 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.

V. ANALYSIS RESULTS

In this paper, EMR data were collected the chronic hepatitis in-patient nursing and passage records from among nursing and passage records preserved by the EMR system at the University of Miyazaki Hospital, and the nursing records from August 2007 to July 2009 were used. The following analysis results are shown:

(patient 145: 1128 total document (nursing records 173 document, passing records 955 document )
1) Nursing record: Fig.4
   - foundation 1,2:
     Doctors observe the drug injection site at the beginning and end.
   - foundation 3:
     The agreement is confirmed at the time of drug administration.

   The foundations are obtained from the text data with event-sets 1: { foreign, noodle, order, end, left hand, outside, administration, start }, 2: { humerus, enforcement, injection, subcutaneous, right hand, right } and 3: { abdominal, liver, hospitalization, consent, test }.

2) Passage record: Fig.5
   - foundation 1:
     Method and duration of drug administration
   - foundation 2:
     Determine whether the administration actually
   - foundation 3:
     A check for side effects

Nurses are being evaluated for conditions such as dizziness, the doctor can evaluate and understand the laboratory results. Doctors observe whether the drug had positive or negative effect and nurses record the administration of the drug.

The foundations are obtained from the text data with event-sets 1: { subcutaneous, week, morning, annotation, weekly }, 2: { continuity, antigen, author, change, state } and 3: { caution, anemia, decrease, leukocyte, individual, platelet, sphere }.

Table I shows the relationship between the top ten words and their frequency in the University of Miyazaki Hospital chronic hepatitis in-patient nursing and passage records. For example, in nursing record, injection, appeared 19 times. Subcutaneous, humerus, test, symptom, mood, please, liver, failure, schedule, outside, complaint, safe, tomorrow, hospitalization, oral, right hand, before, left hand and start also appear often.

VI. CONSIDERATION

The following is an overall evaluation of this research:
- Due to KeyGraph analysis, characteristic keywords were extracted from nursing and passage records and the formation of foundations was confirmed.
- Grasping the characteristics of foundations 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 and doctors.
- 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.

VII. CONCLUSION

In the present study, the chronic hepatitis in-patient nursing and passage records were chosen from among the nursing passage 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 and passage 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 from the in-patient nursing and passage records. We constructed important vocabularies characterizing each nursing and passage 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.
ACKNOWLEDGMENT

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

REFERENCES


Table I Relationship between top 20 words and frequency

<table>
<thead>
<tr>
<th>Rank</th>
<th>Word</th>
<th>Frequency</th>
<th>Rank</th>
<th>Word</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>injection</td>
<td>19</td>
<td>6</td>
<td>outside</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>subcutaneous</td>
<td>12</td>
<td>7</td>
<td>complaint</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>humor</td>
<td>11</td>
<td>8</td>
<td>safe</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>test</td>
<td>10</td>
<td>9</td>
<td>tomorrow</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>symptom</td>
<td>10</td>
<td>10</td>
<td>hospitalization</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>mood</td>
<td>9</td>
<td>11</td>
<td>oral</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>please</td>
<td>8</td>
<td>12</td>
<td>right hand</td>
<td>5</td>
</tr>
<tr>
<td>8</td>
<td>liver</td>
<td>8</td>
<td>13</td>
<td>before</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>failure</td>
<td>7</td>
<td>14</td>
<td>left hand</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>schedule</td>
<td>7</td>
<td>15</td>
<td>start</td>
<td>4</td>
</tr>
</tbody>
</table>

(a) nursing record

<table>
<thead>
<tr>
<th>Rank</th>
<th>Word</th>
<th>Frequency</th>
<th>Rank</th>
<th>Word</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>administration</td>
<td>339</td>
<td>11</td>
<td>platelet</td>
<td>167</td>
</tr>
<tr>
<td>2</td>
<td>today</td>
<td>225</td>
<td>12</td>
<td>leaving hospital</td>
<td>156</td>
</tr>
<tr>
<td>3</td>
<td>hepatitis</td>
<td>190</td>
<td>13</td>
<td>chronic</td>
<td>154</td>
</tr>
<tr>
<td>4</td>
<td>start</td>
<td>189</td>
<td>14</td>
<td>schedule</td>
<td>147</td>
</tr>
<tr>
<td>5</td>
<td>fever</td>
<td>188</td>
<td>15</td>
<td>symptom</td>
<td>147</td>
</tr>
<tr>
<td>6</td>
<td>liver</td>
<td>182</td>
<td>16</td>
<td>abdominal</td>
<td>146</td>
</tr>
<tr>
<td>7</td>
<td>inspection</td>
<td>177</td>
<td>17</td>
<td>oral</td>
<td>136</td>
</tr>
<tr>
<td>8</td>
<td>decrease</td>
<td>173</td>
<td>18</td>
<td>ophthalmology</td>
<td>133</td>
</tr>
<tr>
<td>9</td>
<td>individual</td>
<td>169</td>
<td>19</td>
<td>side-effects</td>
<td>131</td>
</tr>
<tr>
<td>10</td>
<td>course</td>
<td>168</td>
<td>20</td>
<td>sphere</td>
<td>129</td>
</tr>
</tbody>
</table>

(b) passage record


