検査介助をする．検査が短時間で終了しない場合，移送担当者は病棟へ戻り別業務を行い，検査室からの迎え連絡を受けて検査室へ移動する．検査終了後は検査室から患者を申し受け，移送ケア用具へ移乗し，病棟まで移送し，病室のベッドへ再び移乗する．ME 機器や装着中の医療物品をベッド上生活可能なよう整え，患者の状態をアセスメントした後，診察券，カルテといった持参物品を片づけ，移送記録を行う．アクティビティ図により患者移送業務プロセスが 47 のタスクより構成されることが明らかとなった．

アクティビティ図で示した 47 のタスク別に 4 病棟各 1 日間の所要時間合計，発生件数，平均値，中央値，標準偏差，範囲を示す（表 1）．合計所要時間量が最も多いものは，「T29 患者を移送する」（9：15：49）であり，1 件あたり平均 5 分患者移送を行っている．記録された患者移送 213 件中 109 件が実際の患者の移送を伴うものであった．患者移送を伴わない患者移送業務とは，独歩患者への対応や検査予定時間調整のみの業務であった．次いで所要時間量が多いものは「T36 目的地で待機する」（1：57：19）であり，さらに「T28 患者を移乗する」（1：46：59）であった．一方，「T06 ケア実施者を特定する」「T12 ケア実施者を待つ」「T15 ケア実施者を変更する」といった移送担当者の探索や変更に関するタスクは発生件数が少なかった．変動係数を比較すると，「T41 カルテを片づける」「T16 看護師に検査情報を伝える」「T36 目的地で待機する」「T21 状態をアセスメントする」で変動係数が高く，「T29 患者を移送する」や「T43 移送ケア用具を片づける」では変動係数は相対的に低かった．

業務分類別所要時間を表 2 に示す．直接業務とは直接的に患者に対して行う業務であり，間接業務とは直接業務のための準備や片づけを含む，患者に直接接することなく行う業務である．患者移送の約 60% を直接業務が占め，間接業務は 14% 程度であった．

5. 考　　察

タイムスタディデータを利用しオブジェクト指向に基づく業務プロセス可視化により第一に，業務責任者の所在と役割が明らかとなった．機能的側面から患者移送業務の主たる担い手は看護師であるが，連絡調整に医療クラーク，ME 機器や輸液の留置・装着を伴わない患者の移送に看護助手が関与することが示された．医療クラークは検査の連絡を受けカルテにて移送ケア方法を確認するものの，実施者の変更や業務委任の権限は無く，リーダー看護師に委ねられている点が明らかとなった．さらに病棟連絡担当者は独歩患者については，連絡担当者が医療クラークか看護師であるかに関わらず患者への検査情報の伝達まで責任を負うことが示された．また，護送，担送患者の場合連絡を受けてから担当者に検査情報を申し送るまでの責任を負い，連絡担当者がクラークの場合，ケア実施者の変更権限がないため，リーダー看護師へ業務を委ねることが示された．このプロセスで最も時間を要するタスクは「患者への検査情報の伝達」であり，次いで「カルテの準備」「診察券の準備」であった．診察券の利用は入院患者の外来受診，検査受診時に限られており，収納場所も固定されているのに対し，カルテは医師，看護師，医療クラーク，その他医療従事者の複数が様々な用途で利用するためカルテ探しが発生しており，カルテ準備の所要時間が長くなっていた．連絡担当者から移送担当者へ情報伝達がなされた後，移送記録まで移送担当者が全ての責任を持つことが分かった．

第二に広く臨床で用いられている業務手順書と実際の業務プロセスの乖離が示された．対象病院の業務手順書では，「患者を探す」「移送担当者を探す」「移送担当者を変更する」「（検査室で）検査準備をする」「検査介助をする」といったタスクは明示されていない．この理由として業務手順書が標準的手順について書かれており，業務プロセスで注目すべきイレギュラーな事象や重複作業を念頭に置いていない点が考えられる．また，業務手順書は個人の看護師の業務手順として書かれており，先述したような業務責任者の所在と役割が明確ではない．本研究で実際の業務記録に基づく業務プロセス描画により，

第三にプロセスに時間情報を付加することで，業務の稼働効率が示された．稼働効率は業務プロセス改善で最も注目すべき点である．タスク別に所要時間及び時間のばらつきが示されたことにより，患者移送ケアを構成する時間要素が明らかとなった．今後制約条件により所要時間がどのように変化するか詳細に検討していくことが求められる．

第四にリスク分析が可能となった点が挙げられる．本研究により最終的に患者移送を構成する業務タスクとして 47 タスクが抽出され，タイムスタディ記録によりその順序関係が明らかとなった．これにより，各タスクの入力と出力が明確化され，またイレギュラー例の頻度も明らかとなった．「患者を探す」「看護師を探す」といったイレギュラー事象は今回の調査で記録された業務の目標達成を阻害するリスクと言えよう．今後タスクひとつひとつに注目し，それぞれの出力を阻害する因子を明らかとすることで，患者移送業務の抱えるリスクをタスク毎に抽出することが可能であり，安全やケアの質向上といった議論が可能になると考える．

6. 今 後 の 展 望

本研究により患者移送業務構造が可視化された．患者移送業務は患者の状態や検査の種類，業務発生時間により扱うオブジェクト，プロセス，時間効率が大きく異なることが示唆された．また，業務発生が不定期であることが多く，かつ迅速な対応を要するため，看護師は他業務との調整を図りつつ患者移送業務を遂行しなければならないことが明らかとなった．

表1　患者移送プロセス別業務量

タスク項目		TOT	件数	平均値	SD	範囲	CV
T01	検査時間を調整する	0:33:27	28	71.7	65.7	(5-273)	0.92
T02	検査予定を書類で確認する	0:05:24	10	32.4	28.0	(4-100)	0.86
T03	検査呼出しを受ける	0:31:30	45	42.0	51.0	(1-324)	1.21
T04	カルテを探す	0:04:32	11	24.8	21.6	(2-64)	0.87
T05	安静度を確認する	0:09:11	10	55.1	58.6	(6-186)	1.06
T06	ケア実施者を特定する	0:00:58	3	19.3	13.6	(4-32)	0.70
T07	地図を準備する	0:08:27	20	25.4	16.0	(3-70)	0.63
T08	診察券を準備する	0:14:37	31	27.1	23.6	(1-108)	0.87
T09	カルテを準備する	0:28:41	42	41.0	38.9	(5-187)	0.95
T10	ケア実施者を探す	0:01:59	3	39.7	22.6	(16-60)	0.57
T11	患者を探す	0:07:33	11	41.2	45.4	(4-116)	1.10
T12	ケア実施者を待つ	0:00:21	1	21.0			
T13	患者に検査情報を伝える	0:29:55	43	41.8	32.7	(1-144)	0.78
T14	患者に必要物品を渡す	0:00:21	3	7.0	5.6	(2-13)	0.80
T15	ケア実施者を変更する	0:00:37	1	37.0			
T16	看護師に検査情報を伝える	0:26:48	38	42.3	77.7	(1-384)	1.84
T17	フィルムを準備する	0:00:44	2	22.2	10.3	(15-29)	0.46
T18	持参物品を準備する	0:04:02	3	80.7	102.9	(6-198)	1.28
T19	移送ケア用具を準備する	0:22:38	46	29.5	31.2	(1-139)	1.06
T20	移送ケア用具を搬入する	0:21:27	40	32.2	22.4	(1-88)	0.70
T21	状態をアセスメントする	0:24:48	17	87.5	128.3	(2-382)	1.47
T22	患者氏名を確認する	0:02:45	10	16.5	7.8	(6-30)	0.47
T23	ME機器の可動準備をする	0:13:50	19	43.7	49.8	(7-237)	1.14
T24	医療物品の可動準備をする	0:16:43	23	43.6	35.3	(2-117)	0.81
T25	排泄を援助する	0:05:16	5	63.3	54.5	(10-152)	0.86
T26	更衣を援助する	0:12:35	19	39.7	36.2	(10-127)	0.91
T27	環境を整える	0:10:22	13	47.8	52.3	(5-199)	1.09
T28	患者を移乗する	1:46:59	83	77.3	94.8	(3-707)	1.23
T29	患者を移送する	9:15:49	109	306.0	162.5	(1-866)	0.53
T30	検査受付をする	0:08:56	34	15.8	19.3	(1-90)	1.22
T31	患者を受け渡す	0:01:55	8	14.4	10.3	(2-34)	0.72
T32	必要物品を受け渡す	0:10:31	30	21.0	19.1	(1-89)	0.91
T33	情報を受け渡す	0:33:09	31	64.2	42.0	(3-156)	0.65
T34	検査準備をする	0:27:16	26	62.9	84.2	(1-370)	1.34
T35	検査介助する	0:42:01	41	61.5	60.9	(6-255)	0.99
T36	目的地で待機する	1:57:19	35	201.1	299.6	(1-1612)	1.49
T37	患者を申し受ける	0:06:37	7	56.7	74.2	(6-208)	1.31
T38	ME機器を再装着する	0:41:25	18	138.1	184.4	(6-766)	1.34
T39	医療物品を再装着する	0:21:23	14	91.7	94.2	(2-396)	1.03
T40	診察券を片づける	0:04:35	23	12.0	10.2	(1-44)	0.85
T41	カルテを片づける	0:18:34	30	21.8	40.7	(1-214)	1.87
T42	フィルムを片づける	0:00:28	4	7.0	6.1	(3-16)	0.87
T43	移送ケア用具を片づける	0:25:52	40	38.8	25.3	(2-115)	0.65
T44	地図を片づける	0:01:54	5	22.8	31.6	(1-78)	1.39
T45	その他の片づけをする	0:13:24	15	53.6	47.3	(1-159)	0.88
T46	移送を記録する	0:11:10	11	60.9	83.0	(3-247)	1.36
M	移動する	4:36:03	119	139.2	150.9	(2-1068)	1.08

TOT: Time on Task（合計所要時間量），SD: 標準偏差，CV: 変動係数，平均値，標準偏差，範囲の単位は秒である

表2　業務分類別所要時間

分類	タスク数	TOT	(%)
間接業務	21	3:56:23	(14.1)
直接業務	21	16:08:27	(58.0)
情報伝達	2	0:59:57	(3.5)
待機	1	1:57:19	(7.0)
記録	1	0:11:10	(0.6)
移動	1	4:36:03	(16.5)
合計	47	27:49:19	(100.0)

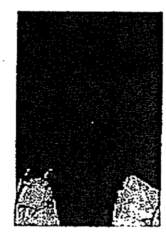
TOT: Time on Task(合計所要時間量)

業務量のみならず業務構造や業務プロセスを明示化するタイムプロセススタディの有用性が示された．今後他業務や複数の対象施設の業務記録に基づき同様の研究を行うことにより本研究の応用可能性を確認していく．

謝辞　本研究の実施にあたりご協力いただきました看護師，看護助手，医療クラークの皆さんに改めて心より御礼申し上げます．また調査者として大阪大学大学院医学系研究科数理保健学研究室ゼミ生の皆様にご協力頂きました．ここに記して感謝申し上げたいと思います．

文　　献

1. Van de Werf E, Lievens Y, Verstraete J, et al. : Time and motion study of radiotherapy deliverly: economic burden of increased quality assurance and IMRT. Radiother Oncol. **93**: 137-140, 2009.
2. Were MC, Sutherland JM, Bwana M et al. : Patterns of care in two HIV continuity Clinics in Uganda, Africa: a time motion study. AIDS care. **20**(6): 677-682, 2008.
3. 高橋陽子, 栗原真弓, 美原盤: 神経難病と脳血管障害後遺症患者に対する直接看護業務の検討. 看護. **59**(2): 87-93, 2007.
4. Tang Z, Weavind L, Mazabob J et al. : Workflow in intensive care unit remote monitoring: a time and motion study. Crit Care Med. **35**(9): 2057-2063, 2007.
5. 村野大雅, 小林寛伊: タイムスタディによるICUの看護業務, 動線分析. 手術医学. **30**(3): 284-287, 2009.
6. 横内光子, 大野ゆう子, 笠原聡子他: 業務スケジューリングからみた看護業務属性の検討. 生体医工学. **43**(4): 762-768, 2005.
7. 白石義人, 中川智永子, 森田耕司他: 看護師の業務量調査から見た手術室の効率的運用. 手術医学. **30**(2): 155-158, 2009.
8. 上野正文: 手術室適正人員配置のための新たな発想. 手術医学. **30**(2): 116-121, 2009.
9. Yen K, Shane EL, Pawar SS, et al. : Time motion study in a pediatric emergency Department before and after computer physician order entry. Ann Emerg Med. **53** (4): 462-468, 2009.
10. Finkler SA, Knickman JR, Hendrickson G, et al. : A Comparison of work-sampling and time and motion techniques for studies in health services research. Health Service Res. **28**(5): 577-597, 1993.
11. Shiki N, Ohno Y, Fujii A, et al. : Time process study with UML a new method for process analysis. Methods Inf Med. **48**(6): 582-588, 2009.
12. Eriksson HE, Penler M: Business modeling with UML. Wiley, New Jersey, 2000.



清水　佐知子（シミズ　サチコ）
大阪大学医学部保健学科看護学専攻卒業．神戸大学大学院国際協力研究科地域協力政策専攻博士前期，後期課程．国立保健医療科学院政策科学部研究員を経て，大阪大学大学院医学系研究科保健学専攻助教．数理保健学研究室所属．

Linguistic Analysis of Large-Scale Medical Incident Reports

KATSUHIDE FUJITA

School of Engineering, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-0021, Japan,

fujita@ipr-ctr.t.u-tokyo.ac.jp

MASANORI AKIYAMA

Policy Alternatives Research Institutes, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan,

makiyama@pp.u-tokyo.ac.jp

KEUNSIK PARK

Osaka City University Hospital, 1-5-7 Asashi, Abeno-ku, Osaka, 545-8586, Japan, kspark@msic.med.osaka-cu.ac.jp

ETSUKO (NAKAGAMI) YAMAGUCHI

Osaka City University Hospital, 1-5-7 Asashi, Abeno-ku, Osaka, 545-8586, Japan, melano@skyblue.ocn.ne.jp

HIROYUKI FURUKAWA

Director of Pharmaceutical Service, Yamaguchi University Hospital, 1-1-1, Minamikogushi, Ube, 755-8505, Japan,

sambista-knz@umin.ac.jp

ICHIRO SAKATA

Policy Alternatives Research Institutes, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan,

isakata@ipr-ctr.t.u-tokyo.ac.jp

Abstract

The analysis of medical incident reports is indispensable for the patient safety management. Most of the incident reports include some free composition formats, therefore, the analysis of free descriptions gives new perceptions. We aimed to accumulate, to interpret information again by structured incident information, and to clarify the improvement point for safe medical treatments in this study.

We employ text mining techniques to the analysis of medical incident reports in this paper. The text mining can find various relationships that are not only direct relationships but also indirect relationships. First, some important characteristic words were extracted in three sections of the accident's details, responses and solutions using TF-IDF measures. By using the TF-IDF, we can get some important characteristic words for analyzing the reports. In addition, we show the co occurrence networks using these extracted words.

In addition, the class of medical treatments and the class of operations are granted. In the medical incident repots categorizing, a major direction of the medical incident analysis is the assimilation of the existing top-down type class grants by specialists. In this study, we evaluated existing categories using the bottom-up analysis and discussing the differences between existing classes granted the top-down analysis. There is a need to investigate whether classes granted here are in accord with the characteristics of each document item. For the bottom-up analysis, we generated a network of documents so that documents belonging to identical classes relating to the identical treatments and operations are grouped together the most. In this network of documents, nodes mean the documents of the medical incident reports and links means high similarities between two documents. After that, we conducted the network clustering in order to carry out labeling of each document based on networks.

By the analysis based on text mining and network clustering, the language networks with the hub of the word "confirmation," thereby revealing that inadequate confirmations on the drug labels, instructions of a physician and patient were very significant causes of accidents. These results suggest the effectiveness of introducing the network analysis method. In addition, the class of patient managements regarding patients' fallings in top-down analysis is created clearly. On the other hand, some categorizes by top-down analysis don't reflect the category by the bottom-up analysis.

Keywords: Medical Safety Management, Incident Reports Analysis, Network Analysis, Natural Language Processing

1. Introduction

"In the shadow of every serious accident, there exist 29 times more minor accidents and 300 times more near misses." This principle was published in 1929 by Herbert William Heinrich, an assistant manager in the technology and research division of an American insurance company (Heinrich, 1931). This principle, which hits home the nature of the occurrence of accidents, is taken up in various fields, such as the study of failure, safety engineering, cognitive psychology as well as the study of reliability, and the incident analysis of minor accidents associated with this is recognized as being important in preventing accidents.

Also, the use of information pertaining to medical accidents is important when implementing medical safety management. The medical safety mechanism of WHO aims to prevent accidents by reusing incident reports through the introduction of IT and Management technologies. Harvard University is engaged in the standardization for the collection of medical accident reports and accident information in the risk management consortium. In England, the National Health Service conducts the medical accident/incident report collection project. Even in Japan, the Ministry of Health, Labour and Welfare began the project to Collect Medical Near-Miss/Adverse Event Information in 2001 (JCQHC, 2009). Through this project, the Ministry conducts analyses based on the collected incident reports.

On the other hand, regarding patient safety management, guidelines for the future deployment of incident analysis are set out in WHO's International Classification of Patient Safety (WHO, 2009). ICPS states the necessity of first investigating the adequacy of classes of incident case studies such as those mentioned above, and second, methods of expressing incidents that adequately reflect these classes, i.e., it states the necessity of ontological construction.

In the Medical Near-Miss/Adverse Event Information including "Details of solutions (Detail)," "Act after incidents by medical staff and patient progress (Response)," and "Problem to solve (Solution)" for a single case are described using a free composition format. In this paper, we analyze the large number of medical incident repots (more than 15,000 reports) provided by Osaka City University using the

Natural Language Processing (NLP) and the Network Analysis (Manning et. al, 2002; Rasmussen, 1992).

In addition, in each case study the class of treatment and the class of operation are granted. There is a need to investigate whether classes granted here are in accord with the characteristics of each document item. In order to achieve the above, this study uses the network clustering (Newman, 2004).

The remainder of this paper is organized as follows. First, we describe the dataset of the medical incident reports provided by Osaka City University. Next, we describe the methodology based on the Natural Language Processing and the Network analysis for analyzing the large number of medical incident reports. Then, we present and discuss the results of analysis of incident reports. Finally, we present our overall conclusions.

2. Medical Incident Reports in Osaka City University Hospital

With increasing social demand for the prevention of medical accidents, the Health, Labour and Welfare Ministry started the Project to Collect Medical Near-Miss/Adverse Event Information from 2001 in order to collect and analyze incident case studies and to provide information conducive to medical safety, such as measures for improvements. When the project was first started, a framework was in place in which the Pharmaceuticals and Medical Devices Agency collected incident case studies from participating medical institutions and then reported these case studies to the Health, Labour and Welfare Ministry, following which a Health, Labour and Welfare Ministry study group conducted aggregate calculations and analysis. The 1st–10th collection of incident case studies were conducted following this framework, and information based on these collected incident case studies was provided by the Health, Labour and Welfare Ministry. From 2004, the Japan Council for Quality Health Care took over the collection of incident case studies, collecting case studies from the 11th collection (JCQHC, 2009). Doctors, nurses and pharmacists in some hospitals are obligated to submit the incident reports when some near-miss/adverse events are happened.

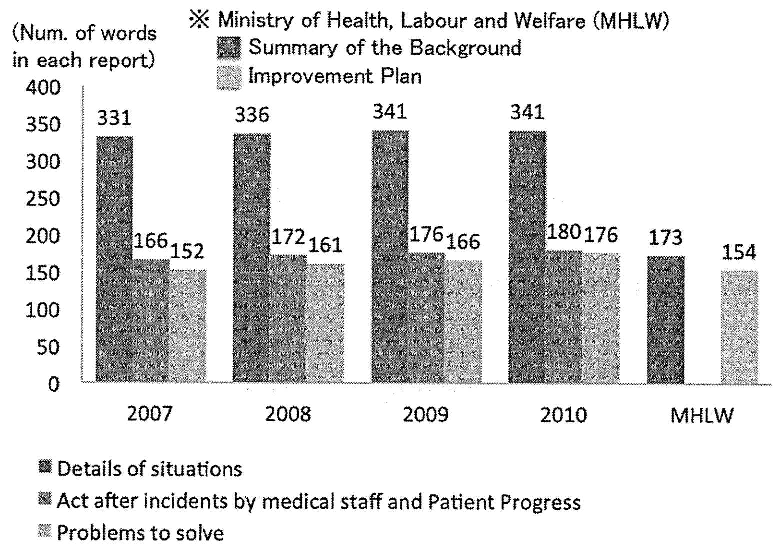


Figure 1: The number of words in each reports between Osaka City University Hospital datasets and MHLW datasets

Osaka City University also collected 18,340 incident reports from 2007 to 2010. In the incident reports provided by Osaka City University, free composition formats are taken quite seriously compared with ones provided by other Hospitals. For instance, the average number of words in the incident reports by Osaka City University is 341 words, on the other hand, the one by the Project to Collect Medical Near-Miss/Adverse Event Information is 173 words in 2010, as Figure 1 showing. In fact, doctors and nurses in Osaka City University have to input the reports for the free descriptions at first because of the Layout of data entry screen.

There are three sections for reporting the incident events: "Details of solutions (Detail)," "Act after incidents by medical staff and patient progress (Response)," and "Problem to solve (Solution)." These sections are different from the ones of Ministry of Health, Labour and Welfare. The combination of "Detail" and "Response" in Osaka City University is a same section as "Summary of Background" in the Ministry of Health, Labour and Welfare. With regard to the class of treatment, there are six classes of incident reports: Medicine and Blood Transfusion, Medical Equipment and Route, Patient Management, Examination, Treatment and Intervention, and Others.

When describing accidents in a free composition format, the reporter makes every effort to include every single circumstance. We can say that extracting important information from these circumstances means creating a foothold for a bottom-up type of ontological construction. Results obtained from this and links with classes granted top-down is in accordance with the future guidelines for incident analysis sought by ICPS.

3. Methodologies of Natural Language Processing and Network Analysis

3.1 Methodologies for analyzing the incident reports

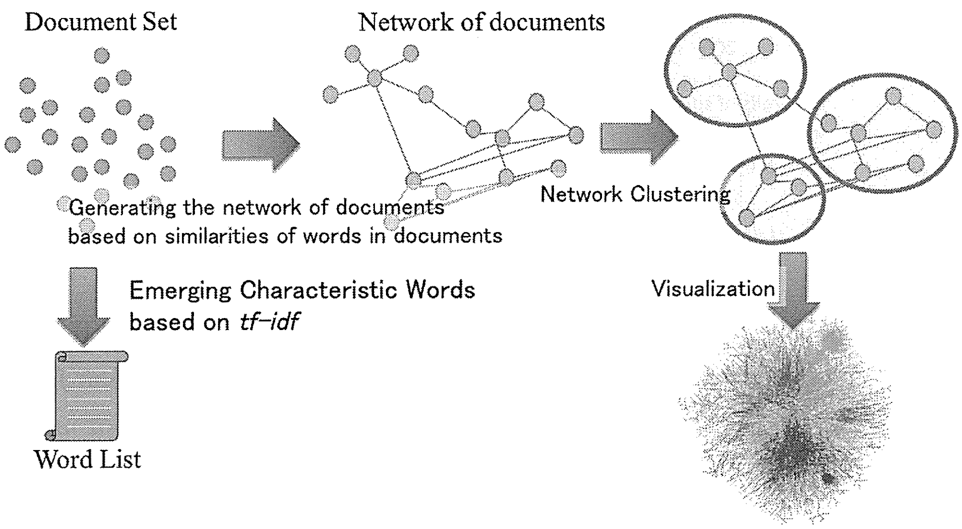


Figure 2. Methodology of analyzing the incident reports

In this paper, natural language processing was first conducted on the incident reports. Keywords that emerge characteristically were then extracted for each category of "Detail," "Response," and "Solution" using the *tfidf* method (Salton, 1983). After that, the semantic tendency of the incident report was investigated in order to create a network of words by calculating the co-occurrence information (Cohen, 2005; Takeda, 2006) of the characteristic words using best 30 based on *tf-idf*.

Also, we show the networks among each document, which are determined by the similarities between documents based on Jaccard coefficient (Jaccard, 1902; Jaccard 1912; Rasmussen, 1992). In generating the networks of documents, we selected best 100 characteristic words based on *tf-idf* measure that can be used in determining links. For determining the links between nodes, we used as a method for finding links from the degree of similarities between words in documents. After that, we conducted the network clustering to them. By the network clustering, the groups which means the categories based on the bottom-up analysis are emerged. Figure 2 shows the workflow of the methodology for analyzing the medical incident reports.

3.2 Japanese language morphological analysis

In the first stage of preprocessing, we conducted morphological analysis in order to break down reports into words. Morphological analysis is a method used to delimit each word in the text where words are not delimited by spaces, such as in languages like Japanese (Manning, 2002). In this research we used MeCab, one of the most common engines for conducting morphological analysis (MeCab, 2006).

In the documents, nominalized verbs, general nouns, and proper nouns were targeted. Focusing solely on nouns is the method generally used in extracting characteristic words. Moreover, in the case of official documents in Japanese, as many of the verbs are nominalized, a lot of information can be obtained regarding action even if using only nouns.

3.3 TF-IDF method

In this study, we calculated a value called *tf-idf* from the frequency of occurrence and conducted filtering based on this values. *Tf-idf* is one of the most widely used indices in extracting characteristic words for document classes and in cases where a certain word occurs several times in a small number of documents, it is defined so as to enlarge that value (Salton, 1983). *Tf-idf* is calculated as follows:

$$tfidf(t,d) = tf(t,d) \times idf(t) - (1)$$
$$tf(t,d) = n(t) / \sum_{k \in T} n(k) - (2)$$
$$idf(t) = \log |D| / \{d : d \in t\} - (3)$$

Here, t is a term, d is a document, *n(t)* is the frequency of occurrence of term *t*, $|D|$ is the total number of documents, and is *{d:d\in t}* the number of documents in which word t occurs. *T* means the set of terms.

The *tfidf* of general words occurring in a large number of documents has a tendency to be of a low value, although words among even general words that have an abnormally high *tf* in some cases exceed the filter effect of *idf* and assume a high value.

3.4 Creation of co-occurrence networks

The co-occurrence index is generally used as a method for finding links between characteristic words in set of documents. Here, the simplest co-occurrence index for finding links between the two characteristic words w_A and w_B is the number of co-occurrence $|w_A \cap w_B|$. Here, $|w_A \cap w_B|$ is the number of documents that exist in both w_A and w_B, $|w_A \cup w_B|$ is the number of documents that exist in either w_A or w_B. In cases where $|w_A \cap w_B| / |w_A \cup w_B|$ exceeded the threshold value (0.1 in this study) then we treated those two words have a link.

3.5 Network of documents based on the Jaccard Similarity

Network analysis is an extremely effective method of looking at the links between documents (Kajikawa, 2007). By conducting network analysis, the discovery of hidden links between two nodes can be expected. In cases where links between only two documents are considered, even if there are no links, there are instances where overall links can be discovered by creating networks.

The co-occurrence index is generally used as a method for finding links from the degree of similarities between words in documents. Here, the simplest co-occurrence index for finding links between the two documents A and B is the number of words' co-occurrences between two documents. Here, $|A \cap B|$ is the number of characteristic words that exist in both document A and B. If considered with only $|A \cap B|$, there are problems such as including as many characteristic words as in long texts and links with other documents being displayed as high. In this paper, we used the Jaccard coefficient (Jaccard, 1902; Jaccard 1912; Rasmussen, 1992) as the similarities measure.

$$\text{Jaccard Coefficient: } |A \cap B| / |A \cup B| \quad \text{-(4)}$$

($|A \cap B|$: The number of words that exist in both Document A and Document B
$|A \cup B|$: The number of word that exist in either Document A or Document B)

In generating the networks of documents, we selected best 100 characteristic words based on *tf-idf* measure that can be used in determining links. A link is established between the two documents in the event that these indices exceed the threshold value. The network changes acutely depending on which of the above indices are selected and how the threshold value is established. As the aim of this study is to investigate the extent to which top-down type of classes are reflected in links in document content that is sought bottom-up, when forming networks, we selected the index and determined the threshold values so as to reflect most the given classes.

3.6 Network Clustering

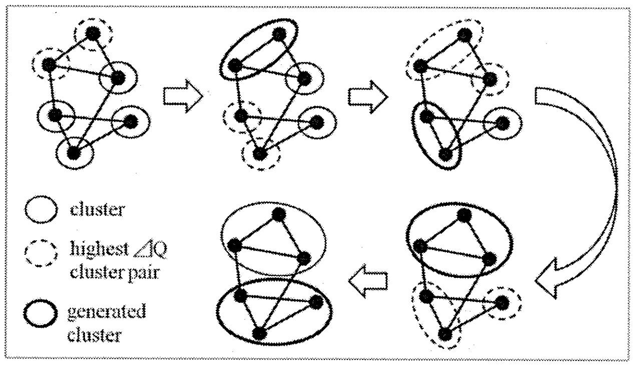


Figure 3. Newman Clustering

In this study, other than given classes, we conducted clustering using the Newman method (Newman, 2004) in order to divide all documents into some groups. Figure 3 shows that method of Newman Clustering. The Newman method is a widely used method for clustering networks. It can be applied even if the number of clusters is unknown; in recent years, it is also being widely applied in large-scale network analysis, such as SNS and blog, due to it being scalable in regard to increases in node numbers. As shown in formula (5), clustering was conducted by maximizing modularity Q, an index for evaluating the modularity of defined networks.

$$Q = \sum_i (e_{ii} - a_i^2) \quad \text{---(5)}$$

$$a_i = \sum_j e_{ij} \quad \text{---(6)}$$

Here, element e_{ij} of line e represents the fraction of the total number of edges of the number of edges that connect cluster i with cluster j and e_{ii} is the fraction of edges with both end vertices in the same community i. Maximizing Q corresponds with maximizing the disparity between the number of edges that exist within clusters and the number of edges that link clusters together.

Table. 1. Best 20 Characteristic Word in Incident Reports (TF: Term Frequency)

	Detail	Reaction	Solution
1	Patient	Report	Confirmation
2	Confirmation	Patient	Time
3	Medicine	Check	Patient
4	Nurse	Confirmation	Order
5	Order	Attending Doctor	Medicine
6	Administration	Notice	Explain
7	Infusion	Doctor	Nursing
8	Pill	Call	Nurse
9	Room Visit	Night Shift	Thoroughness
10	Dose	Doctor	Nurse
11	Route	Order	Check
12	Find	Nurse	Use
13	Time	Explain	Medicine
14	Bed	Condition	Doctor
15	Morning	Apology	Route
16	Operation	Administration	Need
17	Give	Medicine	Time
18	Doctor	Minute	Internal use
19	Self	Measure	Attention
20	Report	Route	Infusion

Table. 2. Best 20 Characteristic Word in Incident Reports (TF-IDF)

	Detail	Reaction	Solution
1	Medicine	Report	Confirmation
2	Patient	Patient	Time
3	Confirmation	Confirmation	Patient
4	Administration	Attending Doctor	Order
5	Order	Notice	Medicine
6	Nurse	Night Shift	Explain
7	Dose	Doctor	Nursing
8	Infusion	Order	Thoroughness
9	Room Visit	Nursing	In
10	Route	Explain	Check
11	Find	Condition	Use
12	Bed	Apology	Medicine
13	Operation	Administration	Doctor
14	Time	Medicine	Route
15	Use	Measure	Need
16	Report	Route	Time
17	Doctor	Dose	Administration
18	Toilet	Examination	After
19	Self	Self	Attention
20	Indication	Infusion	Before

4. Results

4.1. Linguistic Analysis Results

The best twenty characteristic words that appear in the incident report such as "Detail," "Reaction," and "Solution" with *tf* (term frequency) are shown in Table 1. The best twenty characteristic words that appear in the incident report with *tf-idf* (Eq.(1)) are shown in Table 2. Under the category of "Detail," the words "Patient," "Confirmation," "Medicine," and "Nurse," rank high in Table 1. Moreover, the fact that the word "Nurse" ranks high shows that there are many accidents related to nurses. Under the category of "Detail" in Table 2, the word "Medicine" ranks higher than the one in Table 1. Therefore, accidents related to medicines are important for analyzing the reports. Also, the characteristic words related to the medical process such as "before" and "after" are high rank in Table 2. In addition, the characteristic words related to patients' managements such as "Room Visit," "Toilet," and "Bed" are high rank in Table 2.

Figure 4, figure 5 and figure 6 show the networks of characteristic words created using the incident reports. Each node represents a characteristic word, and an edge represents the intensity of the co-occurrence between the words. In viewing the network for "Detail" (see Figure 4), most of the reports are about medicines, infusion and fallings. The accidents of medicines are happened in the morning because the peak of giving patients internal medicines is morning in this hospital. In viewing the network for "Reaction" (see Figure 5), the structure of the network is more complex. This is because that the action after an accident by medical staffs and the patient progresses are various depending on the situations. In viewing the network for "Solution" (see Figure 6) it is clear that the network is created around the characteristic word "Confirmation," and one can see that descriptions of solutions are about confirming the medicines, patients, or timings by nurses.

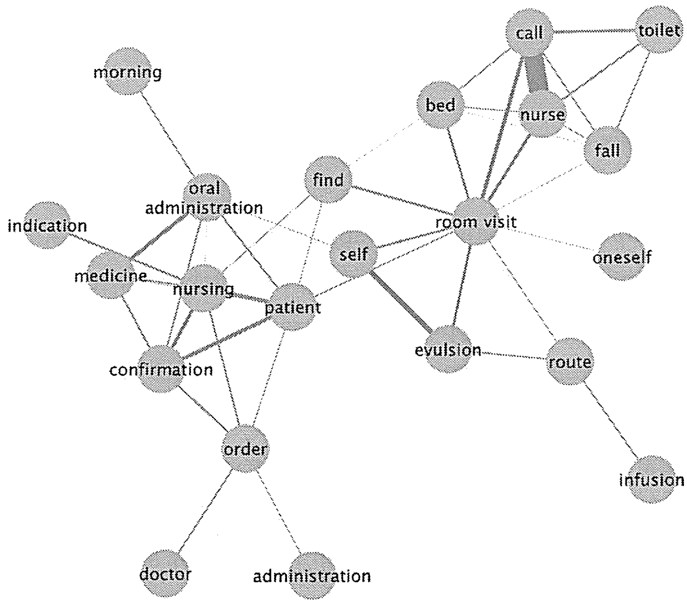


Figure 4. Co-occurrence Network of the Characteristic Words (Detail).

***Nodes mean the best 30 characteristic words based on tf-idf measures, and edges mean the existence of co-occurrences. Heavy lines between characteristic words mean strong connection between words.**

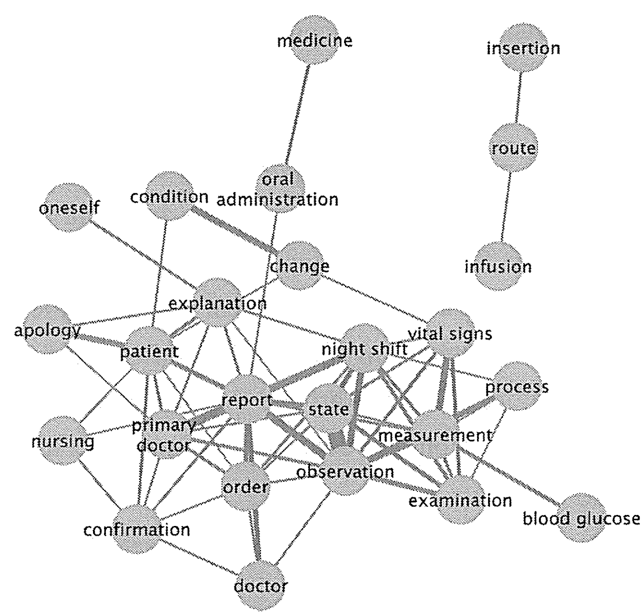


Figure 5. Co-occurrence Network of the Characteristic Words (Reaction)

****Nodes mean the best 30 characteristic words based on tf-idf measures, and edges mean the existence of co-occurrences. Heavy lines between characteristic words mean strong connection between words.**

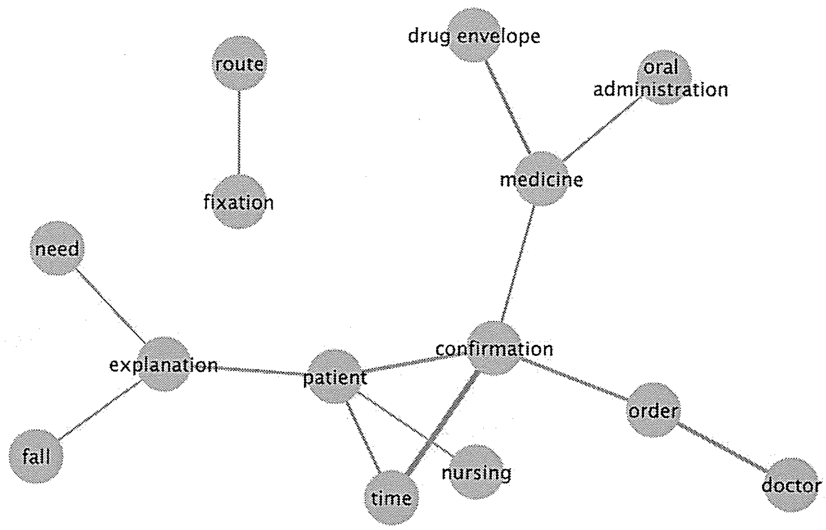


Figure 6. Co-occurrence Network of the Characteristic Words (Solution)

****Nodes mean the best 30 characteristic words based on tf-idf measures, and edges mean the existence of co-occurrences. Heavy lines between characteristic words mean strong connection between words.**

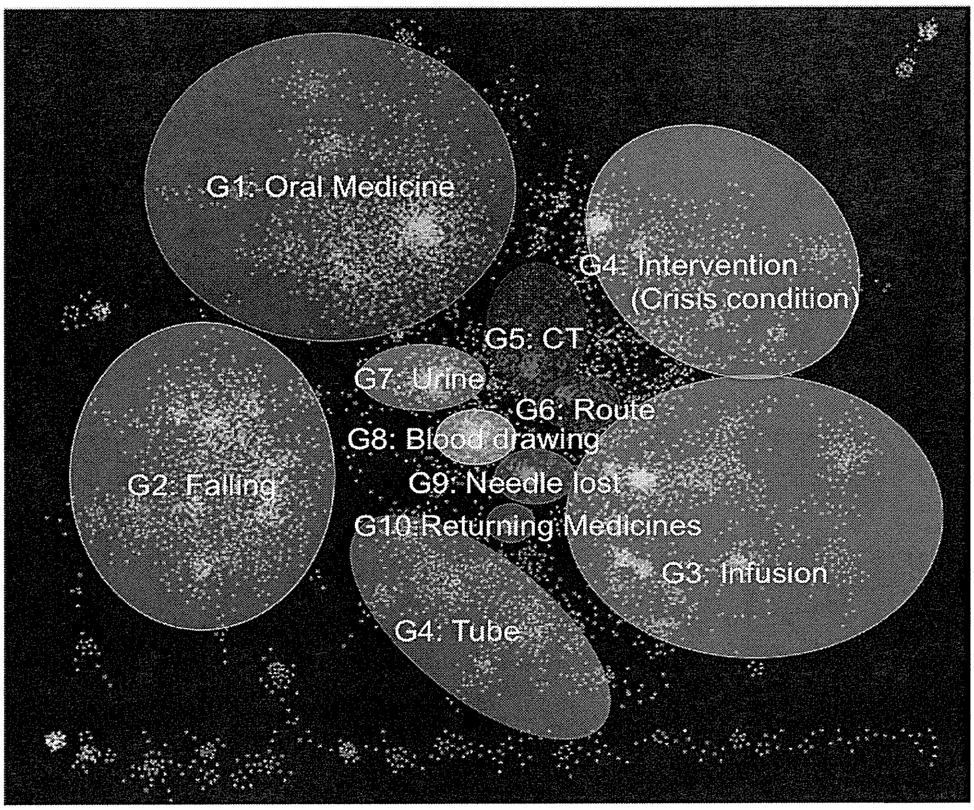


Figure 7. Clustering Results to the Network of incident reports (Detail)

***The name of the group means the rank of the number of nodes**

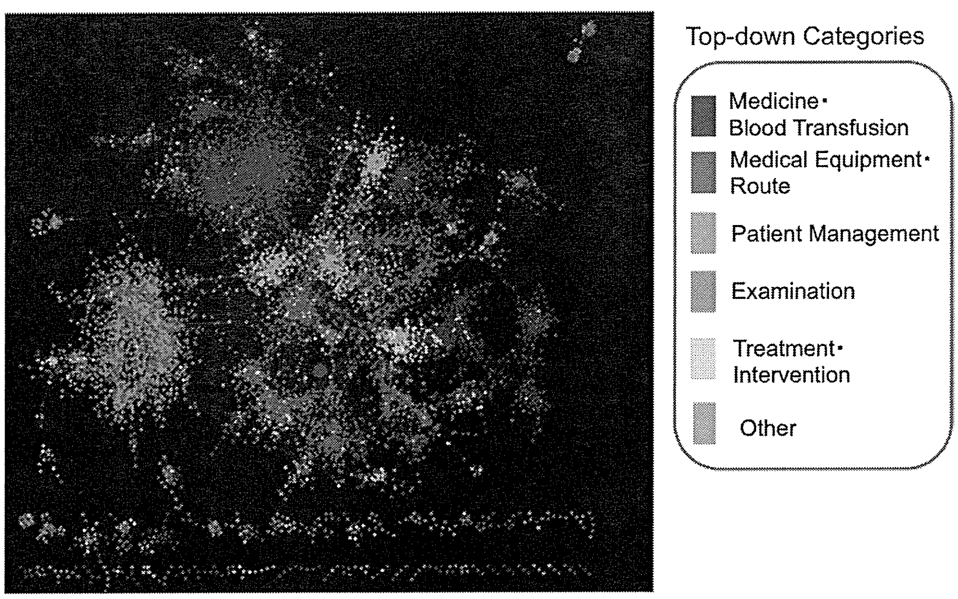


Figure 8. Comparison between Top-down Categories and Bottom-up Categories for Network of Documents (Detail)

*** The color of nodes mean the categories decided by the experts and the positions of nodes are plotted according to similarities between nodes.**

4.2. Categorizing Results based on Bottom-up Analysis

Figure 7 shows the best ten clusters and in "Details" categories. The reason of selecting the "Detail" section is that this category contains the most important information in order to divide the reports into some groups. The doctors in Osaka City University carried out labeling of each cluster based on the networks. The clusters from position 1 to 5 clearly appear the granted classes. In the cluster in position 1, the class of oral medicines is prioritized. In this group, administrations of medicines and preparations of medicines are included; in positions 2, the class of patient managements regarding patients' fallings is created clearly; however, in position from six to ten, the classes that clearly characterize clusters could not be seen.

5. Discussions

Figure 8 shows the comparisons between the top-down categories and bottom-up categories for network of documents in the section of details of situations. In this figure, the color of nodes mean the categories decided by the experts and the positions of nodes are plotted according to similarities between nodes.

The top-down category of "Medicine and Blood Transfusion (Blue)" has two main groups; one is about the oral medicine, the other is about the infusion. This means that top-down categories of "Medicine and Blood Transfusion (Blue)" should be divided into two new categories. In fact, some medical staff indicates that this category is too wide range as the categories of the incident reports in the interview.

The reports with "Patient Management (Orange)" category are mainly about the fallings, and most of them contain the second largest groups labeled as the Falling in figure 7. It is better for the medical staffs that the label of this category is "Falling" because some medical staffs labeled the accidents of fallings to "Others (Gray)." Actually, "G2: Falling" contains orange and gray nodes in figure 7.

The top-down categories of "Examination (Sky)" and "Medical Equipment and Route" are divided into some small groups. In the interview to the medical staff in Osaka City University Hospital, the idea of dividing the top-down category of "Examination (Sky)" and "Medical Equipment and Route (Red)" into the some small categories was appeared.

The top-down category of "Treatment and Intervention" can't be seen the large-sized groups. Most of the medical staffs don't use this top-down category because most of the accidents are divided into other suitable top-down categories. Therefore, this top-down category doesn't make any sense.

However, in position from six to ten, classes that clearly characterize clusters could not be seen. The possible future work is that the section for analyzing the bottom-up categories isn't only the detail section but also the combinations of the detail section and the reaction section.

6. Conclusion

In this paper, the characteristic words were extracted by analyzing incident reports, and the co-occurrence networks of the characteristic words were created. As a result, the language networks with the hub of the word "confirmation," thereby revealing that inadequate confirmations on the drug labels, instructions of a physician and patient were very significant causes of accidents. These results suggest the effectiveness of introducing the network analysis method. In addition, the class of patient managements regarding patients' fallings in top-down analysis is created clearly. On the other hand, some categorizes by top-down analysis don't reflect the category by the bottom-up analysis. These results suggest the effectiveness of introducing the network analysis method.

In the future work, we would like to focus on the medical repots for improving the notational rules for the names of drugs and dosages in incident reports. Also, we would like to analyze the differences of understanding of the incident reports between positions like doctors, nurses, pharmacists.

References

AM, Cohen, WR, Hersh, C, Dubay, K, Spackman (2005).Using co-occurrence network structure to extract synonymous gene and protein names from MEDLINE abstracts" BMC Bioinformatics 6:103.

CD, Manning and H, Schutze (2002). Foundations of statistical natural language processing, The MIT Press.

E, Rasmussen (1992). Algorithms, Information Retrieval: Data Structures and Algorithms. William B. Frakes and Ricardo Baeza-Yates.

G, Salton (1983). Introduction to Modern Information Retrieval. Mcgraw-Hill College.

H, Takada, K, Saito, T, Yamada, M, Kimura (2006). Analysis of Growing Co-occurrence Networks, SIG-KBS, Vol.73, pp.117-122.

HW, Heinrich (1931). Industrial accident prevention: A safety management approach. McGraw-Hill Customer Service.

Japan Council for Quality Health Care: JCQHC (2009), Project to Collect Medical Near-Miss/Adverse Event Information 2009 Annual Report, http://www.med-safe.jp/contents/english/index.html.

MeCab (2006). Yet Another Part-of-Speech and Morphological Analyzer, http://mecab.sourceforge.net/.

MEJ, Newman (2004). Fast algorithm for detecting community structure in networks. Physical Review E, 69, 066133.

P, Jaccard P (1902). Lois de distribution florale. Bulletin de la Socíeté Vaudoise des Sciences Naturelles 38:67-130.

P. Jaccard P. (1912), The distribution of the flora in the alpine zone. New Phytologist 11(2):37-50.

WHO (2009). Conceptual Framework for the International Classification for Patient Safety,http://www.who.int/patientsafety/implementation/taxonomy/icps_technical_report_en.pdf

Y, Kajikawa, J, Ohno, Y, Takeda, K, Matsushima, and H, Komiyama (2007). Creating an academic landscape of sustainability science: an analysis of the citation network. Sustain Sci2:221–231.

PRELIMINARY LINGISTIC ANALYSIS OF LARGE NUMBER OF MEDICAL INCIDNET REPOTS FOR PATIENT SAFETY

Katsuhide Fujita[1)], *Masanori Akiyama*[2)], *Keunsik Park*[3)], *Etsuko Yamaguchi (Nakagami)*[3)], *Hiroyuki Furukawa*[4)]

1) Institute of Engineering of Innovation, The University of Tokyo
2) Policy Alternatives Research Institute, The University of Tokyo
3) Osaka City University, 4) Yamaguchi University Hospital

ABSTRACT

The analysis of medical incident reports is indispensable for the patient safety. Most of the incident reports include some free composition formats, therefore, the analysis of free descriptions gives new perceptions. We aimed to accumulate, to interpret information again by structured incident information, and to clarify the point that should be improved for the cause of the accident and safe medical treatment improvements in the present study. We employ the natural language processing to the analysis of medical incident reports in this paper. The network analysis can find various relationships that are not only direct relationships but also indirect relationships. First, some important characteristic words were extracted in three categories of the accident's background, details, and solutions using TF-IDF measure. By using the TF-IDF, we can get some important characteristic words for analyzing the reports. In addition, we show the co occurrence networks using these extracted words.

1. INTRODUCTION

"In the shadow of every serious accident, there exist 29 times more minor accidents and 300 times more near misses." This principle was published in 1929 by Herbert William Heinrich, an assistant manager in the technology and research division of an American insurance company [1]. This principle, which hits home the nature of the occurrence of accidents, is taken up in various fields, such as the study of failure, safety engineering, cognitive psychology as well as the study of reliability, and the incident analysis of minor accidents associated with this is recognized as being important in preventing accidents.

Also, the use of information pertaining to medical accidents is important when implementing medical safety measures. The medical safety mechanism of WHO aims to prevent accidents by reusing incident reports through the introduction of IT technology. Harvard University is engaged in the standardization for the collection of medical accident reports and accident information in the risk management consortium. In England, the National Health Service conducts the medical accident/incident report collection project. Even in Japan, the Ministry of Health, Labour and Welfare began the project to Collect Medical Near-Miss/Adverse Event Information in 2001[2]. Through this project, the Ministry conducts analyses based on the collected incident reports.

On the other hand, regarding patient safety, guidelines for the future deployment of incident analysis are set out in WHO's International Classification of Patient Safety (ICPS) [3]. ICPS states the necessity of first investigating the adequacy of classes of incident case studies such as those mentioned above, and second, methods of expressing incidents that adequately reflect these classes, i.e., it states the necessity of ontological construction. In this research, in line with WHO guidelines, we conducted an analysis regarding the adequacy of classes in case studies collected in the Project to Collect Medical Near-Miss/Adverse Event Information and the tendencies of description that aim at ontological construction.

In the Medical Near-Miss/Adverse Event Information including the abstract, background, and solution for a single case are described using a free composition format. In this paper, we analyze the large number of medical incident repots (more than 15,000 reports) provided by Osaka City University using the natural language processing and the network analysis. By using natural language processing, an understanding of the tendencies of description as well as guidelines for future ontological construction can be acquired.

The remainder of this paper is organized as follows. First, we describe the dataset of the medical incident reports provided by Osaka City University. Next, we describe the methodology based on the Natural Language Processing and the Network analysis for analyzing the large number of medical incident reports. Then, we present the results of analysis of incident reports. Finally, we present our overall conclusions.

2. MEDICAL INCIDENT REPORTS BY OSAKA CITY UNIVERSITY