

Development of a System that Generates Structured Reports for Chest X-ray Radiography

Y. Hasegawa¹; Y. Matsumura²; N. Mihara¹; Y. Kawakami³; K. Sasai³; H. Takeda²; H. Nakamura¹

¹Department of Diagnostic and Interventional Radiology, Osaka University Graduate School of Medicine, Suita, Osaka, Japan;

²Department of Medical Informatics, Osaka University Graduate School of Medicine, Suita, Osaka, Japan;

³Department of Imaging System R&D Division, Konica Minolta Technology Center, Inc, Takatsuki, Osaka, Japan

Keywords

Chest x-ray photography, radiology report, structured report, resource description framework, information model

Summary

Objectives: Radiology reports are typically made in narrative form; this is a barrier to the implementation of advanced applications for data analysis or a decision support. We developed a system that generates structured reports for chest x-ray radiography.

Methods: Based on analyzing existing reports, we determined the fundamental sentence structure of findings as compositions of procedure, region, finding, and diagnosis. We categorized the observation objects into lung, mediastinum, bone, soft tissue, and pleura and chest wall. The terms of region, finding, and diagnosis were associated with each other. We expressed the terms and the relations between the terms using a resource description framework (RDF) and developed a

reporting system based on it. The system shows a list of terms in each category, and modifiers can be entered using templates that are linked to each term. This system guides users to select terms by highlighting associated terms. Fifty chest x-rays with abnormal findings were interpreted by five radiologists and reports were made either by the system or by the free-text method.

Results: The system decreased the time needed to make a report by 12.5% compared with the free-text method, and the sentences generated by the system were well concordant with those made by free-text method (F-measure = 90%). The results of the questionnaire showed that our system is applicable to radiology reports of chest x-rays in daily clinical practice.

Conclusions: The method of generating structured reports for chest x-rays was feasible, because it generated almost concordant reports in shorter time compared with the free-text method.

reporting system based on it. The system shows a list of terms in each category, and modifiers can be entered using templates that are linked to each term. This system guides users to select terms by highlighting associated terms. Fifty chest x-rays with abnormal findings were interpreted by five radiologists and reports were made either by the system or by the free-text method.

According to the order and sends the image data to a PACS server. A radiologist retrieves the images from the PACS server by a viewer system and makes radiology reports by using a reporting system, which can be seen by the physician who placed the order (► Fig. 1). X-ray findings and diagnoses are so diverse that radiologists usually enter their reports in free-text with or without speech recognition technology [1]. However, reports in free-text cannot be processed by machine, and secondary uses of the entered data are therefore difficult [2]. Recent studies have shown that structured reports have potential benefits for research, education, and diagnosis [3–7] by improving searching, storage, and semantic interoperability. In the field of radiology reporting, the importance of structured reporting has also been acknowledged. The Digital Imaging and Communications in Medicine (DICOM) Structured reporting (SR) was presented in Supplement 23 of the 2000 edition of the DICOM standard. Several studies have suggested that the DICOM SR might be feasible and useful [8–13]. Also, the Radiological Society of North America (RSNA) is building a radiological lexicon (RadLex), a controlled terminology of standardized terms used in radiology [14–16]. RadLex can be the foundation of such a system generating structured report.

To generate a structured report, structured data entry is one of the methods where predefined structure and vocabulary are used at the time of data entry [17]. However, structured data entry is not yet commonly used. Because efficiency in daily clinical practice has priority over the requirements for research and decision

Correspondence to:

Yasushi Matsumura
Department of Medical Informatics
Osaka University Graduate School of Medicine
2-15 Yamadaoka, Suita
Osaka 565-0871
Japan
E-mail: matsumura@hp-info.med.osaka-u.ac.jp

Methods Inf Med 2010; 49: 360–370

doi: 10.3414/ME09-01-0014

received: February 16, 2009

accepted: February 19, 2010

prepublished: July 6, 2010

1. Introduction

In recent years, electronic medical record systems and PACS (picture archiving and communication system) have been implemented in many hospitals. In such a com-

puterized hospital, a physician requests a radiology examination for the patient by a computerized order entry system. The patient information is transferred to a radiology information system. A radiological technologist then takes some images ac-

support, medical staff rejects structured data entry if it takes more time than conventional free-text entry [18, 19].

Natural language processing (NLP) is another solution [20, 21] for extraction of structured data from radiology free-text reports. In the field of radiology reports, the success of automatic structuring of English texts by NLP in thoracic CT and x-ray reports has been reported with accuracy rates over 70% [22, 23]. Also, Dang showed substantial differences between the extraction rates of positive findings for different imaging modalities ranging from 23.1% (mammography) to 85.1% (nuclear medicine) [24]. NLP in Japanese is exceedingly difficult because terms have to be extracted from continuing letter strings [25].

Recent work has focused on extracting medical knowledge using ontology [26, 27]. Baneyx has generated various types of medical ontologies based on terminology extraction from texts using NLP [28]. However, the systems described in these papers which use extracted knowledge by NLP are not widely used at present.

For structured reports in Japanese, structured data entry is thought to be more practical than NLP. However, in order to succeed in structured data entry, it should take equal or shorter time to make a sufficient report compared with free-text typing. In this study, we propose a method of generating structured reports for chest radiography in which the sentences are generated by selecting candidate terms. We initially constructed an information model in the domain of radiology reports for chest x-rays [29], which was described in the Resource Description Framework (RDF) [30–32]. RDF is recommended for presenting metadata in World Wide Web Consortium (W3C). RDF Object-Attribute-Value triples form the basis of RDF, which can be serialized as the eXtensible Markup Language (XML) and represent relationships [33, 34].

We developed a reporting system for chest radiography based on the information model. This system was intended to be used in Japan, so the terms in the system were presented in Japanese. Selection from a list of candidate terms is not easy because a user has to select from a list of many terms in radiology reports. To overcome this

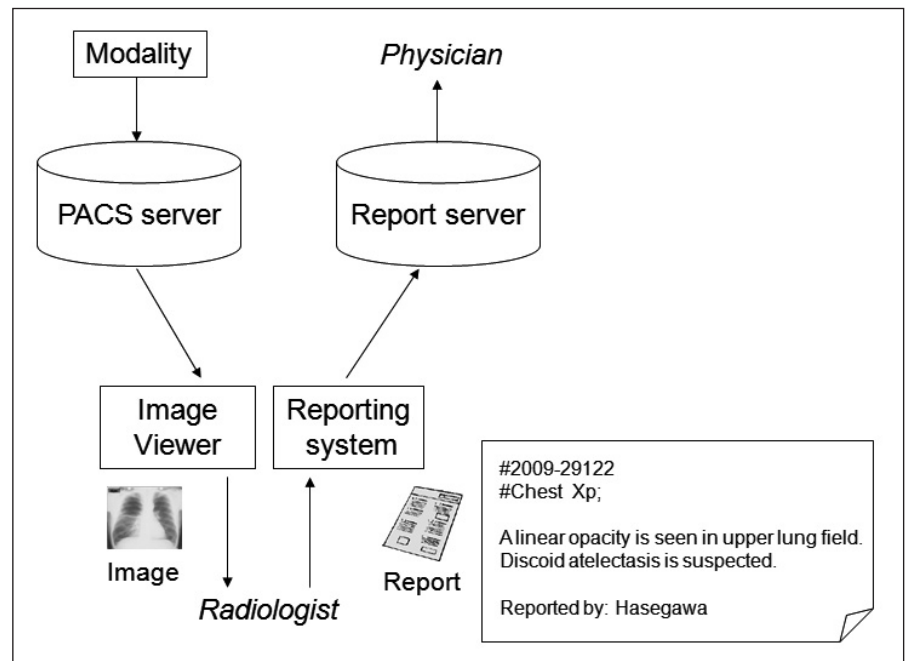


Fig. 1 The operation of x-ray radiology reports in a radiology department. A radiologist retrieves images from PACS server by an image viewer. After reading the images, radiology reports are input by using a reporting system, which are sent to the report server then seen by the physician.

problem, we narrowed the candidate terms in the list according to the terms already selected by a user. After selecting terms, the system uses them to generate a narrative sentence. At the same time, the system also outputs a structured report in RDF. In this paper, we describe the details of the system and the information model constructed for it, and also the results of an evaluation of the system.

2. Methods

2.1 Construction of Information Model for Chest Reports

We created an information model by manually extracting terms and sentences from 1200 de-identified chest x-ray reports made in Osaka University Hospital. These reports had been typed into a free-text-based reporting system, without generated-dictation/transcription and speech recognition.

Radiology reports of x-rays typically consisted of lists of descriptions about observation objects and a summary. The observation objects in chest x-rays could be

categorized into lung, mediastinum, bone, soft tissue, and pleura and chest wall. Sentences regarding an object in a report typically consisted of the concepts of procedure, region, finding, and diagnosis. Among these, region and finding were essential. The description of region was typically composed of basic region, side, and other modifiers. The description of finding was typically composed of characters and modifiers. The terms of diagnosis were often followed by a modification of uncertainty, such as “is suspected” or a negation such as “is denied”.

There were associations among the terms for region, finding, and diagnosis. Some combinations of terms were common, while others were impossible. We asked a chest x-ray radiography specialist about the probability of each combination of a term for region and a term for finding, as well as a term for finding and a term for diagnosis. For each possible combination of terms for region, finding, and diagnosis, the specialist provided the probability. The probabilities were categorized into “frequent”, “possible”, and “impossible”.

The terms for procedure, region, finding, and diagnosis were extracted from re-

using Eclipse Graph Editing Framework (GEF) [36]. We used Jena to handle the RDF, which was linked with PostgreSQL [37] as a database management system.

The information model described by the RDF was parsed by Jena [38] and the terms and the relationships among them were stored into the database. The interface for selecting terms to generate reports was linked to the database via Jena. The system showed a window that presented candidate terms retrieved from the database by Jena. Users could add a term to the corresponding class, which was expressed by the RDF

and stored in the database via Jena. Reports made by selecting terms were also expressed by the RDF and stored in the database by the same method. When a term was selected by a user, “frequent” or “possible” subsequent terms were highlighted according to the probability of the combinations of the selected term and candidate terms. The probability data were drawn directly from the database.

The OS of the client PC was Windows XP. That of the server was Fedora core 6.

2.3 System Operation

When a patient is selected from the patient list, the initial window opens and displays the category list (lung, mediastinum, bone, soft tissue, pleura and chest wall), the list of frequent phrases, and the list of summary phrases. To enter detailed findings, one of the categories is selected. The term selection window for the selected category is opened, and displays the list of terms for procedure, region, finding, and diagnosis (► Fig. 5). The user selects terms in these nodes to generate a sentence in a report.

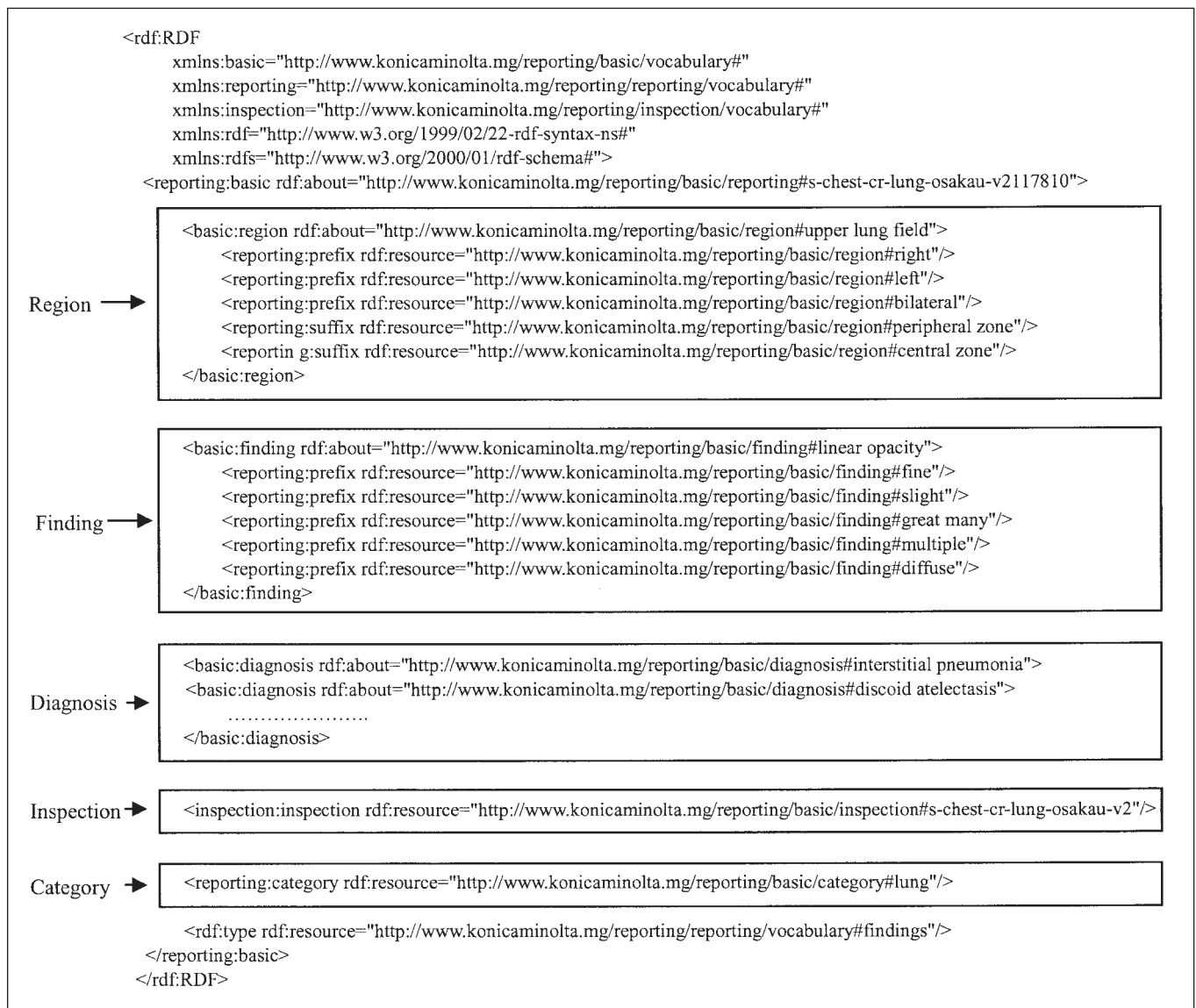


Fig. 3 An example of RDF expression regarding configurational structure. The example shows “upper lung field” in region, and its prefixes and suffixes; “linear opacity” in finding, and its prefixes; and “interstitial pneumonia” and “discoid atelectasis” in diagnosis in the category of “lung”.

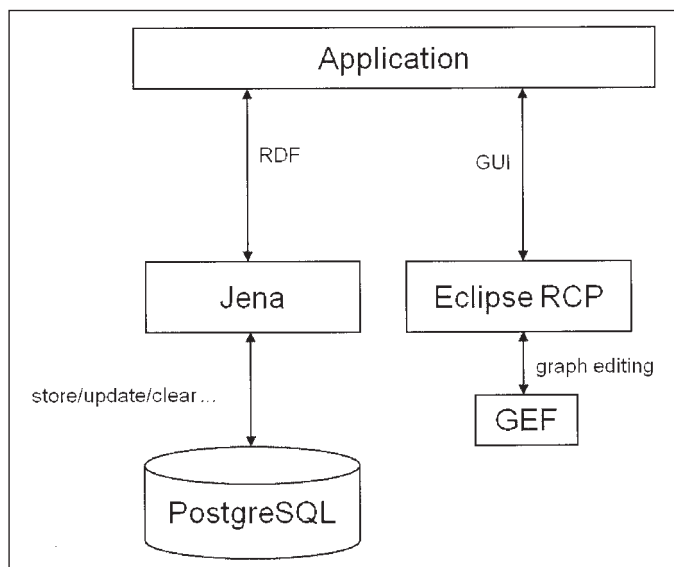


Fig. 4

System architecture. The information model described by the RDF was parsed by Jena and the terms and the relationships among them were stored into PostgreSQL. The system showed the graphical user interface and graph editing using each Eclipse RCP and Eclipse Graph Editing Framework (GEF).

The term for procedure was sometimes omitted. If modifiers need to be added to a term, an input template for modifiers is opened. After selecting terms, the system creates a narrative sentence by composing the terms as an interpretation result. Examples of composed narrative sentences in Japanese are shown in ▶Figure 6. Users can select a sentence from the list of frequent and summary phrases. If plural findings are observed, users repeat this procedure. After completing the report, narrative data is sent to the report system, and at the same time the structured data presented in RDF is stored.

To aid users in selecting a term from many candidates, the system guides users by highlighting the possible terms. For

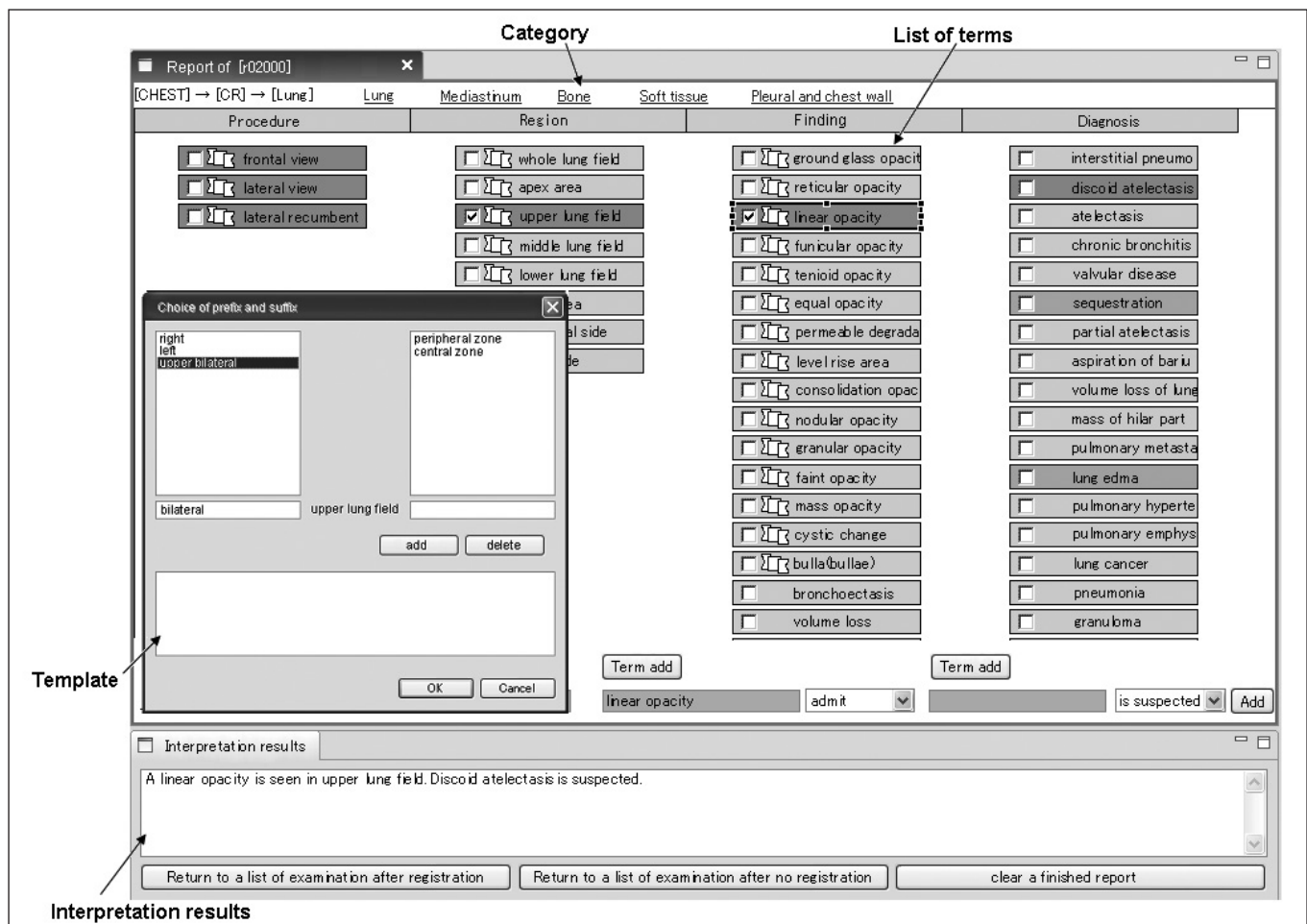


Fig. 5 Example of the reporting system screen. This term selection window opens after selecting a category in main screen. In actual system, the terms are written in Japanese, which are translated into English for this paper. After selection of lung in category, this window presents the lists of procedure, region, finding, and diagnosis. When "upper lung field" is selected in region

and "linear opacity" is selected in finding, "discoid atelectasis" is distinctly highlighted in diagnosis. When "upper lung field" is selected, the input template for its modifiers appears to prompt the user to enter detailed region information. After selecting these terms, a narrative radiology report is generated in the interpretation results field (lower fields).

example, when a user selects “upper lung field” for region and “linear opacity” for finding, the system highlights “discoid atelectasis”, “sequestration”, “lung edema”, “inflammatory change”, “lymphangitis carcinomatosa”, and “cardiac insufficiency” as possible diagnosis terms. Among these possible terms, “discoid atelectasis” is presented in a different color, indicating that it is more likely than the others. When a user first selects “discoid atelectasis” as the diagnosis, “linear opacity”, “funicular opacity”, and “consolidation opacity” are highlighted for “finding”. The highlighting is

according to the data of probability of combination of terms drawn directly from the database. If a term that a user wants to enter is not included in the list, the user can enter the term in the textbox. The terms added by a user can be available for making reports from then.

2.4 System Evaluation

To evaluate the feasibility of this term selection method, five radiologists made radiology reports of chest x-rays using both the

system and the free-text typing. The radiologists had the following experience: “A” was a first year radiologist; “B” was a fifth year abdominal radiologist with little experience in chest x-ray interpretation; “C” was a fifth year chest radiologist; “D” was a seventh year nuclear medicine specialist with little experience in chest x-ray interpretation; and “E” was a seventh year chest radiologist. Each of them interpreted 50 chest x-rays with abnormal findings and made radiology reports by using our system. After a 10-day interval, the radiologists interpreted the same chest x-rays

- 1) An example of a narrative sentence when terms in procedure, region, finding and diagnosis are selected.

正面像にて、右下肺野に結節影を認める。肺癌の疑い。

{prefix}[procedure]にて、{prefix}[region]{suffix}に{prefix}[finding]を認める。[diagnosis]の疑い。

{prefix}[procedure], {prefix}[finding] is seen in {prefix}[region]{suffix}. [diagnosis] is suspected.

In frontal view, a nodular opacity is seen in right lower lung field. Lung cancer is suspected.

- 2) An example of a narrative sentence when terms in region, finding and diagnosis are selected.

左下肺野に複数の腫瘍影を認める。転移性肺癌の疑い。

{prefix}[procedure]にて、{prefix}[region]{suffix}に{prefix}[finding]を認める。[diagnosis]の疑い。

{prefix}[procedure], {prefix}[finding] are seen in {prefix}[region]{suffix}. [diagnosis] is suspected.

Several mass opacities are seen in left lower lung field. Metastatic lung cancer is suspected.

Fig. 6 Example of composing a narrative sentence in Japanese by term selection method. After a user selects terms, the system composes sentences by filling the elements in the prepared sentence templates. The elements consist of procedure prefix, procedure, region prefix, region, region suffix, finding prefix, finding, and diagnosis. All elements are optional. If an element is not filled, the postpositional particle is also not shown. A prefix or a suffix is en-

closed within parenthesis {}, and a skeletal element (procedure, region, finding and diagnosis) is enclosed within parenthesis []. The first line shows narrative sentences in Japanese composed by the system. The second line shows the sentence template. Selected terms are underlined with arrows pointing to the corresponding elements in the template. The fourth and the third lines are the English translation of the first and the second lines, respectively.

and reported by free-text typing. The same 50 x-rays in the same assignment of image were used in both experiments. We measured the time to generate a report from opening the input window to completing it.

We compared the sentences in the reports made by the two methods. We measured the concordance of sentences by using F-measure method [39, 40]. A sentence was separated into phrases about procedure, region, finding, diagnosis and other. If the meaning of a phrase in a sentence generated by the term selection method was concordant with the corresponding phrase in that by free-text method, the phrase was counted as "true positive". If a phrase in a sentence generated by free-text typing did not exist in a sentence generated by term selection method, or if the meaning of it was different from that of the corresponding phrase in the sentence by the term selection method, then it was counted as "false negative". Conversely, if a phrase in a sentence generated by the term selection method did not exist in a sentence by free-text method, or the meaning of it was different from the corresponding one by free-text method, it was counted as "false positive". F-measure was calculated as follows:

$$F\text{-measure} = \frac{2 \times \text{true pos.}}{2 \times \text{true pos.} + \text{false neg.} + \text{false pos.}}$$

To evaluate the sufficiency of the candidate terms displayed according to the information model, we counted the number of terms added to the system by the radiologists because the terms they wanted to use did not exist in the list.

Finally, we provided a questionnaire including 14 questions in four categories: overall assessment; screen layout, flow, and ease of operation; number of terms and facility of selection; and system performance (►Table 1). Four points were allotted to each item: 0 for negative, 2 for neutral, and 4 for positive. After reporting by both the system and free-text method, the radiologists answered the questionnaire. They were informed about the content of the questionnaire before performing the experiment.

2.5 Statistical Analysis

All values are expressed as mean \pm standard error of the mean (SEM). Statistical comparison of input time by the free-text method and by the term selection method was performed by repeated measures of two-way ANOVA using SPSS with significance defined as $P < 0.05$.

3. Results

3.1 System Evaluation

The time to generate a report by free-text was 68 ± 4.6 sec. On the other hand, the time by our term selection method was 59 ± 4.8 sec, which was significantly shorter than the free-text method ($P < 0.05$). This method shortened the time to generate reports by 12.5% compared with the free-text method (►Fig. 7, ►Table 2). Individually, input time was shorter for radiologists A, D, and E, while it was slightly longer for radiologists B and C.

The time it took to generate a report in the first half of reports and the second half of reports by the term selection method were 61 ± 8.5 and 57 ± 3.7 sec, respectively. Those by free-text typing were 67 ± 11 and 69 ± 7.5 sec., respectively.

Each radiologist made 283 ± 15 phrases in 50 reports (5.7 ± 0.3 phrases per report) by using the term selection method and 260 ± 11 phrases in 50 reports (5.2 ± 0.2 phrases per report) by free-text typing. The number of concordant phrases ("true positive") in the reports by both methods was 243 ± 11 in 50 reports. The number of phrases which only existed in the reports by free-text typing was 17 ± 2.2 in 50 reports. The number of phrases which only existed in the reports by the term selection method was 40 ± 9.6 in 50 reports. The F-measure was 90 ± 1.9 %.

We analyzed the 86 phrases which only existed in the manually typed reports. Forty of the 86 phrases (47 %) were modifiers half of which could be entered by the term selection method. Thirty-six of the 86 phrases (42%) were frequently used phrases such as "dilatation of heart" and "aortic elongation", etc. They also could be entered by the term selection method. One

diagnosis was different between the two reports. The diagnosis in the report by term selection method was correct.

The number of terms added by the respective radiologists was 7.2 ± 1.4 for making the 50 reports.

The results of the questionnaire are shown in ►Table 1. The average scores for each category were as follows (0 = negative, 2 = neutral, 4 = positive): 2.4 ± 0.4 for overall assessment, 3.1 ± 0.4 for screen design, flow, and ease of operation, 2.5 ± 0.1 for number of terms and facility of selection, and 2.0 ± 0.5 for system performance.

4. Discussion

Our method shortened the time to generate reports by 12.5% compared with the free-text method. The time to generate a report in the second half of the reports was shortened from that in the first half of them. On the other hand, no significant difference was found between the time to generate a report in the first and the second half of the reports by the free-text typing. Thus if the radiologists had done practices before the experiments, the time to generate a report by our method would be shortened even more. The number of phrases entered for reporting by the term selection method was much more than that by free-text typing. The F-measure value was 90 ± 1.9 %, indicating the phrases generated by our system were reasonably concordant with those made by free-text typing. In F-measure value, a phrase in the report by the term selection method which did not exist in the report by free-text typing was counted as false positive. However, the phrase was not actually "false". For example, diagnosis was sometimes missing in free-text typing. The term selection method prevented this. Thus the actual appropriateness of the report by the term selection method was thought to be more than the appropriateness measured by F-measure value. These results demonstrated that our method did not take more time than the conventional way even though the entered information was not significantly different between the two.

There were some phrases which existed only in the reports generated by free-text

typing. Most of them were modifiers of the basic terms. However, most of the modifiers were predefined in this system. In the latter, although the basic term can be selected easily, additional clicks are needed for entering the modifier. This was the reason why the system frequently lacked modifiers. The frequent phrases of “dilatation of heart” and “aortic elongation” were also often omitted in the report by term selection method. These findings were common abnormal findings in chest x-ray radiography, but not the main ones in the tested radiographies. The radiologists might feel bothered entering these minor findings by several clicks.

In the evaluation method, we set a 10-day interval between the experiments to

erase memories in the radiologists. However, the radiologists might recall the images in the second experiment because the x-ray images were assigned in the same order in both experiments. This factor favored the time to generate reports by free-text typing. On the other hand, the experience of using the system also might be forgotten. This factor might affect the results of the questionnaire.

The results of the questionnaire showed that the score of overall assessment was over neutral (score: 2.4 ± 0.4). One item whose score was under neutral was the question about whether the users could enter what they would like to write for findings (score: 1.8 ± 0.5). This shows the essential weakness of term selection methods

for reporting. Users can add any terms to the system and can add comments to the reports. Although users can enter what they want to write for findings, users would feel bothered if they need to add terms to the system each time while making report. The answers to the question about the preference for making reports were divided: Two radiologists preferred to use this system rather than the free-text typing, and other two preferred to use the free-text typing (score: 2.2 ± 0.6). Most of them, however, were interested in this system (score: 3.2 ± 0.4). The number of terms and their expression were satisfactory (score: 3.2 ± 0.2), and the radiologists did not think that selection of terms from the lists was difficult (score: 2.4 ± 0.2). They were quite satisfied

Table 1 The questions and results of the questionnaire. We provided a questionnaire including 14 questions in four categories: overall assessment; screen layout, flow, and ease of operation; number of terms and facility of selection; and system performance. Four points were allotted to each item: 0 for negative, 2 for neutral, and 4 for positive. After reporting by both the system and free-text method, the radiologists answered the questionnaire.

Question								Score
1. Total assessment								
Are you interested in this reporting system?	Not interested	0	1	2	3	4	Very interested	3.2 ± 0.4
Could you enter what you would like to write for findings?	Dissatisfied	0	1	2	3	4	Very satisfied	1.8 ± 0.5
Which do you prefer for making reports?	Free-text	0	1	2	3	4	This method	2.2 ± 0.6
Average								2.4 ± 0.4
2. Screen layout, flow, and ease of operation								
Were you satisfied with the screen layout and flow?	Dissatisfied	0	1	2	3	4	Very satisfied	3.2 ± 0.2
Did you feel learning the system was easy?	Very difficult	0	1	2	3	4	Very easy	3.6 ± 0.2
Could you operate this system easily and smoothly?	Very difficult	0	1	2	3	4	Very easy	2.4 ± 0.2
Average								3.1 ± 0.4
3. Number of terms and family of selection								
Were you satisfied with the number of prepared terms?	Dissatisfied	0	1	2	3	4	Very satisfied	3.2 ± 0.2
Could you easily find the term you wished to enter?	Very difficult	0	1	2	3	4	Very easy	2.4 ± 0.2
Were the frequent phrases or summary phrases useful?	Quite useless	0	1	2	3	4	Very useful	2.4 ± 0.2
Did you feel it was useful to select category initially?	Quite useless	0	1	2	3	4	Very useful	2.2 ± 0.4
Did you feel it was useful to select terms from the lists of region, finding, and diagnosis?	Quite useless	0	1	2	3	4	Very useful	2.6 ± 0.5
Did you feel it was useful to select prefix and suffix by template?	Quite useless	0	1	2	3	4	Very useful	2.8 ± 0.2
Was the guide function that highlighted candidate terms useful?	Quite useless	0	1	2	3	4	Very useful	2.2 ± 0.4
Average								2.5 ± 0.1
4. System performance								
Were you satisfied with the performance of the system?	Dissatisfied	0	1	2	3	4	Very satisfied	2.0 ± 0.5
Average								2.0 ± 0.5

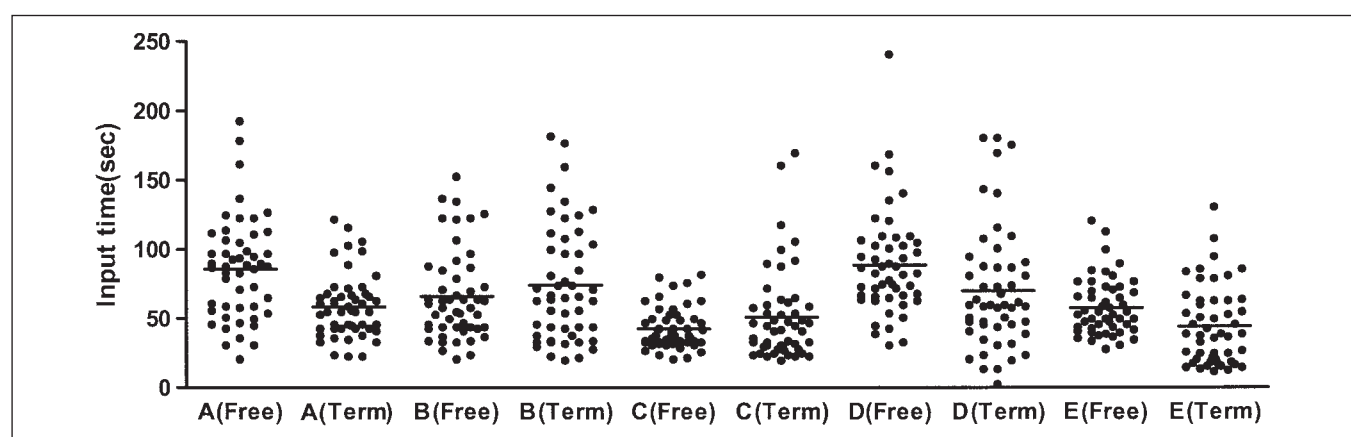


Fig. 7 Comparison of input time of the free-text method (Free) and the term selection method (Term). Dots indicate the time to generate a report. Lines indicate average time for 50 reports. The experience of each radiologist was as follows; A was a first year radiologist, B was a fifth year abdominal radiologist with rare experience in chest x-ray interpretation, C was a fifth year

chest radiologist, D was a seventh year nuclear medicine specialist with rare experience in chest x-ray interpretation, E was a seventh year chest radiologist. The time using the term selection method was shorter than that using the free-text method for radiologists A, D, and E.

with the arrangement of the terms. (score: 3.2 ± 0.2). These results show that making reports by this system was not more troublesome than by free-text method. Up to now structured data entry has been seen to be time-consuming, thus unacceptable in daily clinical practice even though it is beneficial for research or decision support. But our system is applicable to radiology reports of chest x-rays in daily clinical practice and it rather shortened the time to enter reports comparing with free-text typing.

The system particularly benefited radiologists who did not have significant experience in interpreting chest x-rays, as the time to generate reports was shortened and the concordance scores were high. The guide function for term selection also works as decision support, because when a user selects region and finding, the system shows candidate diagnoses. This function may be helpful for beginning radiologists or doctors who are not radiologists. In one case radiologist A issued an incorrect diag-

nostic statement in the free-text typing scenario but his diagnosis was correct using the term selection method.

The system may become well-accepted for each user after a certain period of usage, because terms that users wish to use can be added by themselves and the system acquires the frequency of combination of terms from the stored generated reports. This amendment function was not evaluated in this study.

Table 2 Results of system evaluation in comparison of the free-text method (Free) and the term selection method (Term). Time to generate a report by the free-text method and by the term selection method by each radiologist is presented. The number of phrases existing in the reports generated by both systems (true positive), the number of phrases existing only in the reports generated by the free-text method (false negative), the number of

phrases existing only in the reports generated by the term selection method (false positive) in 50 reports and the value of F-measure in each radiologist are shown in the middle column of this table. The number of terms added to the system by each radiologist is shown in the last column. The identification of radiologists corresponds to those in Figure 7.

Radiologist	Input time (sec)		Concordance score				Number of terms added/ 50 reports
	Free-text	Term selection	true positive (number of phrases existing in both system/ 50 reports)	false negative (number of phrases existing only in free-text/ 50 reports)	false positive (number of phrases existing only in term selection/ 50 reports)	F-measure (%)	
A	86 ± 5.2	58 ± 3.3	244	24	77	83	8
B	66 ± 4.6	74 ± 6.0	270	16	29	92	7
C	42 ± 2.1	51 ± 4.7	219	14	23	92	5
D	89 ± 5.4	70 ± 6.3	265	12	33	92	12
E	58 ± 2.9	44 ± 4.0	218	20	39	88	4
Average	68 ± 4.6	59 ± 4.8	243 ± 11	17 ± 2.2	40 ± 9.6	90 ± 1.9	7.2 ± 1.4

This study was based on a Japanese lexicon (controlled vocabulary). We compared the terms in our lexicon with those in RadLex. Fifteen percent of the terms in our lexicon were not included in RadLex, showing that the lexicon in Japanese is different from that in English even in the domain of radiology reports. Procedure, region, and diagnosis in our lexicon were well matched. The observation objects of bone, soft tissue, pleura and chest wall were well matched also. However, some terms of finding (imaging observation characteristic) in lung and mediastinum could not be correlated with RadLex terms because they weren't defined in the 2.1 version of RadLex. These terms were categorized into Visual Feature in the reports. We believe that the maintenance of RadLex terminology is essential.

This system is applicable to a radiology report of chest x-ray, because it is composed of simple sentences. CT or MRI reports, however, tend to be more complicated, especially when describing tumors. The information model adopted in this system should be revised for these complicated reports and a more advanced system is required.

Our system allows a radiologist to generate a structured report at a detailed level. The structured report can be used as a means of indexing to facilitate image retrieval, which will allow similar images to the image being diagnosed to be retrieved from databases according to the reports. The reports of other advanced examinations such as CT for the same patient will be helpful for diagnosis. Such collections of similar images will also contribute to the education of radiographic interpretation. Furthermore, a clinical decision support system will be able to interpret the reports made by our system in order to extract certain problems of a patient (e.g. suspect of interstitial pneumonitis).

This system will also enhance the potential for standardized reporting. By using this system, the terms necessary for radiology reports can be collected automatically, which will be helpful to build a standard lexicon. Although the structured reports output from our system are not a standard form, it is not difficult to transform it to the DICOM SR or Clinical Document

Architecture in Health Level 7 formats [41–43].

6. Conclusions

In order to realize structured reporting, we developed a system that generates radiology reports for chest radiography by a term selection method. We constructed an information model for it and categorized five observation objects. Modifiers could be entered using templates linked to each term. The system could guide users to select terms by highlighting associated terms. This method decreased the time to generate reports, and the sentences made by this system were well concordant with those by the free-text method. The results of the questionnaire showed that our system is applicable to radiology reports of chest x-rays in daily clinical practice. These results demonstrated that this method was feasible for generating structured reports.

Acknowledgments

We thank Kayoko Isohashi, Masahiro Yanagawa, Atsushi Nakamoto, and Akihiro Tsukabe for giving advice on system design and evaluation.

References

1. Tierney WM, Overhage JM, McDonald CJ. Toward electronic medical records that improve care. *Ann Intern Med* 1995; 122: 725–726.
2. Langlotz CP. Automatic structuring of radiology reports: harbinger of a second information revolution in radiology. *Radiology* 2002; 224: 5–7.
3. Cheung NT, Fung V, Chow YY, Tung Y. Structured data entry of clinical information for documentation and data collection. *Medinfo* 2001; 10: 609–613.
4. Langlotz CP, Meiningner L. Enhancing the expressiveness and usability of structured image reporting systems. *Proc AMIA Symp* 2000. pp 467–471.
5. Williamson KB, Steele JL, Gunderman RB. Assessing radiology resident reporting skills. *Radiology* 2002; 225: 719–722.
6. Moorman PW, van Ginneken AM, van der Lei J, Van Bommel JH. A model for structured data entry based on explicit descriptive knowledge. *Methods Inf Med* 1994; 33 (5): 454–463.
7. Los RK, van Ginneken AM, van der Lei J. OpenSDE: A strategy for expressive and flexible structured data entry. *Int J Med Inform* 2005; 74: 481–490.
8. Association National Electrical Manufactures: Digital Imaging and Communication in Medicine (DICOM) 3.0, supplement 23 DICOM Structured Reporting. Last accessed February 16, 2009. Available at: ftp://medical.nema.org/Dicom/final/sup23_ft.pdf.
9. Rada H, Uwe E, Andre S. DICOM Structured Reporting Part 1. Overview and Characteristics. *RadioGraphics* 2004; 24: 891–896.
10. Graig A, Usha S, Ricky T. Structured Reporting in Neuroradiology. *Ann N Y Acad Sci* 2002; 980: 259–266.
11. Luyin Z, Kwok PL, Jingkun H. Generating XML Schemas for DICOM Structured Reporting Templates. *J Am Med Inform Assoc* 2005; 12: 72–83.
12. Rita N. Benefits of the DICOM Structured Report. *J Digit Imaging* 2006; 19: 295–306.
13. Kamau AW, Duvall SL, Liimatta AP, Wiggins RH, 3rd, Avrin DE. Translating the IHE Teaching File and Clinical Trial Export (TCE) Profile Document Templates into Functional DICOM Structured Report Objects. *J Digit Imaging* 2007, Sep 4.
14. RadLex, Radiology Society of North America (RSNA). Available at <http://www.radlex.com>. Accessed March 2008.
15. Langlotz CP. RadLex: a new method for indexing online educational materials. *Radiographics* 2006; 26 (6): 1595–1597.
16. Marwede D, Schulz T, Kahn T. Indexing Thoracic CT Reports Using a Preliminary Version of a Standardized Radiological Lexicon (RadLex). *J Digit Imaging* 2007; 28: 1–8.
17. van Ginneken AM. Structured data entry in ORCA: the strengths of two models combined. *Proc AMIA Symp* 1996. pp 797–801.
18. Johnson SB, Bakken S, Dine D, Hyun S, Mendonça E, Morrison F, Bright T, Van Vleck T, Wrenn J, Stetson P. An electronic health record based on structured narrative. *J Am Med Inform Assoc* 2008; 15: 54–64.
19. O'Connell BT, Cho C, Shah N, Brown KM, Shiffman RN. Take note(s): differential HER satisfaction with two implementations under one roof. *J Am Med Inform Assoc* 2004; 11: 43–49.
20. Friedman C. A broad-coverage natural language processing system. *Proc AMIA Symp* 2000. pp 270–274.
21. Taira RK, Soderland SG. A statistical natural language processor for medical reports. *Proc AMIA Symp* 1999. pp 970–974.
22. Friedman C, Alderson PO, Austin JH, Luyin Z, Kwok PL, Jingkun H. A general natural-language text processor for clinical radiology. *J Am Med Inform Assoc* 1994; 1 (2): 161–174.
23. Taira RK, Soderland SG, Jakobovits RM, et al. Automatic structuring of radiology free-text reports. *RadioGraphics* 2001; 21: 237–245.
24. Dang PA, Kalra MK, Blake MA, et al. Use of Radcube for Extraction of Finding Trends in a Large Radiology Practice. *J Digit Imaging* 2008; 16 (June 10).
25. Nishimoto N, Terae S, Uesugi M, Ogasawara K, Sakurai T. Development of a medical-text parsing algorithm based on character adjacent probability distribution for Japanese radiology reports. *Methods Inf Med* 2008; 47 (6): 513–521.
26. Ceusters W, Spyns P, De Moor G. From syntactic-semantic tagging to knowledge discovery in medical texts. *Int J Med Inform* 1998; 52 (1–3): 149–157.
27. Nardon FB, Moura LA. Knowledge sharing and information integration in healthcare using ontologies and deductive databases. *Stud Health Technol Inform* 2004; 107 (Pt 1): 62–66.

28. Baneyx A, Charlet J, Jaulent MC. Methodology to Build Medical Ontology from Textual Resources. *AMIA Annu Symp Proc* 2006. pp 21–25.
29. Qamar R, Kola J, Rector AL. Unambiguous data modeling to ensure higher accuracy term binding to clinical terminologies. *AMIA Annu Symp Proc* 2007; 11: 608–613.
30. Lassila O, Swick R. World Wide Web consortium: Resource description framework (RDF) model and syntax specification, 1999. Last accessed December 18, 2008. Available at: <http://www.w3.org/TR/PR-rdf-syntax>.
31. Berners-LT, Hendl J, Lassila O. The semantic web. *Scientific American* 2001; 5: 34–43.
32. Berners LT. Information management: A proposal (The original proposal of the WWW; March 1989, May 1990). Last accessed February 16, 2009. Available at: <http://www.w3.org/History/1989/proposal.rtf>.
33. Prud'hommeaux E, Seaborne A. World Wide Web consortium: SPARQL Query Language for RDF, 2008. Last accessed July 1, 2009. Available at: <http://www.w3.org/TR./rdf-sparql-query>.
34. Boulos MN, Roudsari AV, Carson ER. Towards a semantic medical Web: HealthCyberMap's tool for building an RDF metadata base of health information resources based on the Qualified Dublin Core Metadata Set. *Med Sci Monit* 2002; 8 (7): MT124–36.
35. Eclipse.org: Eclipse universal tool platform. Last accessed February 16, 2009. Available at: <http://www.eclipse.org>.
36. Eclipse Foundation, Eclipse Graph Editing Framework: Last accessed February 16, 2009. Available at: <http://www.eclipse.org/gef>.
37. PostgreSQL, PostgreSQL Global Development Group: Last accessed February 16, 2009. Available at: <http://www.postgresql.org>.
38. Jena semantic web framework, Sourceforge.net: Last accessed February 16, 2009. Available at: <http://jena.sourceforge.net>.
39. Hripcsak G, Rothschild AS. Agreement, the f-measure, and reliability in information retrieval. *J Am Med Inform Assoc* 2005; 12 (3): 296–298.
40. Roberts A, Gaizauskas R, Hepple M, et al. The CLEF corpus: semantic annotation of clinical text. *AMIA Annu Symp Proc* 2007. pp 625–629.
41. Robert HD, Liora A, Sandy B, Calvin B. HL7 clinical document architecture, release 2.0, 2004. Health Level Seven. Last accessed February 16, 2009. Available at: <http://xml.coverpages.org/CDA-20040830v3.pdf>.
42. Gerdson F, Müller S, Jablonski S, Prokosch HU. Standardized exchange of clinical documents, towards a shared care paradigm in glaucoma treatment. *Methods Inf Med* 2006; 45 (4): 359–366.
43. Klein A, Prokosch HU, Müller M, Ganslandt T. Experiences with an interoperable data acquisition platform for multi-centric research networks based on HL7 CDA. *Methods Inf Med* 2007; 46 (5): 580–585.