

However, there is an issue to achieve perfect traceability of blood products that is tradeoff between public safety and privacy data protection in this situation. In contrast to drug traceability, perfect traceability of blood products is including highly private information such as infectious information of donors and there is a possibility to like the information to a specific name. Collecting information on blood products has a possibility to be a threat for donor's privacy. Solution for this tradeoff is also required to implement traceability system and our target in this study.

1) *Point of Act System*: POAS captures complete data on each medical action including 6W1H information (When, Where What, Why, for What, to Whom and How) conducted in the hospital. The units of data recorded by the system are: Who—the implementer (the person who initiated the order, or the person who carried it out), to Whom—the patient, How—medical activities and changes in them, What—materials used (pharmaceuticals, medical materials and others), How much—amount of materials used and number of applications, for What—name of patient receiving medical services, When—the order was placed, implemented and discontinued and the activities that were implemented, and Where—place of implementation (department, hospital, ward, etc.). The collection of complete data including 6W1H information is an innovative source in understanding actual situations directly without estimation or bias, and enables the investigation of solutions to prevent error [7,8].

2) *Complete data*: POAS data is “Complete data” that capture every action by real time and quite accurately. This means the data captured by the system has full traceability of drugs and materials and can be used for analyses on healthcare management. Complete data provide us great opportunity to analyze situation of healthcare management, quality and safety without any sampling methodologies to estimate original value. That makes reliability of analyses higher. In addition, complete data is especially useful for patient safety researches, because complete survey is necessary to estimate medical error and accident rate.

3) *Process Management*: Structure of POAS data capturing is based on process management of each medical action. Process management structure requires every medical workers capture data at their point of action. Without capturing data on completion of activities, medical workers can't do next activities on the medication process. It enables POAS to acquire every result of medical action and assure capturing complete data.

4) *Settings*: Our experimental project was enforced in Iwate Red Cross Blood Center and Morioka Red Cross Hospital as Table1. We created the system for auto identification and data capturing from blood collection in the blood center to administration in the hospital with POAS and RFID. The system put time stamp with the data to ensure rightness of information and consistency of process order in capturing data.

5) *Single item management from production to consumption with SGTIN*: Serialized number was put on

RFID to distinct each blood product with single item level. Serialization of blood products is essential factor to distinguish one blood from others uniquely. If a number was used for more than two objectives, it makes difficult to confirm an object uniquely.

6) *Certification system for safe blood transfusion and electrical data capturing with RFID*: This system was aimed to confirm 5 rights of transfusion at each point of transfusion. 5 Rights in blood transfusion is right patient, right blood, right unit, right root and right time. Right blood in this setting includes five additional components with checking product ID. At the point of checking, this system certified types of product including Blood Red Cell, plasma and blood plate, blood type appropriateness, completion of cross matching, result of cross matching and adverse event information of products from same donor. In concrete, system that is possible to verify information of infected blood products founded just 2 minutes before in other hospitals. Table I shows comparison of verification component with other blood transfusion systems. Existing systems had focused on Blood type certification and some systems had tried to integrate transfusion system with blood test laboratory system to check the results of cross matching at laboratory.

Experimental project with the system described above was implemented in Morioka Red Cross Hospital and Iwate Red Cross Blood center. The experimental period is from 30/07/2007 to 30/11/2007 for 4 month. Object departments in Iwate Red Cross Blood Center are department of Testing, Preparation and Delivery. Objective departments and wards in Morioka Red Cross Hospital are wards of digestive tract internal medicine, General Medicine, Surgery and Testing Department. The number of blood products used in these 3 wards is 75% of total usage in all hospital. Though the object wards are three, it is enough to investigate feasibility of the system in these three wards. We operated 377 blood products with RFID during the term.

TABLE I

COMPARISON OF AUTO IDENTIFICATION WITH OTHER SYSTEMS

	Existing Administration System	POAS System
Blood Type Certification	Possible	Possible
Completion of Cross Matching	Partially Possible	Possible
Checking results of Cross Matching	Impossible (Need additional procedures)	Possible
Checking adverse event information by database located outside hospitals	Out of Focus	Possible

III. RESULTS

A. Process management

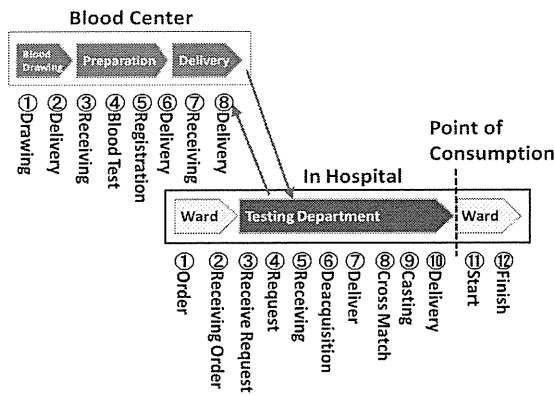


Fig. 1. Process of blood transfusion

We analyzed process of transfusion in Morioka Red Cross Hospital and Iwate Red Cross Blood Center to identification and track appropriately. Figure 1 shows result of process analysis of a transfusion in time series. Transfusion Process could be divided into two major parts, blood center and hospital, based on the place. In blood center, staffs in blood center collect blood from donors and deliver it to department of preparation. Department of preparation receive the blood and test blood for screening whether the blood is appropriate for blood products or not. Department of preparation prepare the blood passed screening for products and form products to deliver to hospitals. At this time, blood is ready for use for blood transfusion and wait for requests from hospitals.

On the other side, in hospital physicians order transfusion for patients and nurses receive the order and request blood products to department of testing. This order was made junction with blood delivered from blood center and department of testing in hospital operate cross matching. If the result of cross matching shows appropriateness for transfusion, the blood delivers to the point of transfusion. And in a ward or operating room, nurses or physicians administer the blood to patient. In this sequential process, there are movements of places and many actors engage to this process to operate transfusion. Figure 1 shows normal process of blood transfusion in the hospital. This is not only case to be treated by transfusion system. We analyzed transfusion process in the hospital to find every type of process to cover all case of transfusion. Each process shaped into use case with UML and there are 14 types of use case for transfusion process. Table II shows 14 types of use case.

These use cases can be classified with 4 major categories; Ordinary, Cancellation, Warning and ex post

TABLE II
COMPARISON OF AUTO IDENTIFICATION WITH OTHER SYSTEMS

Classification	Case Number	Usecase Name
Ordinary	1	Transfusion Operation (without stocks)
	2	Transfusion Operation (with stocks)
	3	Allocated Products to Stock
	4	Cross unmatched
Cancell	5	Order cancelled
	6	Order cancelled after allocating blood
	7	Scrap blood products
	8	Request cancelled
Warning	9	Prohibition of usage in entering test results)
	10	Prohibition of usage in starting transfusion)
	11	Taking wrong products in deliver
	12	Wrong patient in transfusion
	13	Wrong blood products in transfusion
ex post facto	14	Usage after office hour

facto. In ordinary process, transfusion operation with blood stocks and without blood stocks are regarded as different use case, because interactions and movements on information and products are different in each use case. Similarly, in cancellation, the activities and information to be interchanged are different based on the timing of cancellation. Ex post facto means information was entered after injection because they are used after office hour and testing department was closed.

The important thing to achieved 5 rights transfusion with IT is feasibility of the system and information. In this system all certification was operated with just one data capturing by the point of action. In verifying information, system refers original data captured or entered at the point of actions.

In this system, we can check completion of cross matching without fail by process management technique. In designing system, we analyzed process of medical activity and described as nonreversible process that is a series of medical activities.

B. Use Case analysis

We described all patterns of blood transfusion process by process analysis methods. In normal transfusion process, flow of process from physician's order to administration goes thought without hitch. However there are other patters including emergency cancellation and rejection of blood products. By describing all patterns of use case, it is possible to construct the system can handle all occasions without any exception. This feature is especially important for traceability. Use case 1 (Transfusion operation without stocks) is the most common use case, because the blood center is located on next to the hospital and they don't need to have a lot of stocks. Use case 2 is usually the most common case in hospitals. In use case 1, three actors including physician, technician, nurse and staff in blood center operate the process. At first, physician makes a decision on transfusion and order transfusion. Nurse receives the order and delivers the order to testing

department. Technician starts preparation for transfusion by request to blood center, because they don't have a stock in the hospital. Technician receives blood products from blood center, operates cross matching and delivers it to ward or operation room. Nurse operates the transfusion.

Our system was created based on these analyses on process of transfusion with use cases. Every type of transfusion except use case 14 was target to ensure traceability of blood products. Figure 4 shows overview of our system. In hospital, transfusion process was managed by Transfusion management server and Hospital information/CPOE server. In blood center, transfusion process was managed by public server and donor server. At he each point of process, actors read RFID to capture data on 6W1H and auto identification with PDA and computer. Transfusion server connected to public server in hospital through internet (VPN). This connection makes possible to manage whole process from production to bedside.

The ID on RFID was rewritten after preparation. Information to link donor ID to product ID was securely managed inside blood center and blocked physically and nobody could refer to this server from outside. The process of rewriting was also under process management and the process can't be processing without rewriting.

By connecting hospital system to public server in blood center through internet, it is possible to certify availability of blood products by original product database including adverse event information in real time. When nurse identified blood product at the point of care, PDA checked adverse event information in blood center database trough middle ware as well as patient information and product information. All transaction for identification to ensure 5 rights and right process was completed within 2 seconds. If some infected blood were found at other hospital and the information was putted in public server, right after the time PDA warn usage of the blood products prepared from same source. By this system, nurses can check safety of transfusion from the various points of view by just one reading RFID with PDA within 2 seconds. It is effective to improve patient safety and operation of nursing works.

C. Evaluation

We evaluated the system based on data captured by this system. We proved that the system with the RFID tag and SGTIN was able to manage the pharmaceutical drugs at the single item level in real time, and improve patient safety. For all 377 blood transfusions captured by this system, data was perfectly collected and there is no inconsistency on the data. For patient safety, it is very important that the response of the information processing is quick to operate the real system. We accessed the data center in Nagoya City, central Japan, away at 900km from the hospital in Morioka City in the Tohoku region, north-east Japan, through the Internet line. As a result, the processing time of the system was within 0.4 seconds, and

thought to be a response enough by practical use also at each stage. Moreover, both access times were the responses within one second, and it was thought enough though the access from the wholesale enterprise in Morioka City was an access to the data center in Nagoya City that used the connection of the Internet of a very narrow band up to 402 kbps by Personal Handy-phone System. To evaluate data captured by the system, we drew the data as traceability graph. An example of traceability graph is Figure 2. Horizontal axis shows time flow and vertical axis shows flow of blood products from production to consumption. There are two lines before delivery. Left line shows blood product flow in blood center. Before delivery, blood products don't timed with any special patients. Right line shows flow of transfusion order by doctor. Followed by the order, a blood product was matched with a transfusion order at the point of delivery. From the point, Product ID is associated with Patient ID. The advantage of describing traceability graph is showing result of data capturing visually and easily. If data capturing system worked correctly, data capturing point is going right with process progressions. If the line is going left, there are some problems on data capturing such as ex post data entry, delay of the system and time lag among systems.

The result of describing 377 traceability shows the system ensures traceability for all blood products from production to consumption. Certification at each process was successfully done. During the experimental period, there is no accident and medical mistakes on blood transfusion. At the level of the business load, in the entire work, each person in charge is skilled in a new system and we improved the operating effectiveness of the person in charge more than existing business. In terms of checking prescription and mixing injection drugs, we have double check system now. In near future, we can make up single check system with RFID tag system for backup.

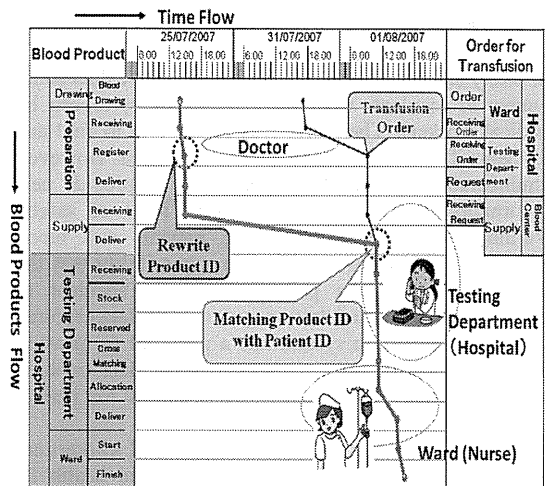


Fig. 2. Traceability Graph of Blood Products

Moreover, we proved the expectation of the effect of the medicine of the abandonment amount reduction, and contribution also improved management and the CO₂ exhaust amount reduction of the medical institution.

As a result of the operation switch verification at the failure, we confirmed the operation switch was able to be done promptly, and there was no big influence on the hospital work. We expect that applicability of the RFID tag is able to have the good effect of the batch reading in the business of confirming a large amount of medicine.

IV. DISCUSSION

We constructed system with internet and RFID to manage whole processes from production to consumption of blood products to expand the capability of certification system and ensure traceability. Many previously published literatures have been tried to construct certification system at the point of care for blood transfusion or at blood center to make right documentation of blood products for blood safety and management [1-6]. Compare to these systems, this system has several advantages that other systems don't have. This system ensures 5 rights of transfusion and right process and information with real time original information. By checking original data base through middle ware at the point of care, the correctness of information for certification is highly secured. This technique makes us possible to check the original database to certify patient information and blood products with electrical medical records and computerized order entry system directly and find adverse events information on blood products through internet.

We tried to evaluate improvement on blood transfusion safety and traceability with this system. The number of medical accidents and incidents on blood transfusion from April to June 2007 (before implementation of the system) is zero and the number of them during experimental period is also zero. This information didn't provide us evidence of improvement on safety based on number of accidents. These data about number of accidents were based on voluntary reports by medical workers. Therefore, it is impossible to find any accidents if they don't report medical accidents. However, administration systems have possibility to provide new opportunity to evaluate and measure level of safety. Warning logs by reading wrong patients and blood products RFID means that there is a possibility that the administration for the patient would be accident or incident without administration system. The data captured by the administration system has a potential to measure the level of safety and comparison of the data between before and after implementing interventions for patient safety is our next target for researches.

Costs including work burdens of medical workers are sometimes the highest obstacle to introduce health IT system [9-10]. It is useful to investigate feasibility of the system by evaluating change of time to finish each

activity [11-18]. We investigated time to finish each activity by collecting data observationally and computed average length of time from around 10 observationally data of each activity. We compared length of time to finish each activity between using this RFID based administration system and using paper based communication and documentation. We chose six activities for comparison and six activities are blood receiving, decussation testing and stock taking in the blood center and delivering to wards, certification before administration and recording administration in the hospital.

For all 6 activities, the time to finish each work with RFID is shorter than with paper based system. Works of nurses and technicians would be also effective as well as safer by introducing RFID based traceability system. Especially time for administration of transfusion is decreasing. This system has a possibility to improve productivity of transfusion process as well as transfusion safety.

Another way for identification and data capturing used widely is barcode technology. RFID is superior to traditional barcode technology in numerous ways [15,16]. RFID does not require line-of-sight, allows simultaneous read of multiple tags, is able to store more information on the chip, can include sensors for condition monitoring such as time and temperature, and enables automatic identification and data capture [15]. In addition to these operative advantages, RFID enable rewriting information and it is significantly important to construct a solution for privacy data protection and future extensions.

We investigated and focused improvement of transfusion safety with auto identification and data capturing system. In addition to these advantages, it also has possibility to provide significant advantages on hospital management and regional health system management. The ways to storage blood products were strictly regulated, because quality of blood products is easy to change with affects from outside. Red blood cell products must be stored inside refrigerators having a function to record temperature and blood platelet must be inside storage with vibration system. They require strict methods to store and blood products delivered once were regarded as consumption and sometimes wasted. Blood products are scarce and valuable resource from human blood, and waste of blood products might cause safety and management problems in hospitals and regional healthcare system. Traceability data can contribute to solve these issues by visualizing data of distributed blood products. That would enhance effective use of blood products by connecting hospitals to blood center and among hospitals.

V. CONCLUSION

In this study, we focused on identification and data capturing for patient safety. Capturing data and alibi management of materials including blood products leads

to effective use of resources as well as improve patient safety as mentioned above. We can certificate each medication and capture those data at the same time, contribute to patient safety and improve health care delivery. To be a trusted system, the systems have to use right information and consider securities of people. Trusted IT system can contribute to patient safety, effective use of blood products, reducing waste that might be essential factors for trusted health care system.

ACKNOWLEDGMENT

This work was supported by Health Labour Sciences Research Grant of Japan.

REFERENCES

- [1] J.C. Chan, R.W. Chu, and B.W. Young, "Use of an electronic barcode system for patient identification during blood transfusion: 3-year experience in a regional hospital," *Hong Kong Medical Journal*, vol.10, pp166-171, 2004.
- [2] A. Davies, J. Staves, J. Kay, A. Casbard, and M.F. Murphy, "End-to-end electronic control of the hospital transfusion process to increase the safety of blood transfusion: strengths and weaknesses," *Transfusion*, vol.46, pp352-364, 2006.
- [3] M.F. Murphy, "Application of bar code technology at the bedside: the Oxford experience." *Transfusion*, vol.47, pp120-124, 2007.
- [4] A. Ohsaka, K. Abe, T. Ohsawa, N. Miyake, S. Sugita, and I. Tojima, "A computer-assisted transfusion management system and changing transfusion practices contribute to appropriate management of blood components," *Transfusion*, vol.48: pp1730, 2008.
- [5] R.W. Askeland, S. McGrane, J.S. Levitt, S.K. Dane, D.L. Greene, J.A. VandeBerg, K.Walker, A. Porcella, L.A. Herwaldt, and L.T. Carmen, "Kemp JD. Improving transfusion safety: implementation of a comprehensive computerized bar code-based tracking system for detecting and preventing errors," *Transfusion*, vol.48, pp1308, 2008.
- [6] R. Davis, B. Geiger, A. Guitierrez, J. Heaser, and D. Veeramani, "Tracking blood products in blood centers using radio frequency identification: a comprehensive assessment," *Vox Sanguinis* vol.97, pp50-60, 2009.
- [7] M. Akiyama, "Risk Management and Measuring Productivity with POAS-Point of Act System-A Medical Information System as ERP (Enterprise Resource Planning) for Hospital Management," *Method Inf Med*, vol.46, pp686-693, 2007.
- [8] M. Akiyama and T. Kondo, "Risk Management and Measuring Productivity with POAS--point of act system," *Stud Health Technol Inform*. Vol.129, pp208-12, 2007.
- [9] S.G. Sandler, A. Langeberg, and L. Dohnalek L, "Bar code technology improves positive patient identification and transfusion safety," *Adv Transfusion Safety*, vol.120, pp19-24, 2005.
- [10] S. Allen, "System Targets Blood-Type Mix-Ups", February 24, 2005. *Boston Globe Health/Science*; available at http://www.boston.com/news/globe/health_science/articles/2005/02/24/system_targets_blood_type_mix_ups/
- [11] R.W. Askeland, S. McGrane, J.S. Levitt, S.K. Dane, D.L. Greene, J.A. VandeBerg, K. Walker, A. Porcella, L.A. Herwaldt, L.T. Carmen, and J.D. Kemp, "Improving transfusion safety: implementation of a comprehensive computerized bar code-based tracking system for detecting and preventing errors," *Transfusion*, vol.48, pp.1308-1317, 2008.
- [12] D. Watson, J. Murdock, C. Doree, M. Murphy, M. Roberts, A. Blest, and S. Brunskill, Blood transfusion administration one- or two- person checks, "which is the safest method?" *Transfusion*. vol.48, pp783-789, 2008.
- [13] E. J. Thomas , D. M. Studdert, H.R.Burstin, E.J. Orav, T. Zeena, and E.J. Williams, "Incidence and types of adverse events and negligent care in Utah and Colorado," *Med Care*, vol. 38, pp261-271, 2000.
- [14] R. Sharma, S. Kumar, and S.K. Agnihotri, "Sources of preventable errors related to transfusion," *Transfus Med Prod*, vol.81, pp37-41, 2001.
- [15] R. Davis, B. Geiger, A. Guitierrez, J. Heaser, and D. Veeramani, "Tracking blood products in blood centers using radio frequency identification: a comprehensive assessment," *Vox Sanguinis*, vol. 97, pp50-60, 2009.
- [16] S.G. Sandler, A. Langeberg, and L. Dohnalek, "Bar code technology improves positive patient identification and transfusion safety," *Adv Transfusion Safety* vol.120, pp19-24, 2005.
- [17] M. Akiyama, A. Koshio, and N. Kaihotsu, "Analysis of data captured by barcode medication administration system using a PDA; aiming at reducing medication errors at point of care in Japanese Red Cross Kochi Hospital," *Stud Health Technol Inform*. Vol.160, pp774-8, 2010.
- [18] C. Huckvale, J. Carl, M. Akiyama, S. Jaafar, T. Khoja, A. B. Khalid, A. Sheikh, and A. Majeed, "Information technology for patient safety," *Qual Saf Health Care (BMJ)* vol.19, pp i25-i33, 2010.

Detection of Precarious Situations in Medical Care with Mining Track record of Dosing

SATORU YAMAMOTO

Policy Alternatives Research Institutes, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan
yamamoto@biz-model.t.u-tokyo.ac.jp

YINGZI JIN

The Institute of Engineering Innovation, The University of Tokyo, 2-11-16, Yayoi, Bunkyo-ku, Tokyo, 113-8656, Japan
Yzjin2006@gmail.com

YUTAKA MATSUO

The Institute of Engineering Innovation, The University of Tokyo, 2-11-16, Yayoi, Bunkyo-ku, Tokyo, 113-8656, Japan
matsuo@biz-model.t.u-tokyo.ac.jp

ICHIRO SAKATA

Policy Alternatives Research Institutes, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan,
isakata@jpr-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, poas@mit.edu

Abstract

We propose a new approach to detect the precarious situation in medical care analyzing tracking record. Attention is being drawn to the use of incident reports as a means of increasing patient safety. Research teams are being formed across the world by WHO and in Japan by the Health, Labour and Welfare Ministry. In this instance, what is being emphasized as a major direction for future incident analysis is the assimilation of the existing top-down type class grants and bottom-down ontological construction. In this research, targeting incident case studies collected in Japan, we evaluated the degree of similarities between incident documents obtained bottom-up and the links between existing classes granted top-down. In doing so, we made it possible to evaluate overall similarities regarding incident documents through the method of network analysis. In addition, it became clear that the use of the Cos coefficient or the Jaccard coefficient is appropriate in creating networks.

As a result of this analysis, existing classes correspond comparatively well with the characteristics of reports regarding the abstract and solution; on the other hand, regarding the background, it demonstrated that existing classes are inadequate in representing the characteristics of documents and that there is a need to improve classes. By the way, we can upgrade patient safety and quality of health care service.

Key word: healthcare service, incident report, eHealth, WHO

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, ergonomics, 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. Therefore, patient safety with eHealth becomes growing field of research recently (Gaudhi and Lee, 2010, Mcloughlin et al., 2006, Huckvale et al., 2010, Kaushal and Bates, 2002).

In order to eradicate medical malpractice, medical institutions break down barriers between departments, collect and analyze incidents, and work out countermeasures. With this background, the Health, Labour and Welfare Ministry in Japan started the Project to Collect Medical Near-Miss/Adverse Event Information from 2001. This project collects, analyzes, and publishes incident reports. The guidelines for analysis are to calculate the total of each class, including related medical departments, occurrence factors, and time periods, and to root out the causes of accidents.

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). 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 data provided in the Project to Collect Medical Near-Miss/Adverse Event Information, the abstract, background, and solution for a single case study is described using a free composition format. 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 research used the techniques of natural language processing and 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. Moreover, by networking the reports obtained from this, discoveries of overall links that could not be found from comparing only two reports are expected.

2. MEDICAL INCIDENT REPORTS

Here, we will explain approaches and issues relating to medical incidents and characteristics of the data used in this research.

2.1 Overview of Incident Reports Sought by ICPS

ICPS's general description sets out past activities and future guidelines relating to incident reports (WHO). Until now, the main work of ICPS has been the granting and maintenance of classes to accidents by specialists. By granting this kind of top-down "agreed upon class," it becomes possible to convey a summary of incidents and accidents to even those who are not medical specialists. On the other hand, top-down type of classes created from present conditions are not detailed enough to provide satisfactory explanations of the characteristics of individual incident case studies. In addition, as class is granted in advance, opportunities to find valid unknown classes for patient safety are lost. Consequently, ICPS has stated that it will introduce ontological thought as part of future guidelines. Ontology in medicine refers to conducting from the bottom up and based on actual data the construction of methods necessary for describing individual case studies without misinterpretation as well as the discovery of classes of case studies that use these methods. ICPS indicates that the granting

of top-down type categories by specialists as well as the granting of information that uses bottom-up type of ontology are necessary.

2.2 Collection of Incident Information in Japan

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. The results of aggregate calculations and analysis are published on the website of this organization.

2.3. Data Sets

From among incident data provided by the Project to Collect Medical Near-Miss/Adverse Event Information, in this research, we used data relating to medical agents from 2005 to 2010 that was published on the Internet. In order to conduct a detailed analysis, from the case studies provided, we used only 1,067 documents that included the information of the abstract, background, and solution. Each case study is in a free composition format, with the abstract, background, and solution being approximately 300 characters long, respectively. In addition, the two classes of medicine and accident are granted to each case study. With regard to the class of treatment, there are six classes of general drug, preparation of drugs, drowsy of drugs, contraindicated drug, chemo treatment, and other drug; with regard to the class of operation, there are the nine classes of name of drug, amount of drug, regimen, amount and regimen, flow rate, drug sensitivity, diapedesis, forget to dose, and object person. With regard to the class of treatment, as all the classes of operation do not exist, there are 32 cross classes that cross calculate the class of treatment and the negligent class of operation.

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. NATURAL LANGUAGE PROCESSING

In this research, we extract characteristic words using natural language processing as a first step in extracting important information that characterizes each document with the aim of ontological construction. The links between each document are determined from similarities between characteristic words obtained here. As natural language processing contains a lot of noise, there is a need to conduct preprocessing in order to obtain characteristic words that can be used in determining links. Preprocessing mainly comprises three stages of “breaking down into words in reports,”

“connecting words that have been broken down too much,” and “filtering the obtained words.” Details of these are set out below.

3.1 Breaking Down into Words

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 and Schütze, 2002). In this research we used MeCab, one of the most common engines for conducting morphological analysis.

3.2 Connecting Phrases

There is the possibility that words obtained using MeCab are too finely classified to conduct the analysis of links. Therefore, in this research we connected words using the two methods set out below and used them as new words.

First, we connected words using information on the parts of speech. The above-mentioned MeCab not only breaks down words but also grants major classes and minor classes relating to parts of speech. In cases where the minor class of parts of speech of certain words was a suffix and the word before it was a noun, these two words were treated as one word.

Next, we connected words based on the number of word occurrences (Matsuo and Ishizuka, 2002). Let us envisage a situation in which two words—hereafter called A and B—appeared consecutively. If we designate the number of word occurrences in instances where each word is considered separately as $n(A)$, $n(B)$, then the number of word occurrences in which they appear consecutively is expressed as $n(A \cap B)$. In cases where $n(A \cap B) / \min(n(A), n(B))$ exceeded the threshold value (0.8 in this research) then we treated those two words as one word.

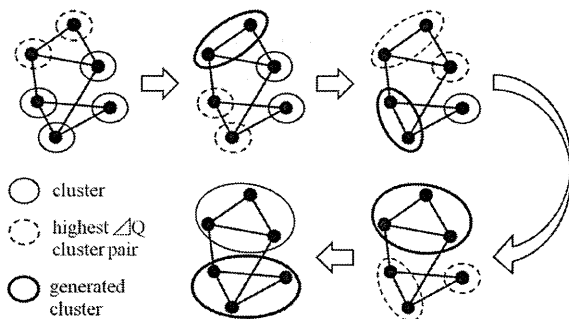


Fig1. Model of Newman clustering

3.3 Filtering Words

Words obtained through the above two processing methods still contain a lot of noise, which can be expected to exert a bad influence on the calculation of links in documents. Thus, it is necessary to select words to be used in calculating links. The following sets out details on filtering.

First, we conducted filtering using the class of parts of speech. As stated above, major and minor classes are granted to words. Nouns were the targets of research on this occasion as major classes of parts of speech. Nominalized verbs, general nouns, and proper nouns were also targeted as minor classes of parts of speech. 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.

Next, we conducted filtering based on the frequency of occurrence. In this research, we calculated a value called tfidf from the frequency of occurrence and conducted filtering based on this values. tfidf 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 (Saltin,). tfidf is calculated using the following formulas.

$$\text{tfidf} = \text{tf} \cdot \text{idf} \quad (1)$$

$$\text{tf}_i = \frac{n_i}{\sum_k n_k} \quad (2)$$

$$\text{idf}_i = \log \frac{|D|}{|\{d: d \ni t_i\}|} \quad (3)$$

Here, n_i is the frequency of occurrence of word i , $|D|$ is the total number of documents, and $|\{d: d \ni t_i\}|$ is the number of documents in which word i occurs.

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. In this research, we set the maximum value of tf to 50 and eliminated the noise from words with an abnormally high frequency of occurrence. On the other hand, words that make a small number of appearances also have an extremely small value for idf and, as a result, the tfidf has a tendency to increase. Therefore, this time we treated all tf under 10 as 0.

4. NETWORK ANALYSIS

Network analysis is an extremely effective method of looking at the links between documents (Kajikawa et al., 2007, Uchida et al., 2009, Shibata et al., 2010, Shibata et al., 2011 as examples). 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.

4.1 The Creation of 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 co-occurrence $|A \cap B|$ for two documents. Here, $|A \cap B|$ is the number of characteristic words that exist in 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. Consequently, a number of co-occurrence indices that improve on these points have been proposed, with representative indices including the Jaccard coefficient, the Simpson coefficient, and the Cos coefficient (Rasmussen, 1992). Each formula is shown in (4), (5), and (6) and is generally Simpson coefficient > Cos coefficient > Jaccard coefficient.

$$\text{Jaccard coefficient: } \frac{|A \cap B|}{|A \cup B|} \quad (4)$$

$$\text{Cos coefficient: } \frac{|A \cap B|}{\sqrt{|A||B|}} \quad (5)$$

$$\text{Simpson coefficient: } \frac{|A \cap B|}{\min(|A|, |B|)} \quad (6)$$

A link is established between the two documents in the event that these indices exceed the threshold value. The network created changes acutely depending on which of the above indices are selected and how the threshold value is established. As the aim of this research 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.

4.3 Network Clustering

In this research, other than given classes, we conducted clustering using the Newman method in order to carry out labeling of each document based on networks. The Newman method is a widely used method for clustering networks (Newman, 2004). 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 (7), clustering was conducted by maximizing modularity Q , an index for evaluating the modularity of defined networks.

$$Q = \sum_i (e_{ii} - a_i^2) \quad (7)$$

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 a_i represents the sum of row i of line e . 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.

5. RESULTS

5.1 Extraction of Characteristic Words

We conducted the extraction of characteristic words using the methods outlined in section 3. Table 1 and 2 show the number of occurrences by category for the 10 words in which tfidf is at the top. We see from these tables that all characteristic words extracted here occur selectively in each class. In addition, when focusing on the combination of class and occurring word, we see that words that aptly reflect the characteristics of the class are being extracted, such as “injected part” occurring largely in the diapedesis class and “allergia” occurring largely in the drug sensitivity class. On the other hand, as the index in this research uses tfidf as an index of the degree of characteristics, general words do not appear at the top, even if the frequency of occurrence is high. For example, although “patient” had the highest number of occurrences, as this word appears in most case studies, tfidf as an index of class capability of documents takes on a low value and does not appear at the top.

Table.1 characteristic word and class of operation

characteristic word	diapedesis	flow rate	object person	forget to dose	composition
insulin	0	0	4	0	0
furosemide	0	1	0	0	0
injected part	14	0	9	0	1
allergia	0	0	5	0	0
setup	0	51	0	0	0
flow rate	1	38	2	0	0
coinjection	0	3	7	0	3
anticancer	3	3	7	0	0
periphery	1	2	10	0	2
set	0	5	1	13	1

characteristic word	regimen	name of drug	amount of drug	drug sensitivity	amount and regimen
insulin	0	22	16	0	0
furosemide	4	13	17	0	0
injected part	0	0	1	0	0
allergia	0	1	0	19	0
setup	0	0	4	0	0
flow rate	2	0	9	0	0
coinjection	4	22	9	1	0
anticancer	0	2	5	0	7
periphery	4	4	2	3	0
set	1	9	16	0	6

Table2. Characteristic word and class of treatment

characteristic word	chemo treatment	contraindicated drug	dowry of drugs	preparation of drugs
insulin	0	0	3	2
furosemide	0	2	10	1
injected part	14	1	0	0
allergia	0	23	0	0
setup	14	0	0	0
flow rate	8	0	0	0
coinjection	5	7	0	0
anticancer	26	0	2	0
periphery	5	5	0	0
set	0	1	29	3

5.2 Network Analysis

In this research, we aim at uncovering links between classes granted top-down and clusters discovered bottom-up. Thus, we created a network so that documents belonging to identical classes relating to the identical treatment and operation are grouped together the most. To achieve this, we used the index of Class Closeness (CC) defined in formula (8).

$$CC = \sum_i (D_{ii} - \overline{D}_i) \quad (8)$$

Here, D_{ij} of line D calculates the distance from all of the nodes within category i to all of the nodes within category j and takes the averages of these. Also, \overline{D}_i is the average of the i row of D. The classes here refer to 32 cross classes that cross calculate the class of treatment and the class of

operation. CC taking a high value indicates that the nodes of identical classes come close to cases seen across the entire network. If the threshold value of the co-occurrence index is high, there is a tendency for CC to become high as only links with strong links remain. On the other hand, if the threshold value is made too high, a large proportion of the links are lost, the maximum number of connections (LC) of nodes within the network decreases, and the analysis of overall links becomes impossible. Thus, in this research, in order that the product of the maximum number of connections of CC and nodes (CCLC) are at the maximum, we conducted the selection of co-occurrence indices and the determination of threshold values. Table 3 shows the maximum value of CCLC for each co-occurrence index and the values of indices in these instances. By doing so, we discovered that a network that reflects given classes could be obtained by using the Cos coefficient for networking the content and the Jaccard coefficient for networking the background and solutions. Although the Cos coefficient and Jaccard coefficient demonstrate almost identical CCLC in the abstract, background, and solution, it is markedly low regarding the Simpson coefficient. This is because documents that include large numbers of characteristic words are connected with many documents and are close to documents that belong to other classes in the network.

Table 3. Class closeness and co-occurrence index

abstract				
Index	Value	CC	LC	CCLC
jaccard	0.290	3.162	510	1612.4
cos	0.459	2.908	594	1727.9
simpson	1.000	0.430	750	322.5

back ground				
Index	Value	CC	LC	CCLC
jaccard	0.334	1.441	580	835.8
cos	0.580	3.348	241	806.8
simpson	0.860	0.192	970	186.5

solution				
Index	Value	CC	LC	CCLC
jaccard	0.338	1.441	550	1444.7
cos	0.544	2.955	428	1264.9
simpson	0.700	0.279	1023	285.3

Table 4. cluster and characteristic class

cluster index	characteristic class	share of the class	ratio of the class in the cluster	number of the node
1	dowry of drugs	84.80%	60.50%	147
2	preparation of drugs	58.80%	51.30%	111
3	flow rate	55.30%	70.30%	37
4	diapedesis	55%	30%	27
5	-	-	-	25

6. DISCUSSION

The characteristic words displayed in Table 1 and 2 were all selected based on actual data. Networks obtained from similarities with these characteristic words remove the effects of similarities shared in common with all documents and are formed from the overall combination of an independent degree of similarity between two documents. This kind of network is first realized by extracting characteristic words bottom-up. In cases where keywords that should be checked top-down are decided, there are instances when, after having conducted class, there is no guarantee that that keyword is not valid and a network of documents linked only by the characteristic similarities such as those described above cannot be obtained.

Figure 2 shows the abstract network from among the networks created based on the characteristic words obtained bottom-up. In addition, Table 4 shows the abstract network resulting from having conducted clustering using the Newman method. Table 4 shows the top five clusters of the number of nodes and the main classes within these. The clusters from position 1 to 4 clearly reflect the granted classes. In the cluster in position 2, the class of treatment is prioritized and collects together documents classified as drowsy of drugs and preparation of drugs; in positions 3 and 4, the class of operation is prioritized and clusters reflecting flow rate and diapedesis are created; however, in position 5, classes that clearly characterize clusters could not be seen.

Moreover, regarding the solution network, links between clusters and classes such as those described above were extracted. On the other hand, clusters that clearly reflected the given class could not be seen among the clusters of the background network. The fact that networks created through background do not reflect overall given classes is in accord with the results of analysis in which the value of CC corresponding to the co-occurrence index used in creating networks in Table 3 is around 3 in abstract and solution, while in background it is less than 1.5.

So, is the class of background not related in some way to networks created bottom-up? In order to investigate this, we conducted an investigation into how many identical classes exist within the k step from certain nodes. The results of this investigation are shown in Fig. 3. We can see that in no matter which network, the ratio of nodes with identical class decreases to the extent of the number of steps increasing. Here, attention should be drawn to the point that the proportion of identical classes existing within a single step is almost the same in each of the categories of abstract, background, and solution. This indicates that the same classification also takes place in case studies with an extremely high degree of similarity in background. On the other hand, when looking at the overall network, given classes relating to the background are scattered compared with abstract and solution. This concept is outlined in Fig. 4.

With regard to the abstract and solution, links determined from ontology reflect comparatively given categories; however, regarding the background, when looked from the perspective of ontology, there is the possibility that it is completely different from the classes of abstract and solution, even those matters in proximity to step numbers, i.e., it is difficult to predict accidents that occur from information on background. Therefore, it shows that the content of descriptions of background is inadequate from the perspective of preventing reoccurrences. On the other hand, regardless of differences between abstract, background, and solution, the reporter is likely to have taken care to use descriptions that express accurately the maximum limits and the reality of the situation. Class methods that give appropriate suggestions to on-site reporters are required. We can say that there is a strong desire for an ontological construction relating to descriptions of background.

Considering the similarities between only two documents without using network analysis is equivalent to looking at the similarities of the nodes within step one in Fig. 3. In this case, in order to

obtain almost identical values between abstract, background, and solution, differences relating to the power of expression of descriptions in the three divisions are not seen. The discovery that descriptions of background are inadequate was made possible for the first time by considering the overall similarities through the networking of case studies.

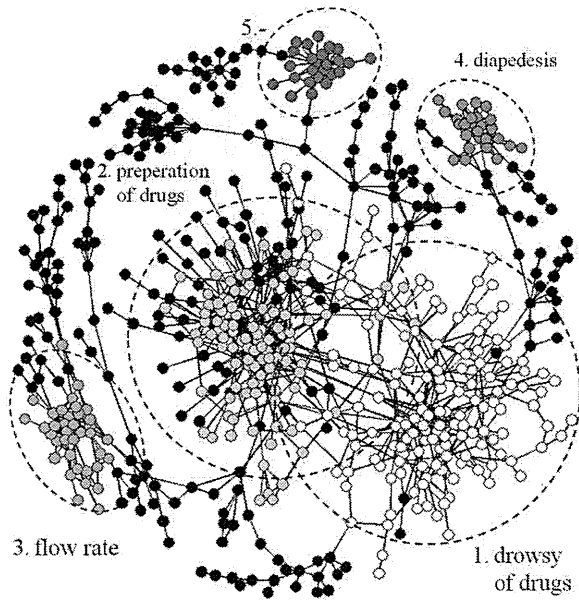


Figure 2. Network of an abstract in Near-miss/ Averse Event

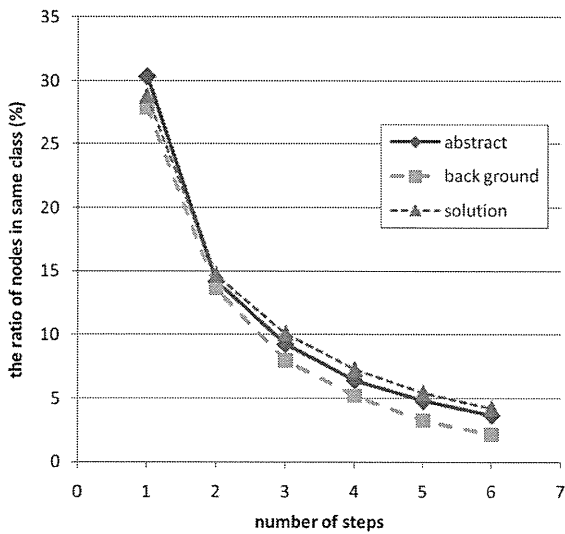


Figure 3. the ratio of nodes in same class

7. CONCLUSION

In this research, we evaluated the degree of similarities between incident documents obtained bottom-up and the links between existing classes granted top-down. In doing so, we made it possible to evaluate overall similarities regarding incident documents by using the method of network analysis. Moreover, it became clear that the use of the Cos coefficient or the Jaccard coefficient is appropriate for determining similarities in creating networks.

With regard to the background, the results of the analysis demonstrated that, compared with abstract and solution, existing classes are inadequate for representing the characteristics of documents and that there is a need to improve classes (figure 4).

By using the methods employed in this research, suggestions otherwise unobtainable through conventional methods can be made regarding the investigation of how classes should be.

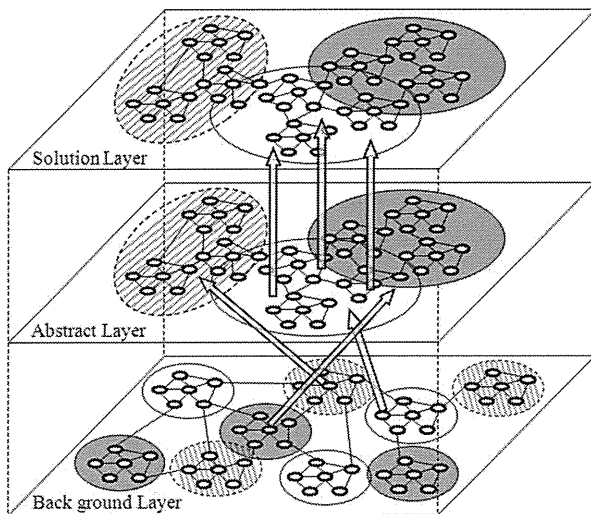


Figure 4. Structure of network in each layer

8. REFERENCES

- Gandhi, T.K. and T. H. Lee (2010). Patient safety beyond the hospital. *The New England Journal of Medicine*, vol.363, pp.1000-1003.
- Heinrich, H W. (1931). *Industrial accident prevention: A safety management approach*. McGraw-Hill Customer Service.
- Huckvale, C., J. Car, M. Akiyama, S. Jaafar, T. Khoja, B. Farlow, S. Yaron, G. Locke and S. Whittaker (2010). Information technology for patient safety. *Quality & Safety in Health Care*, vol.19(2), pp. i25-i33.
- Kajikawa, Y., J. Ohno, Y. Takeda, K. Matsushima, and H. Komiyama (2007). Creating an academic landscape of sustainability science: an analysis of the citation network. *Sustainability Science*, vol.2, pp.221–231.
- Kaushal, R. and DW. (2002). Bates, Information technology and medication safety: what is the benefit ? *Quality & Safety in Health Care*, vol.11, pp.261-265.
- Majeed, A., J. Car and A. Sheikh (2008). Accuracy and completeness of electronic patient records in primary care, *Fam Pract*, vol.25, pp.213-214.
- Manning, C. D., and H. Schutze (2002). *Foundations of statistical natural language processing*. The MIT Press, London.
- Matsuo, Y. and M. Ishizuka (2002). Keyword extraction from a document using word co-occurrence statistical information. *JSAI*, vol.17(3), pp.213-227.
- McLoughlin, V., J. Millar, S. Mattke, M. Franca, P.M. Jonsson, D. Somekh and D. Bates (2006). Selecting indicators for patient safety at the health system level in OECD countries. *International Journal for quality in Healthcare*, Sept 2006, pp.14-20.
- Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E*, 69, 066133.
- Rasmussen, E. (1992). *Clustering algorithms, information retrieval: data structures and algorithms*. William B. Frakes and Ricardo Baeza-Yates.
- Salton, G. (1983). *Introduction to Modern Information Retrieval*. McGraw-Hill College.
- Shibata, N., Y. Kajikawa and I. Sakata (2011). Measuring relatedness between communities in citation network. *Journal of the American Society for Information Science and Technology*, in press
- Shibata, N., Y. Kajikawa and I. Sakata (2010). Extracting the commercialization gap between science and technology. *Technological Forecasting and Social Change*, vol.77(7), pp. 1147-1155.

Uchida, M., N. Shibata, Y. Kajikawa, Y. Takeda, S. Shirayama and K. Matsushima (2009).
Identifying the large-scale structure of the blogosphere. *Advances in Complex Systems*, vol.12,
pp. 207-219.

<http://mecab.sourceforge.net/>

<http://www.med-safe.jp/contents/english/index.html>

<http://www.who.int/patientsafety/implementation/taxonomy/en/>

http://www.who.int/patientsafety/implementation/taxonomy/icps_statement_of_purpose.pdf

Information Science Linkage of Service Innovation

ICHIRO SAKATA

Policy Alternatives Research Institutes, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan,
isakata@ipr-ctr.t.u-tokyo.ac.jp

JUNICHIRO MORI

Innovation Policy Research Center, The University of Tokyo, 2-11-16, Yayoi, Bunkyo-ku, Tokyo, 113-8656, Japan
jmori@ipr-ctr.t.u-tokyo.ac.jp

NAOKI SHIBATA

Innovation Policy Research Center, The University of Tokyo, 2-11-16, Yayoi, Bunkyo-ku, Tokyo, 113-8656, Japan
shibata@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

YUYA KAJIKAWA

Innovation Policy Research Center, The University of Tokyo, 2-11-16, Yayoi, Bunkyo-ku, Tokyo, 113-8656, Japan
kajikawa@ipr-ctr.t.u-tokyo.ac.jp

Abstract

It is with this recognition that policies for strengthening international competitiveness regarding service innovation are being adopted by many countries. While planning and implementing these policies, what is required in essence is an objective analysis regarding the current status of knowledge related to this field and the linkage between science and innovation. However, the knowledge infrastructure of this kind is inadequate. The aim of this paper is to develop the way to identify the meta structure of knowledge and measure "information science linkage of service innovation". We use bibliometrics, co-citation network analysis and co-occurrence analysis to objectively identify them.

In the field of service innovation, our results show that there are mainly two groups of elements related to service innovation: applications of service innovation and basic theories for service innovation. In the field of service science, we also identified major knowledge groups such as machine learning, and information retrieval. Then we calculated the co-occurrence of characteristic terms in journal paper abstracts belonging to the SSME sub-clusters and author keywords for papers belonging to the sub-sub-clusters related to information science. We clarified which information sciences are used heavily in service innovation. We also determined which areas of innovation make heavy use of information science and which do not. In the field of medical care, the high value we obtained demonstrates that digital health and EHR research is being vigorously conducted worldwide. By contrast, in other fields related to medical treatment, such as mental healthcare and patient satisfaction, we found a big room to promote the spread of information science.

Overall, we have demonstrated the possibility of using bibliometrics to objectively identify the meta structure of knowledge and measure the semantic relationships between science and technology.

Keywords: SSME, Information Science, Science Linkage, Technology Roadmap

Introduction

It is widely recognized that the concept of service innovation is significant for innovation strategy and economic growth. It is with this recognition that policies for strengthening international competitiveness regarding service innovation are being adopted by many countries. While planning and implementing these policies, what is required in essence is an objective analysis regarding the current status of knowledge related to this field and the technological linkage between science and innovation. However, knowledge infrastructure of this kind is inadequate. The sense of concept SSME is so broad that there is not the common understanding about what is service innovation even among experts. Another reason is the lack of suitable indicator or the way of measurement. Even though, the indicators of science linkage which uses citation data of patents or corporate studies have been developed (Motohashi and Yun, 2007, Bonaccorsi and Thoma, 2007 as examples), they provide us limited information of semantic relationships of technologies. Therefore, using bibliometric methods, we shed light on the structure of knowledge pertaining to service innovation and the linkage or semantic similarity between service science and innovation.

The first aim of this paper is to create a knowledge landscape of service innovation and service science from a number of academic publications. In this paper, of the many definitions of service innovation, we focus on the concept “Service Science, Management, and Engineering (SSME)” proposed by IBM. This paper uses same data with our previous study of SSME (Shibata et al., 2009). As an indicator of service science, we use the data in the fields of information science, library science, and computer science, which are the most important basic knowledge for service innovation. We collect data of academic papers, create citation networks regarding papers as nodes and citations as links, categorize papers into sub-clusters or sub-sub-clusters, extract topics of each sub-cluster, and finally discuss the results with experts. The second aim is to calculate the degree of distance in the semantic connections between service innovation sub-clusters on the one hand and information science, library science, and computer science sub-clusters on the other. The degree of distance in these semantic connections may be employed as an indicator of the depth of the relationship between innovation and individual scientific technologies. This is referred to as “Information Science Linkage of Service Innovation.” In addition, the relationship between