Shibata,N., Y. Kajikawa, J. Mori, I. Sakata, and K. Fujisue, "Creating an academic landscape of service innovation," proc. of the 19th International Conference for Management of Technology (IAMOT2010), in Cairo, Egypt (8-11, March, 2010)

Shibata, N., Y. Kajikawa, Y Takeda, I. Sakata and K. Matsushima, "Detecting emerging research fronts in regenerative medicine by the citation network analysis of science publications," *Technological Forecasting and Social Change* (2010), in press

Bibliometric Analysis of Service Innovation Research: Identifying Knowledge Domain and Global Newwork of Knowledge

Ichiro Sakata[1], Hajime Sasaki[1], Masanori Akiyama[1], Yuriko Sawatani[2], Naoki Shibata[3]
[1] The University of Tokyo, Todai Policy Alternatives Research Institute, Tokyo, Japan
[2] JST, RISTEX, Tokyo, Japan
[3] The University of Tokyo, Innovation Policy Research Center, Tokyo, Japan

Abstract—**It is widely recognized that the concept of service innovation is significant for innovation strategy and economic growth. However, since the term "service innovation" represents a broad sense, there does not exist common understanding about what is service innovation even among experts. We developed a methodology to determine the structure and geographical distribution of knowledge, as well as to reveal the structure of research collaboration in such an interdisciplinary area as service innovation by performing journal information analysis, network analysis and visualization. Our results show that there are mainly two groups of elements relating to service innovation. Knowledge in these areas has been growing rapidly in recent years. In particular, the fields of ecosystem and IT and Web are exhibiting a high growth. We also demonstrated that the global network of knowledge is formed around the powerful hub of the US. The research competency of Asian countries lags behind that of the US and EU. With respect to research collaboration, we indentify a big room left for enhancing international collaborations. Our methodology could be useful in forming policies to promote service innovation. Finally, we proposed creation of an international collaboration fund.**

I. INTRODUCTION

The concept of service innovation or service science, management and engineering (SSME) proposed by IBM is widely recognized as a key driver for the economic growth. Then service science is emerging area of research [1, 2] Maglio and Spohrer [3] defines service science as the study of service systems, which are dynamic value co-creation configurations of resources. Spohrer et al.[4] argues that service science can be thought of as a mashup or integration of many areas of study known as service management, service marketing, service operations, service engineering, service computing, service human resources management, service economics, management of service innovation and others. Wu [5] discusses that the concept of SSME is an emerging interdisciplinary approach that combines fundamental management, and engineering theories. Such concept plays a significant role in policy making in many countries. In Japan, the government established a roadmap named Technology Roadmap of Service Engineering, which describes the goal of service innovation. However, the sense of concept SSME is so broad that there is not the common and deep understanding about what is service innovation even among experts [3, 5]. Although the roadmap mentioned above describes forty nine technology elements and the relationship between technology and industry, these descriptions lack concreteness because it is so conceptual. Therefore, prior to developing roadmap, it is necessary to make the academic landscape of SSME in order to understand what have been researched relating to this topic. Then, the first aim of this paper is to identify the way to create the academic landscape. Service science, service innovation or SSME has the interdisciplinary nature of approach [5, 6]. Tracking the evolution of interdisciplinary research domain, such as SSME, is a significant but difficult task by its nature. Previous study argued that interdisciplinary research should not be conceptualized with discipline [7]. Existing categories like journal categories may not matter because interdisciplinary researches vary beyond the boundaries of journals. Some indicators measuring interdisciplinary such as diversity of classifications and topological measures are proposed and evaluated in the previous papers [8,9]. Experts are not able to track the entire trends in such research areas as each research specializes and is segmented.

In such a situation, for policy makers, creating an academic landscape of interdisciplinary research and effective investment on those technologies has become a significant task in order to develop their competitive competence and also to realize the economic growth. In this paper, we develop a computational tool to support them to create an academic landscape among a pile of academic publication. There are two types of computer-based methodology, which can complement the expert-based approach: text mining and citation mining. As an example of the former, Kostoff et al. analyzed multi-word phrase frequencies and phrase proximities, and extracted the taxonomic structure of energy research [10,11]. In previous works, citation-based approaches, latter ones, were used to describe the network of energy-related journals using journal citation data or journal classification data [12]. In the citation-based approach, it is assumed that citing and cited papers have similar research topics. In this paper, we adopt the latter one.

Citation-based approach is useful to make an overview of research domains globally. Klavans and Boyack illustrated how to map science overall using journal citation interactions [13]. Rinia et al. pointed out the importance to consider the process how bibliometric measures are created [14]. By clustering the citation network, we can divide academic papers into groups of papers. Previous research investigated citation networks of academic publications relating to another interdisciplinary research, sustainability science, and extracted the major topics relating to this topic [15]. The first aim of this paper is to create an academic landscape in SSME research domain by using citation network analysis.

Although it is possible to grasp the structure of the intellectual world from the academic landscape, it is also meaningful to know the overall picture of geographical distribution of research and partnerships in research from the perspective of policy making. By understanding regional distribution and partnerships, it is possible to discuss the relationship between the number and nature of policies adopted in a particular region and the study of SSME, and it is also easy to develop a plan for global partnership in the field of research. Hence, the second aim of this paper is to draw a research network diagram that includes information on geographical distribution of knowledge and inter-regional collaboration. To create a research network diagram, we will use author information such as organisations to which authors are affiliated, nationalities of such organisations and co-authors from the same database used for the creation of academic landscape. There are several studies that use co-authorship as a quantitative indicator [16,17]. Co-authorship is used as an indicator of international collaboration [18,19,20]. Katz and Martin point out four key advantages of using co-authorship as an indicator of collaboration including its verifiability, statistical significance, data availability, and ease of measurement [21]. On the other hand, bibliometric analysis of multiple-author papers is not accurate as it can only be used to measure collaborative activities where the collaborating participants have entered their names on joint papers. We are aware of a bias where each research paper published separately despite the collaboration cannot be correctly identified. Nevertheless, this unique analytical method and data provides useful and clear empirical evidence, and when used with appropriate caution reveals new insights for international science policy. Our results can offer an intellectual basis for constructing a policy and strategy.

As stated above, the concept of SSME is yet to be clearly defined. This concept is mainly used in the US so far. It is possible that by using the term SSME as a query keyword, study fields growing in countries other than the United Stated will be underestimated. In evaluating academic landscape and research network diagram, it will be necessary to consider this possibility.

II. METHODOLOGY

First of all, the methodology for creating academic landscape is shown. Analyzing schema is depicted in Fig. 1. The step (1) is to collect the data of the knowledge domain. We collect citation data from the Science Citation Index Expanded (SCI-EXPANDED), the Social Sciences Citation Index (SSCI), and the Arts & Humanities Citation Index (A&HCI) compiled by the Institute for Scientific Information (ISI), which maintains citation databases covering thousands of academic journals and offers bibliographic database services, because these are three of the best sources for citation data. The problem, how we should define a research domain, is difficult to solve. One solution is to use a keyword that seems to represent the research domain. When we collect papers retrieved by the keyword, we can make the corpus for the research domain.

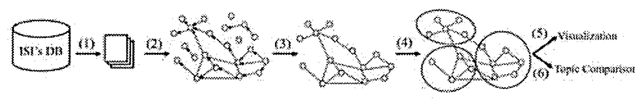

Figure 1. Methodology proposed in this paper.

The step (2) is to make citation networks for each year. We construct citation networks by regarding papers as nodes and intercitations as links. The network created for each year facilitates a chronological analysis of citation networks. According to a previous study, intercitation, which is also sometimes known as direct-citation, is the best way to detect emerging trends [22]. In network analysis, only the data of the largest component on the graph was used, because our study focuses on the relationships among documents, and we therefore want to eliminate from our study those not linked with any others in step (3).

After extracting the largest connected component, in step (4), the network is divided into clusters using the topological clustering method [23], which does not need the number of clusters by users. Newman's algorithm discovers tightly knit clusters with a high density of links within cluster. After the clustering, we visualize the citation networks and named the major clusters of emerging topics as in steps (5) and (6), respectively. In step (5), in order to visualize citation maps, we apply a large graph layout (LGL), an algorithm developed by Adai et al. [24], capable of dynamically visualizing large networks comprised of hundreds of thousands of nodes and millions of links. We visualize the citation network by expressing intra-cluster links in the same color, in order that the clusters are intuitively understood. In step (6), experts in the research domain assign a name to each cluster manually after they had seen titles and abstracts of the papers in each cluster.

Second, we create research network diagram by referring to the same database used for the creation of academic landscape, and for the extraction of data related to organisational affiliation of authors, geographical location of such organisations and co-authorships. Two types of data structure are developed: the data of research competency and of co-authorship. The data of research competency is obtained from the number of papers in each country or organization. The data of co-authorship is led by calculating all combinations of co-authors based on information about the author's organization. For example, if one paper is written by four different authors, and each author belongs to different organizations, the paper is considered to include six co-authorship relations. In addition, a co-authorship is defined as an international co-authorship if the authors belong to organizations in different countries. Authors in co-authored papers are not weighed by the order listed. Then, the data is visualized as a "research network diagram" with the author's organization as a node and co-authorship relation as a link between the nodes. In the diagram, organizations are grouped into the country they belong. In addition, combinations of organizations that have more co-authorship relations are identified. The hub of international co-authorships is also obtained.

III. RESULTS

In step (1) for creating academic landscape, we searched the papers using the terms "service* and (science* or management* or engineering*)" as the query. As a result, we obtained the data of 54,928 papers published until the end of 2008. The number of annual publications was shown in Fig. 2.

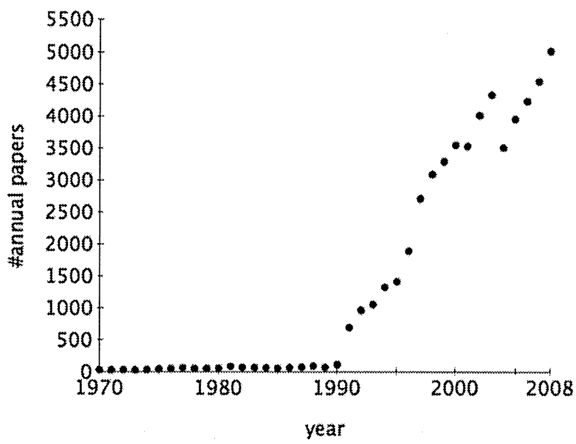

Figure 2. Number of annual papers relating to SSME.

After constructing largest connected component, as step (4), we divided papers into clusters with topological clustering method. With this clustering, citation networks as of 2008 were divided into specific clusters in step (4) and visualized as Fig. 3 in step (5). Focusing on the visualization in 2008, there were eight major clusters emerged in shown table1. Each contains more than 400 papers. The clusters #1, #2, #3, #4, #5, #6, #7 and #8 contained respectively 1,818, 1,681, 1,314, 914, 906, 866, 632, and 459 papers. Their combined publication dates averaged were 2003.0, 2002.7, 2000.8, 2004.1, 2002.7, 2002.2, 2001.8, and 2003.4. In the final step, step (6), our experts named each cluster, using semantic information such as the titles and abstracts of highly-cited documents in each cluster, shown in Table. 1. The clusters #1, #2, #3, #4, #5, #6, #7 and #8 related to management, medical care, mental health care, ecosystem, QOS, public service, public medical care, and IT & Web, respectively.

TABLE1. MAJOR8 CLUSTERS

Id	#papers	Average Publication Year	Name
S1	1818	2003.0	Management
S2	1681	2002.7	medical care
S3	1314	2000.8	mental health care
S4	914	2004.1	Ecosystem
S5	906	2002.7	QOS
S6	866	2002.2	public service
S7	632	2001.8	public medical care
S8	459	2003.4	IT & Web

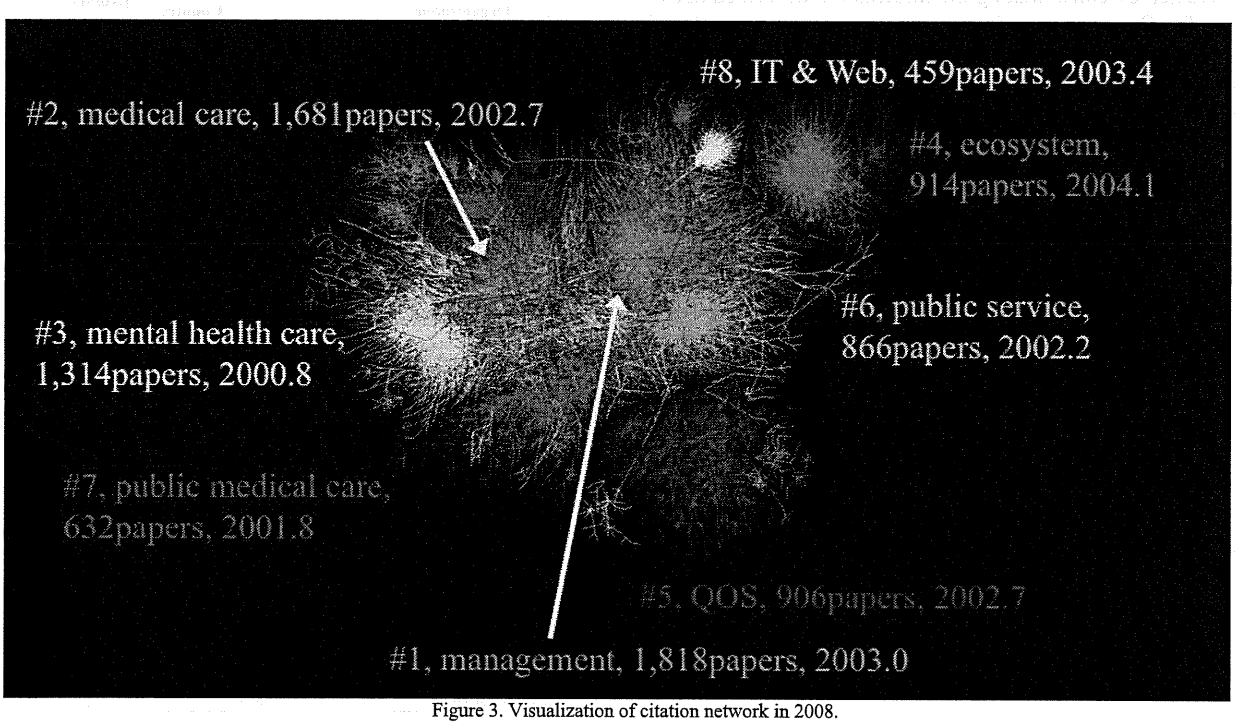

Figure 3. Visualization of citation network in 2008.

Next, a research network diagram is created. It was determined that the number of organisations to which authors of SSME related research papers are affiliated is 20,549, and the number of links among organisations based on the co-authorship is 72,484. Thus, a collaborative research network of 20,549 nodes and 72,484 links is formed in the field of SSME. The top five countries in the research competency are the US, England, Germany, France and Australia (Table 2). As a single country, it is notable that the US leads others by a significant margin. The research competency of EU as a total is slightly higher than that of US. Asian countries (China, Korea, India and Taiwan) are ranked between 10 and 18, although they show a rapid increase in their competencies. By analyzing the data of organisations (Table 3), it is clear that organisations with strong competency are often found in the US. Harvard University ranks first in the number of papers. Table 4 shows combinations of organisations that exhibit a high number of co-authorships. There are more co-authorship relations between organizations with high research competency. In addition, the number of collaborations between universities and their affiliated hospitals located nearby is notable. Geographically, many co-authorship relations are found in organisations within the same country, while international co-authorship is rare. In general, it has been noted that collaborative researches are often conducted by research organisations located within the same geographical region [19]. Nevertheless, collaboration in the SSME field is characterized by the dominance of domestic relationship compared to studies in the renewable energy field [25], which is growing rapidly. A research network diagram was created by consolidating the information described above (Fig. 4). Organizations in the same country are placed together and shown as a node. The size of each node shows the number of papers written by authors from the country. Each link between two nodes of different countries indicates that there is a co-authorships between countries. The Breadth of lines connecting the countries is proportional to the number of co-authorship. In a geopolitical sense, it is clear that the US is a powerful hub of the network. This is completely different from the network structure of renewable energy in which there is a well-balanced structure between North America, Europe and Asia [25]. In particular, there are thick lines between the US and countries such as Canada (1,319 links), England (1,319 links), Australia (506 links), Germany (475 links), China (414 links), Netherlands (325 links) and France (331 links). Among the relationships that do not involve the US as a hub, the thickness of lines between England and countries such as Germany (294 links), Australia (291 links) and Netherlands (281 links) is notable, illustrating that England is another major hub behind the US.

TABLE2. THE TOP 30 COUNTRIES IN THE RESEARCH COMPETENCY

Country	Number_of_Papers
USA	7649
ENGLAND*	2572
GERMANY*	1466
FRANCE*	1116
AUSTRALIA	965
CANADA	923
ITALY*	788
SPAIN*	625
JAPAN	598
PEOPLES R CHINA	469
NETHERLANDS*	453
INDIA	425
SOUTH KOREA	339
SWITZERLAND	339
BRAZIL	319
SCOTLAND*	304
SWEDEN*	295
TAIWAN	287
FINLAND*	232
GREECE*	200
SOUTH AFRICA	192
NEW ZEALAND	180
NORWAY	180
AUSTRIA*	179
ISRAEL	177
MEXICO	177
BELGIUM*	169
RUSSIA	151
DENMARK*	150
IRELAND*	145
* Total of EU in the top30	8694

TABLE3.THE TOP 30 ORGANIZATIONS IN THE RESEARCH COMPETENCY

Organization	Country	Number of Papers
HARVARD UNIV	USA	544
UNIV CALIF LOS ANGELES	USA	487
UNIV TEXAS	USA	453
UNIV MANCHESTER	ENGLAND	449
UNIV TORONTO	CANADA	448
UNIV MICHIGAN	USA	393
UNIV MARYLAND	USA	392
UNIV N CAROLINA	USA	390
UNIV ILLINOIS	USA	374
UNIV MINNESOTA	USA	371
JOHNS HOPKINS UNIV	USA	344
UNIV PENN	USA	344
UNIV WISCONSIN	USA	319
YALE UNIV	USA	315
UNIV PITTSBURGH	USA	309
COLUMBIA UNIV	USA	304
UNIV COLORADO	USA	289
UNIV CALIF SANFRANCISCO	USA	287
UNIV CALIF BERKELEY	USA	277
UNIV SYDNEY	AUSTRALIA	275
STANFORD UNIV	USA	274
UNIV MELBOURNE	AUSTRALIA	271
DUKE UNIV	USA	262
UNIV SO CALIF	USA	252
UNIV NEW S WALES	AUSTRALIA	243
MONASH UNIV	AUSTRALIA	242
OHIO STATE UNIV	USA	242
UCL	ENGLAND	230
UNIV QUEENSLAND	AUSTRALIA	228
INDIANA UNIV	USA	223

TABLE4. THE TOP 30 PAIRS OF CO-AUTHORED ORGANIZATIONS

Organization1	Country1	Number of Co-authored	Organization2	Country2
BRIGHAM & WOMENS HOSP	USA	59	HARVARD UNIV	USA
UNIV CALIF LOS ANGELES	USA	56	RAND CORP	USA
HARVARD UNIV	USA	47	MASSACHUSETTS GEN HOSP	USA
MONASH UNIV	AUSTRALIA	34	UNIV MELBOURNE	AUSTRALIA
UNIV NEW S WALES	AUSTRALIA	33	UNIV SYDNEY	AUSTRALIA
GRP HLTH COOPERAT PUGET SOUND	USA	33	UNIV WASHINGTON	USA
UNIV CALIF SAN FRANCISCO	USA	32	UNIV CALIF LOS ANGELES	USA
UNIV TORONTO	CANADA	29	ST MICHAELS HOSP	CANADA
UNIV SO CALIF	USA	27	UNIV CALIF LOS ANGELES	USA
UNIV N CAROLINA	USA	27	DUKE UNIV	USA
UNIV TORONTO	CANADA	26	INST CLIN EVALUAT SCI	CANADA
BOSTON UNIV	USA	25	HARVARD UNIV	USA
UNIV WASHINGTON	USA	24	UNIV CALIF LOS ANGELES	USA
MCGILL UNIV	CANADA	24	UNIV MONTREAL	CANADA
USDA	USA	23	USDA ARS	USA
YALE NEW HAVEN MED CTR	USA	23	YALE UNIV	USA
HOSP SICK CHILDREN	CANADA	22	UNIV TORONTO	CANADA
HARVARD UNIV	USA	21	CHILDRENS HOSP	USA
VA PUGET SOUND HLTH CARE SYST	USA	21	UNIV WASHINGTON	USA
MCMASTER UNIV	CANADA	20	UNIV TORONTO	CANADA
UNIV CALIF LOS ANGELES	USA	20	HARVARD UNIV	USA
CASE WESTERN RESERVE UNIV	USA	19	UNIV HOSP CLEVELAND	USA
UNIV CONNECTICUT	USA	19	YALE UNIV	USA
JOHNS HOPKINS UNIV	USA	19	UNIV MARYLAND	USA
YALE UNIV	USA	18	VA CONNECTICUT HEALTHCARE SYST	USA
ROYAL PRINCE ALFRED HOSP	AUSTRALIA	18	UNIV SYDNEY	AUSTRALIA
BETH ISRAEL DEACONESS MED CTR	USA	18	HARVARD UNIV	USA
NEW YORK STATE PSYCHIAT INST & HOSP	USA	18	COLUMBIA UNIV	USA
UNIV LIVERPOOL	ENGLAND	18	UNIV MANCHESTER	ENGLAND
UNIV TORONTO	CANADA	17	UNIV HLTH NETWORK	CANADA

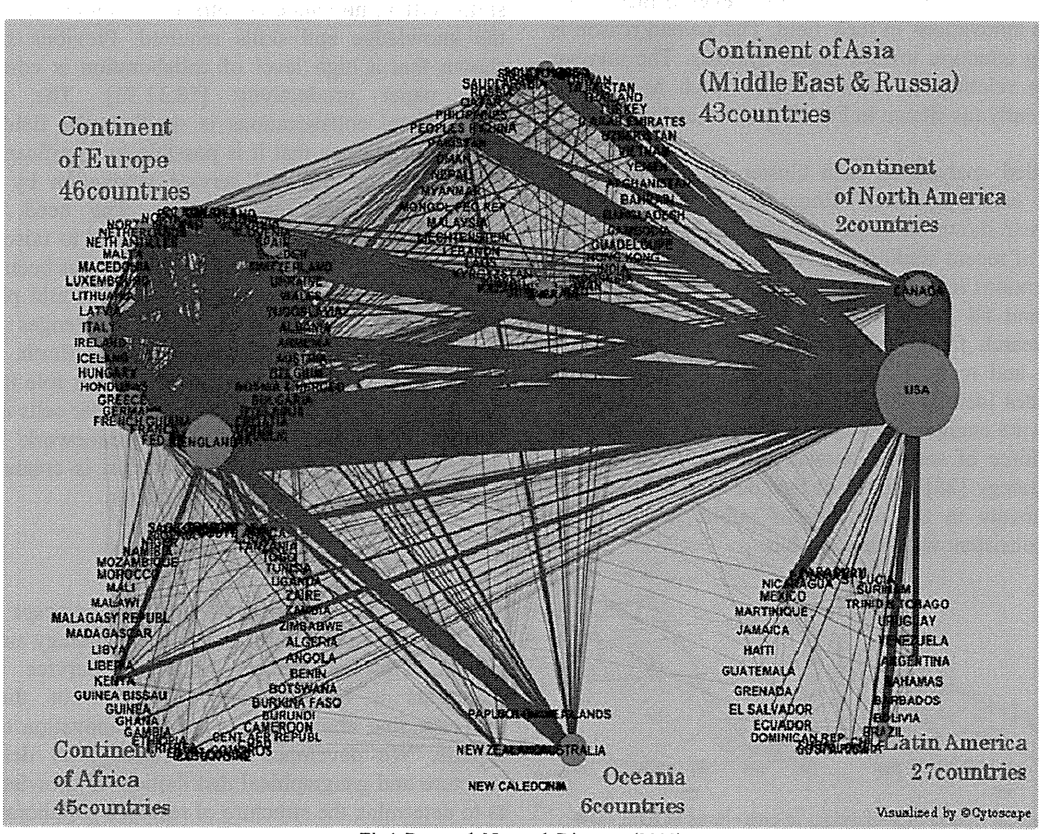

Fig4. Research Network Diagram (2009)

IV. DISCUSSION

As described above, we performed citation network analysis on SSME research domain. Our basic idea was papers dealing with a similar topic cite each other and are strongly connected, and papers dealing with different topics are weakly connected. Therefore, the division of a knowledge domain into strongly connected clusters by citation analysis can detect what kinds of topics are discussed in the SSME research domain. In the result, we could find there were mainly eight clusters. Moreover, this SSME research domain is so interdisciplinary that each of eight is not so strongly related to others.

The eight major clusters we extracted can be divided into two groups; basic research (#1 management, #4 ecosystem, and #5 QOS) and application for society (#2 medical care, #3 mental health care, #6 public service, #7 public medical care and #8 IT & Web). It is worth to be pointed out that SSME tends to deal the topics of public social systems, such as #2 medical care, #3 mental health care, #6 public service, and #7 public medical care, in terms of service innovations. This point is different from the definition by Spohrer et al[4]. As long as we discussed with the experts, there might be two reasons. The first one is that the lack of popularity of the concept SSME. Especially in the research fields which have clear boundary and have been already industrialized, the researcher might not mention about SSME even if they wrote about service innovations in their field. The second reason is the increasing attention toward public systems. The number of researches relating to public service, such as #6, has increased recently (as shown in Fig. 5(b) described in the next paragraph).

The detailed analysis of each cluster can reveal which clusters are emerging ones. Regarding to the average publication year shown in Table 1, #1, #4 and #8 seemed to contain a lot of recent studies. Fig. 4 indicates the number of annual publications in each cluster. In this figure, clusters #1, #2, #4, #6 and #8 are still so growing that they can be emerging research fronts, while #3 and #5 seems to peak around 2000 and to be mature at the end of 2008. Many previous studies identify that IT and web such as computer science, software engineering and grid computing are the base and driving force of service innovation [4, 26, 27]. OECD innovation strategy [28] points out that there is considerable scope to innovate in the delivery of public service. Our findings are consistent with these studies.

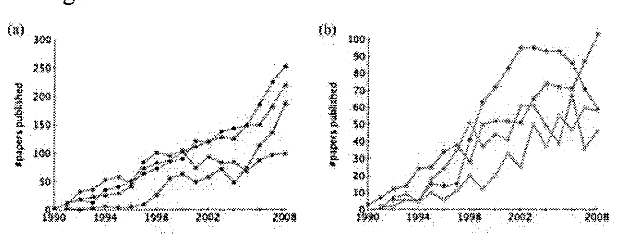

Figure 5. The annual number of publications in each cluster in 2008; (a)●: #1, ▲: #2, ■: #3, ★:#4 and (b)◆:#5, ×:#6, +: #7, O:#8.

With this method, we can extract the research topics in interdisciplinary research area such as service innovation by computational calculation. Currently, some research activities tend to research common concepts, crossing the boundaries of existing research areas or journals. However, we face increasing difficulty to create an academic landscape of these diverse research domains. Our topological approach can become a tool for future "Research on Research" (R on R) and can meet a commensurate increasing need as scientific and technical intelligence to discover emerging research fronts in an era of information flooding.

The research network diagram shows the knowledge distribution and collaborative relationships in the field of SSME objectively. In terms of the structure of co-authorship, there are more co-authorship relations between organizations in the same country or a close spatial proximity. This corresponds with other previous studies identifying the relationship between co-authorship of organizations and spatial proximity, culture, and language [18,19,20,29,30]. Furthermore, there are more co-authorship relations between organizations with high research competency. The motivation for this may include some of what Bozeman and Corley points out: access to expertise and equipment, to obtain prestige or visibility, to gain tacit knowledge, and to enhance productivity [31]. Modern technology is increasingly complex and demands an ever-widening range of knowledge and skills. Often, no single country or institution will possess all the knowledge and skills required. Previous studies have shown that a high level of collaboration is correlated with high paper productivity [18,32,33]. The number of international collaborations is small in the field of SSME. This fact indicates that it is possible to significantly enhance the efficiency of global service innovation by adopting a policy to promote collaboration. The need for service innovations to fuel economic growth and to raise the quality and productivity levels of services has never been greater [6]. In addition, SSME is a technology that could play a major role in finding solutions to global challenges such as an ageing society and sustainability of the Earth. Framework Programmes of Europe had played a major role in promoting collaborative researches in the field of solar cells in and out of Europe. We hope that a similar framework to promote researches by international cooperation is created based on the methodology discussed in this paper.

V. CONCLUSION

It is widely recognized that the concept of service innovation is significant for innovation strategy and economic growth. However, since the term "service innovation" represents a broad sense, there is not the common understanding about what is service innovation even among experts. We developed a methodology to determine the structure and geographical distribution of knowledge, as well as to determine the structure of research collaboration in such an interdisciplinary area as service innovation by performing journal information analysis, network analysis and visualization. Our results showed that there were mainly two groups of elements relating to service innovation: applications of service innovation such as health and medical care, IT and web, and public service; and basic theories for service innovation such as management, ecosystem, and QOS. Knowledge in these areas has been growing rapidly in recent years. In particular, the fields of ecosystem and IT and Web are exhibiting a high growth. We also demonstrated that the global network of knowledge is formed around the powerful hub of the US. On the other hand, in research collaboration, we demonstrated that most research is conducted within same country or a close spatial proximity and, therefore, there is big room left for enhancing international collaborations. These MAPs and diagrams could be useful in forming policies to promote service innovation. We proposed creation of an international collaboration programme to solve global challenges such as an ageing society and sustainability of the Earth.

In this study, the term "SSME" was used in developing queries. It cannot be denied that the definition of the term affects analysis. Determining a better query setting suitable in the interdisciplinary area is another subject to be studied in the future.

ACKNOWLEDGMENT

This paper was partly supported by a grant from the Economic and Social Research Institute (ESRI), Cabinet Office, Government of Japan. The Views contained here reflect those of authors.

REFERENCES

[1] Chesbrough, H., "Toward a science of services," *Harvard Bus. Rev.* vol.83, pp. 16-17, 2005.

[2] Chesbrough, H.and J. Spohrer, "A research manifesto for service scinece. Commun," *ACM* vol.49(7), pp.35-40, 2006.

[3] Maglio, P.P. and j. Spohrer, "Fundamentals of service science," *Journal of the Acad. Mark. Sci.*, vol.36, pp.18-20, 2008.

[4] Spohrer, J., L.C. Anderson, N.J. Pass, T. Ager and D. Gruhl, "Service science," *Journal of Grid Computing*, vol.6, pp.313-324, 2008.

[5] Wu, Liang-Chuan and Liang-Hong Wu, "Service engineering: an interdisciplinary framework," *Journal of Computer Information*, vol.51(2), pp.14-23, 2010.

[6] Spohrer, J. And P.P. Maglio, "The emergence of service science: Toward systematic service innovations to accelerate co-creation of value," *Production and Operations Management*, vol.17(3), pp.238-246, 2008.

[7] Chubin, D.E. , A. L. Porter and F.A. Rossini, "Interdisciplinary research: the why and the how," in *Interdisciplinary Analysis and Research*, eds. D E Chubin, A L Porter, F A Rossini and T Connolly. Mt Airy, MD: Lomond Publications, 1986.

[8] Porter, A.L. and D.E. Chubin, "An indicator of crossdisciplinary research," *Scientometrics*, vol.8, pp.161-176, 1985.

[9] Porter, A.L. , A.S. Cohen, J.D. Roessner and M. Perreault, "Measuring researcher interdisciplinarity," *Scientometrics*, vol.72, pp.117-147, 2007.

[10] Kostoff, R.N., R. Tshiteya, K.M. Pfeil, J.A. Humenik and G. Karypis, "Science and technology text mining: electric power sources," *DTIC Technical Report* No. ADA421789, 2004, National Technical Information Service, Springfield, VA.

[11] Kostoff, R.N., R. Tshiteya, K.M. Pfeil, J.A. Humenik, and G. Karypis, "Power source roadmaps using bibliometrics and database tomography," *Energy*, vol.30, pp.709–730, 2005.

[12] Tijssen, R.J.W., "A quantitative assessment of interdisciplinary structures in science and technology: co-classification analysis of energy research," *Research. Policy*, vol.21, pp.27–44, 1992.

[13] Klavans, R. and K. W. Boyack, "Identifying a better measure of relatedness for mapping science," *Journal of the American Society for Information Science and Technology*, vol.57(2), pp.251-263, 2006.

[14] Rinia, E. J., T.N. van Leeuwen, h.G. van Vuren, and A. F. J. van Raan, "Influence of interdisciplinarity on peer-review and bibliometric evaluations in physics," *Research Policy*, vol.30, pp.357-361, 2001.

[15] Kajikawa, Y., J. Ohno, Y. Takeda, K. Matsushima and H. Komiyama, "Creating an academic landscape of sustainability science: an analysis of the citation network," *Sustainability Science*, vol.2, pp.221-231, 2007.

[16] Smith, M., "The trend toward multiple authorship in psychology," *American Psychologist*, vol.13, pp.596–599, 1958.

[17] Frame, J.D. and M.P.Carpenter, "International research collaboration," *Social Studies of Science* vol.9, pp.481–497, 1979.

[18] Narin, F, K.Stevens, and E.S.Whitlow, "Scientific co-operation in Europe and the citation of multinationally authored papers," *Scientometrics*, vol.21, no.3, pp.313–323, 1991.

[19] Moed, H.F., R.E.de Bruin, A.J.Nederhof and R.J.W.Tijssen, "International scientific co-operation and awareness within the European Community: Problems and Perspectives," *Scientometrics*, vol.21, no.3, pp.291–311, 1991.

[20] Hoekman, J.,K.Frenken, and R.J.W.Tijssen, "Research Collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe," *Research Policy*, vol.39, pp.662–673, 2010.

[21] Katz,J.S., and B.R.Martin, "What is research collaboration?" *Research Policy*, vol.26, pp.1–18, 1997.

[22] Shibata, N., Y. Kajikawa, Y. Takeda and K. Matsushima, "Comparative study on methods of detecting research fronts using different types of citation," *Journal of the American Society for Information Science and Technology*, vol.60(3),pp. 571-580, 2009.

[23] Newman, M. E. J., "Fast algorithm for detecting community structure in networks," *Physical Review E*, vol.69, 066133, 2004.

[24] Adai, A.T., S.V. Date, S. Wieland and E.M. Marcotte, "LGL: Creating a map of protein function with an algorithm for visualizing very large biological networks," *Journal of Molecular Biology*, vol. 340(1), pp.179-190, 2004.

[25] Sasaki, H., Y. Kajikawa, K. Fujisue and I. Sakata, "Detecting the valley of international academic collaboration in renewable energy," *Proc. of the IEEE International Conference on Industrial Engineering and Engineering Management* (IEEM2010), 2010.

[26] Bardhan, I.R., H. Demirkan, P.K. Kannan, R.J. Kauffman and R. Sougstad, "An interdisciplinary perspective on IT services management and service science," *Journal of Management Information*, vol. 26(4), pp.13-64, 2010.

[27] Demirkan, H., R.J.Kauffman, J.A. Vayghan, H.G. Fill, D. Karagiannis and p.P. maglio, "Service-oriented technology and management: Perspectives on research and practice for coming decade," *Electric Commerce Research and Applications*, vol.7(4), pp.356-376, 2008.

[28] OECD, "*Innovation Strategy: Getting a head start on tomorrow*," OECD publicatiion, 2010.

[29] Glänzel, W., "National characteristics in international scientific co-authorship relations," *Scientometrics*, vol.51, no.1, pp.69–115, 2001.

[30] Schott,T., "International influence in science: Beyond center and periphery," *Social Science Research*, vol.17, pp.219–238, 1988.

[31] Bozeman, B. and E.Corley, "Scientists' collaboration strategies: implications for scientific and technical human capital," *Research Policy*, vol.33, pp.599–616, 2004.

[32] Lawani, S.M., "Some bibliometric correlates of quality in scientific research," *Scientometrics*, vol.9, pp.13–25, 1986.

[33] Pao, M.L., "Co-authorship and productivity," *Proc. of the American Society for Information Science*, vol.17, pp.279–289, 1980.

JJCO Japanese Journal of Clinical Oncology

Jpn J Clin Oncol 2011;41(5)723–729
doi:10.1093/jjco/hyr036
Advance Access Publication 29 March 2011

Meeting Report

The 6th Asia Cancer Forum: What Should We Do to Place Cancer on the Global Health Agenda? Sharing Information Leads to Human Security

Norie Kawahara[1,*], Haruhiko Sugimura[2], Akira Nakagawara[3], Tohru Masui[4], Jun Miyake[5], Masanori Akiyama[6], Ibrahim A. Wahid[7], Xishan Hao[8] and Hideyuki Akaza[1]

[1]Department of Strategic Investigation on Comprehensive Cancer Network, Research Center for Advanced Science and Technology (RCAST), The University of Tokyo, Tokyo, [2]Department of Pathology, Hamamatsu University School of Medicine, Hamamatsu, [3]Chiba Cancer Center, Chiba, [4]Department of Disease Bioresources Research, National Institute of Biomedical Innovation, Osaka, [5]Graduate School of Engineering Science, Osaka University, Osaka, [6]Policy Alternatives Research Institute, The University of Tokyo, Tokyo, Japan, [7]Malaysian Oncological Society, Malaysia and [8]Tianjin Medical University, China

*For reprints and all correspondence: Norie Kawahara, Department of Strategic Investigation on Comprehensive Cancer Network, Research, Center for Advanced Science and Technology (RCAST), 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan. E-mail: norie.kawahara@med.rcast.u-tokyo.ac.jp

Received January 12, 2011; accepted February 8, 2011

This forum discussed issues relating to the inclusion of cancer on the global health agenda, with the ultimate aim of achieving human security for all people. The forum discussed what methods are available to the cancer community in attempts to create a common data system for the rapidly growing Asian region. Discussions also focused on the preparations that can be made to consider and respond to the obstacles to the creation of an Asia-wide data and information network. It was also noted that in order to create a cancer information network, support would need to be provided to low- and middle-income countries and efforts made to ensure that data are comparable.

Key words: cancer information network – MDGs – human security – data comparability

OVERVIEW

The Asia Cancer Forum is a grouping that aims to discuss cancer science and policy issues among Asian countries. The basic concept of the forum is that discussion will enhance sharing and awareness of the issues and each of the participants will gain their own take-home message to apply to their own activities as the outcomes of the forum. The forum is operated through the research funds of the participating members and receives support in the form of Health and Labour Sciences Research Grants from the Ministry of Health, Labour and Welfare of Japan, as part of the Third Term Comprehensive Control Research for Cancer or its ongoing work to create an Asian network. The organizer of the forum is N.K. and it is chaired by H.A., both of the Research Center for Advanced Science and Technology (RCAST), the University of Tokyo.

The origins of the Asia Cancer Forum date back to 2004 when a group of Asian researchers launched a platform called the Asia High Technology Network to discuss issues in the field of medicine. The grouping engaged in discussions on the formation of an Asia Cancer Information Network. From 2008, the name of the research platform was changed to the Asia Cancer Forum and the first two meetings were held thereafter. The third meeting was held in February 2009, on the theme of 'Health, Information and Development'. The third meeting was held jointly with SciDev.Net and saw discussion focus closely on issues relating to the setting of the global health agenda. The fourth meeting was held in April 2009 under the theme of

'Asian Challenges in Shifting the Disease Burdens'. In November 2009, the fifth meeting was held in collaboration with the 20th Asia Pacific Cancer Conference (APCC) under the theme of 'What Should We Do to Raise Awareness on the Issue of Cancer in the Global Health Agenda?' The meetings to date have concentrated on ways to share information among Asian research colleagues, thus raising awareness of the importance of including cancer on the global health agenda.

The Sixth Asia Cancer Forum was held in conjunction with the World Cancer Congress UICC 2010, on 21 August 2010, in Shenzhen, People's Republic of China. The meeting consisted of two sessions consisting of six special presentations, followed by detailed discussions. Approximately 60 people were present and active discussions took place. The forum was organized by H.A. and N.K. (RCAST, the University of Tokyo). Invited speakers included H.S. (Hamamatsu University School of Medicine), A.N. (Chiba Cancer Center), T.M. (National Institute of Biomedical Innovation) and J.M. (Osaka University). Also in attendance were Joe Harford [National Cancer Institute (NCI)], Julia Schneider (NCI), X.H. (Chinese Anti-Cancer Association, President of World Cancer Congress UICC 2010), Andreas Ullrich (World Health Organization), David Hill (UICC) and I.A.W. (Asian and Pacific Federation of Organization for Cancer Research and Control).

INTRODUCTION TO THE ASIA CANCER FORUM

N.K. (RCAST) gave an introduction to the ongoing activities and initiatives of the Asia Cancer Forum. She noted that the incidence of infectious diseases in developing countries and the delay in formulating measures to respond to these diseases are recognized as issues requiring the attention of industrialized nations. Accordingly, infectious diseases are given due recognition on the global health agenda. However, cancer has still to gain the recognition it rightly deserves in the world of global health. This is due to the fact that it is generally viewed as a disease specific to individuals in industrialized nations, which occurs as a result of the individual's approach to personal health management.

Last year, the Fifth Asia Cancer Forum discussed issues relating to cancer and concluded that the highest priority should be for expert groups to share a common recognition of the necessity for cancer to be raised on the global health agenda. In the international community, there has also been increasing recognition of the necessity to 'begin discussion on placing cancer on the global health agenda', as evidenced by the Resolution of the United Nations on 13 May 2010 to hold a United Nations General Assembly Summit on Non-Communicable Diseases (NCDs).

However, the results of a survey implemented by the Asia Cancer Forum in April 2010, on the occasion of the 101st Annual Meeting of the American Association for Cancer Research (AACR), entitled 'Survey on Inclusion of Cancer in the Global Health Agenda', showed that interest in this issue is not particularly high among a great majority of specialists. Discussion on the inclusion of cancer on the global health agenda does not stop merely at the advocacy of humanitarian principles. In fact, what is needed now is a move away from the linear debate such as that which has dominated discussions of aid to developing nations in the past, and a move toward more complex projections. Therefore, it is necessary to gain the broad participation of cancer researchers in working to decipher the current challenges faced by industrialized nations, which could then be utilized in assistance to developing nations. In other words, it is necessary to establish a framework for resolving issues that face industrialized nations.

The world is now in an era in which developments in health innovation have a significant impact on the direction for global health.

In the initial stages of the genome-wide association study, it was thought that genetic differences by race increased the predisposition to the occurrence of a particular disease. However, as research has advanced, it has shown that although there are some statistics differences among races according to the genetic background, the genetic factors predisposing a person to the occurrence of disease are clearly shared by all humankind. In other words, any careful observations made in one specific region of the world are relevant to other regions.

Owing to the tremendous improvements in genome analysis capabilities, it is now possible to analyze genetic information to an incredibly detailed level. Furthermore, information technology (IT) has enabled quantitative tracking of the vast amounts of medical-related data that are created in the modern world. By continuously and automatically collecting and gathering information from various sources, including clinical data and medical records, and using this information to realize the creation of a system that would produce the required evidence for the purpose of providing each patient with the most appropriate and latest medical treatment, research could be used in a synergetic partnership with treatment. Through such technological breakthroughs, it would be possible to search out information relating to the culturally diverse acquired lifestyle customs that exist in Asia, even in persons of similar race, and work to reduce risk factors and even help to prevent further epidemics. It is for this reason that rather than basing research on persons of ethnicities removed from Asia, careful study of the fine differences that exist among the races and nationalities of the Asian region would result in a closer understanding of the nature of diseases in humanity as a whole.

Infectious diseases are characterized by their tendency to infect many people, while the variation in the disease itself is not so great. However, non-communicable diseases, and cancer in particular, have the characteristic of presenting differently from person to person. In other words, it can be seen that 'to understand cancer it is important to look at the differences among individuals'. In a region with genetic similarities, in which a diversity of acquired lifestyle customs co-exist, would it not be possible therefore to gather significant data through cohort research in the region?

Progress in science bestows upon people the promise of limitless possibilities and the means to live longer. Humankind has devoted much time and effort in the fight against disease.

In the near future, the international community is likely to face an unjust situation in which some people with the same disease will be cured while others will suffer and die.

It is this grave reality that must be addressed.

The Asia Cancer Forum bases its activities on the Universal Declaration of Human Rights, which states that everyone has the right to share in scientific advancement and its benefits equally. In aiming to utilize scientific advancement to address the issue of what we can do to ensure that the challenges that have been faced by industrialized nations are not faced by developing nations, the Asia Cancer Forum is engaging in discussion on the challenge common to both industrialized and developing nations, namely the inclusion of cancer in the global health agenda.

SESSION 1: INFORMATION FROM THE HUMAN BODY

H.A. (RCAST) opened the meeting by requesting comments from Andreas Ullrich (WHO) and David Hill (UICC). Andreas Ullrich noted that the WHO is working very hard to include cancer in the context of NCDs on the global health agenda. What needs to be done on a global scale is exactly what is happening in the Asia-Pacific region in fora like the Asia Cancer Forum, and these activities are very much in line with WHO strategies and policies. David Hill noted that the Asia Cancer Forum is a series of important discussions on the issue of cancer. He stated that the UICC is a global organization, but has a particular concern about cancer control in low- and middle-income countries, many of which are in Asia. There is enormous potential for cancer control, which is currently not being fully implemented. As a species, human beings are very wasteful of the benefits of discoveries. We are not very good at implementing the benefits of discoveries as rapidly, effectively and equitably as we should. Forums such as the Asia Cancer Forum, which focus not only on research and discovery, but more importantly focus on delivering the benefits of research and discovery to populations, are extremely important and are to be commended. The solutions to cancer control lie in people connecting with each other and with their communities to implement the benefits of knowledge that we already have, and that is exactly what the Asia Cancer Forum is doing here.

H.A. presented the concept for this Asia Cancer Forum. The previous APCC was held in Japan and resulted in the issuance of the Asian-Pacific Consensus Statement by working groups, which aims to improve cancer health science in the Asia-Pacific region. At this discussion last year, it was concluded that cancer must be on the global health agenda and the Asia-Pacific region is ready to work toward this goal. The issues being currently faced are a rapid increase in population in Asian countries, an aging society and increased longevity, together with increased speed in diagnosis. For example, the population of China has a different demographic to that of Japan, but it will gradually come to look like the demographic pyramid of Japan in the future. Expenditure is also rapidly increasing in Asia. Comparisons between the E7 and G7 countries show that medical expenditure is rapidly increasing in E7 countries. Japan has a track record of good healthcare and low spending in terms of GDP. The low spending in Japan has created a number of issues, particularly with regard to the quality of life for medical staff. In other words, Japan has faced a number of cancer issues ahead of other Asian countries and Japan could provide a source of reference for other countries that will face these issues in the future. The aim of the Asia Cancer Forum is to come up with good proposals.

N.K. noted that 'Genetic Solidarity and Altruism' is a powerful phrase that features in the 'Inside Information' documents of the Human Genetics Commission (HGC) of the UK. The progress of innovation means that the significance of holding information and data is changing greatly. What is most important, however, is to ensure that each and every person transforms their awareness about the importance of information in an innovative world.

The Asia Cancer Forum is a body that is committed to strategic analysis in the area of cancer research. The current objective of the Forum is to achieve the inclusion of cancer in the Millennium Development Goals (MDGs) of the United Nations. A long-term perspective must be taken that looks ahead to the issues that will face future generations. It is important to start to consider the design of a social system for collecting and storing the information and data we ourselves posses.

Pathology Networking in Asia

H.S. (Hamamatsu University School of Medicine) noted that cancer diagnosis is based on histopathological pictures and human pathology and cancer diagnosis is a mature scientific field. Histopathological language is common to all oncologists and other cancer specialists and it is now possible technologically to present histopathological pictures. Data can be stored and uploaded on a virtual slide website for joint use. Using this website, scientists worldwide could input their own opinions. Archives stored in digital format can last for almost forever. The virtual slide website is easy to use and browsable. There are many folders on the website featuring histopathological archives, for educational and research purposes, as well as for quality control. Each hospital can send images to a central hospital for diagnosis and compare images among multiple hospitals. The quality of the pictures is much higher than conventional cameras. With high-speed Internet, it is possible to scan images to high resolutions. For virtual slides, no microscope is necessary, only a high-resolution CCD camera. The problem at the moment we face concerns Internet speed. Eventually, with the dissemination of broadband, this system will be able to be further improved around the region. Scanners are installed in 300 institutions at the moment. Histopathological diagnosis can therefore be performed 24 h around the clock using the worldwide network. In order to expand the network further, it will be necessary to develop infrastructure, including high-speed broadband Internet.

Urgent Demand to Establish Asian Network of Pediatric Bio-resource and Tumor Banks for Better Cure of the Sick Children

A.N. (Chiba Cancer Center) talked about the urgent demand to establish bio-resource and tumor banks in order to better cure sick children. The cure rate of pediatric cancer is very low in many countries in Asia. Epidemiology of childhood cancer in developing countries is largely unknown. It is not known what genetic and environmental factors affect pediatric cancers, in contrast to the knowledge available on adult cancer. It is important to establish a standardized therapeutic and diagnostic system, which would be helpful for the development of epidemiology of pediatric cancers. In 2008, at the meeting of the Advances in Neuroblastoma Research (ANR2008) held in Chiba, Japan, the Steering Committee and the Advisory Board Committee of the ANR Association decided to take an action to establish the international neuroblastoma tumor bank (INTB). The INTB task force includes the establishment of a standardized diagnostic and database system. Neuroblastoma is a very enigmatic tumor, with most being very aggressive. Prognosis is very poor, even now. In order to solve this problem, a staging system was proposed. In order to promote new translational research in the field of cancer, it is necessary to establish a tumor bank system. More than 90% of neuroblastoma tumors in Japan are being sent to Chiba University for analysis. Chiba Cancer Center engages in genomic analysis of these tumors. Efforts are being made to propagate our standardized system to other countries in Asia. All countries agreed to establish a tumor bank; however, the central tumor bank and molecular diagnosis systems are still immature in Asian countries.

Why Do We Need Global Collaboration in Cancer Research? Establishing Cross-border Transfer of Research Materials and Information

T.M. (National Institute of Biomedical Innovation) introduced one example of networking and commented on why a network is required, particularly in the Asian context. NCI is working to develop a bio-bank system in the USA. This is a very important attempt to share information and materials among cancer researchers, although it is currently limited to within the USA. Best practices are also issued by the NCI, the first version being issued in 2007. Diagnosis and treatment is not the end of a process, it should be the start for the next generation of research. It is therefore important to achieve integrity between clinical practices and research activities. The NCI also focuses on biomarkers, with the aim of providing transcripts for future use. The common practice for conventional medical research requires a large number of medical researchers and specialists. Researchers tend not to see the bigger picture behind research and it is therefore important to provide transparency in large projects so that researchers can understand their place in the research context. The creation of an international network would therefore be very important. A greater degree of cross-border fluidity is required, working on the already good level of interaction between cancer specialists across borders.

SESSION 2: INFORMATION AS IT SIGNALS

Global Strategies for Genome and Cell-based Informatics: High-performance DNA Sequencing and Expression Analysis Open a New Area

J.M. (Osaka University) explained the need for an Asian network from the viewpoint of engineering. Fighting against cancer is not simple. Everyone in the pharmaceutical industry is now seeking how to control the pathways and molecular systems of cancer cells. We require huge knowledge in order to achieve this aim, as cancer molecules have an enormous number of variations. Four-dimensional data are required to identify cancer pathways. In our laboratory, we have 200TB of data processing capacity, in order to engage in DNA processing, which provides us with a great deal of data. In Okinawa, we have 10 GB sequencers. We know that medical research is already at a very high level, but R&D remains at a low level, as a part of total expenditure. We therefore have to have more information-based medical systems. We need a system that all stakeholders would be able and ready to use. We have been working on the creation of a network and would like to ask you to join us in our efforts.

Tackling the 'Lifestyle-related Cancer' with Cutting-edge IT

M.A. (University of Tokyo) talked about how to build consensus and share information using IT. Aging society is a serious issue as people are susceptible to other diseases in addition to cancer. In general, the collection of information data is generally done from the bedside. The next-generation system would have to be an interactive system. Cutting-edge systems including bar-code systems and wireless devices would help to create and disseminate data. Another issue is how to gather verbal information using IT. Next-generation data entry systems will need to incorporate measures for gathering verbal information in data format. Cloud computing could solve issues of data storage in the future, as the storage capacity using cloud computing is virtually limitless and would enable further collaboration, including data entries from patients' homes, etc. If cloud computing is to be used, it is essential that the systems are secure and trusted.

DISCUSSION

H.A. (RCAST) noted that it is essential that all Asian countries share information, technology and knowledge. He invited comments from other participants.

X.H. (Chinese Anti-Cancer Association) noted that Asia needs a forum to focus on the problems facing Asia. Fifty percent of new cancer cases annually occur in Asia, and from the presentations made at the 21st UICC World Cancer Congress, it is known that 80% of new cases of cancer are from low- and middle-income countries, like China, India and Pakistan and other countries in Asia. The issues raised by the presenters are very important and require action. Although there is a lot of knowledge and consensus on most cancers, we still need further information and consultation on some forms of cancer, including pediatric cancers, leukemia and central nervous system cancers, for example. The possibilities for medical consultation through the Internet would be of benefit not only for Asia but for the world, and would facilitate diagnosis for patients and help to diagnose and identify the correct therapies for patients and save their lives. The issue of a tissue bank is also very important. Six years ago, with the support of the National Foundation for Cancer Research (NFCR) from the United States, a Joint Tissue Banking Facility was opened at the Tianjin Medical University Cancer Institute and Hospital in China. Right now there are about 40 000 specimens. An Asian network is essential and Japan is leading the way on this project.

H.S. (Hamamatsu University School of Medicine) noted that Chinese pathologists have many more cases than ordinary Japanese pathological institutions, maybe due to the numbers of people who have variations of tissues. The Internet is a very comfortable way of developing relationships and colleagues in China and Asia should be encouraged to continue to develop such consultation systems.

Joe Harford (NCI, USA) pointed out that through the practice of tele-pathology, it is possible to have samples read in the USA that were collected in Japan during the night and thereby operate around the clock. In contrast, it is instructive to look at the situation that was encountered with pathology services in Ghana. When the Breast Health Global Initiative visited Ghana, the breast pathology reports were taking 6 months to complete, from the time the samples were collected, until the pathology report was submitted. The idea of getting a report in 18–24 h is very different from waiting for 6 months. Tele-pathology does have a great deal of potential for assisting low- and middle-income countries, where there are few pathologists. It is therefore incumbent on the USA and the Asian region to be thinking about how these technologies can be used to assist the low- and middle-income countries where there are no or few pathologists. This could be in the form of training, or it could be in the form of reading the samples. In the case of Ghana, there was a pathologist in Norway who agreed to train the Ghanaian pathologists so that it became possible to get a much quicker pathology report as a result of training. However, in this case, it required North–South cooperation. Efforts should be made to share resources with the low- and middle-income countries.

A second issue raised by Joe Harford was that of tumor banking. The exchange of samples across borders presents significant problems. Each country has its own restrictions on how samples flow across borders. Hypothetically, there is no need to ever ship a sample across a border. All that is required is to have comparable sample collection everywhere, and the equipment to analyze those samples everywhere, and then the information could be shipped across borders. It ought not to be necessary to ship samples across borders, theoretically. This would require a certain amount of standardization. One of the things that the NCI has been engaged in with the bio-banking effort is best practices and standardization, which is an ongoing effort. In order to ensure that there is comparability across borders requires a small number of samples collected in Japan, for example, to be tested in China or the USA, so that you can assure yourselves that comparability has been achieved. Once comparability has been assured then you ought not to need to ship samples. All of the countries that are involved in a network of collaborative bio-banking should be encouraged to work with governments, and perhaps with the WHO, to make these provisions that would at least allow for these small studies in comparability to be implemented.

The term 'comparability' is an interesting word, but it does not necessarily mean uniformity. This particularly applies to informatics platforms and cancer registries and the software that is used for cancer registries. These are not uniform, but they can still be comparable. Databases in particular do not have to be uniform, but it is important to create 'adaptors' that would enable data gained in one country to be usefully compared in another country. It is not expected that the world will uniformly follow US or other standards, but in the interests of collaboration, the opportunity to adapt between systems and be able to compare is essential.

Julia Schneider (NCI, USA) congratulated the Asia Cancer Forum for specifically talking about developing platforms for enhancing collaboration within and outside of Asia. There is tremendous potential in the age of genomics and proteomics to do meta-analysis of large collections of specimens. It is important to ensure that specimens are comparable. After the initial quality control is implemented, it makes sense for specimens to be analyzed in the country in which they were gathered. It is very exciting that these sorts of issues about creating and developing platforms and infrastructure are being discussed in this forum.

With regard to the NCI best practices, the new version is now published and is available on the website for comment. NCI is very actively interested in receiving comments on this new version. The process that was used for developing the NCI best practices was very focused on the USA. It would be good to continue the dialogue about developing standards that can be implemented effectively in both Asia and the USA and other parts of the world. In the USA alone, many challenges were encountered in terms of the way that different institutions were engaging in analysis, both from the technical side and also the ethical and legal issues (informed consent, privacy protection etc.). These issues become even more complex in the context of cross-border collaboration, but it is extremely important to develop and facilitate such international collaboration.

M.A. (University of Tokyo) noted that with cutting-edge IT, it is possible to create information not only for cancer but also for diabetes and other diseases. Lifestyle-related cancer is a chronic disease. The cost for hemodialysis and treatment of cancer is very expensive. It would be possible to use cutting-edge IT to create systems that would be applicable to a variety of lifestyle-related diseases.

I.A.W. (Asian and Pacific Federation of Organization for Cancer Research and Control) reported that in the Southeast Asian context, it is necessary to look at more fundamental issues, because there are discrepancies in the region with standards of health care. There are some parts of the region where there are no people who diagnose or even treat patients. In order to look at the cancer agenda, we need to look at the issue in global terms. For example, take a country like Malaysia, in Kuala Lumpur, there are 15–20 cancer centers within a radius of 25 km, but in other regions, there are no physicians who are qualified to provide cancer care. These are issues that need to be examined. Hospitals treating cancer in the Southeast Asian region have to endure a tremendous burden, where, in some cases, patients have to share beds in a cancer hospital and 200–300 patients are having chemotherapy in a single day. It is therefore important to examine the manpower problem. Part of the issue here is improving the standards of diagnostic care, sharing of pathology and maybe radiology reports through the Internet, but we must also consider how we address the issue of manpower shortage. There are parts of the region where there are no cancer specialists. It is important to think about these important issues of manpower and consider how we can improve this from an Asian perspective.

Andreas Ullrich (WHO) noted that it is important that the Asia Cancer Forum is an open platform for all countries, including low-, middle- and high-income countries. Linking all these countries toward a common goal is extremely important. One of the major drivers in decision-making in the political circles is the availability of data. It is important not only to know how many cases of cancer are occurring, but also to know about the number of staff who are available in each country. Also, we must consider the availability of technology, including diagnostic devices, essential medicines etc. The Asia Cancer Forum could be one that goes beyond the diagnosis of cancer and could be a forum for collecting data about infrastructure. It could provide information through the Internet and other tools could be developed (or are already developed by the WHO) about capacity in countries. This information could then be combined not only for academic purposes but also for a policy forum, where intelligence is translated into policy proposals to politicians. The politicians could then be shown data about incidence of cancer, mortality and survival rates etc. Survival data are very strong drivers in political decisions, as we have seen in Europe. They are not available universally across the Asian region. There is great potential for this forum to set an agenda for what needs to be achieved in terms of political decision-making and will be required to achieve that target.

Massoud Samiei [International Atomic Energy Agency (IAEA)] noted that in order for donors to invest in cancer, it is important to have convincing projects to show that something can be done about cancer. Cancer is perceived as a very expensive disease. The IAEA works with the WHO in many developing countries, including in Asia, to establish cancer centers, and often donors ask about investing in cancer as it is a very expensive disease. In order to get cancer on the MDGs, it is essential to show that there are strategies and solutions that are cost-effective. With a little investment, progress can be made in terms of prevention, screening programs and focusing on specific types of cancer. For this, we have already created examples through the IAEA programs across the globe. The IAEA could collaborate with the Asia Cancer Forum to provide information for the creation of a proposal to submit to the UN. Donors are only interested in cost-effective solutions. The IAEA has pilot projects in eight countries currently and could share these results with the Asia Cancer Forum.

CLOSING

H.A. and N.K. thanked the speakers and participants for their insightful comments and active participation. In closing, it was noted that the ultimate goal is to utilize advances in innovation to create a large database of knowledge and a global network for analyzing data and sharing information. To this end, it is essential to make efforts to collect all kinds of medical information. The opinions raised at the forum concerning means of sharing data and raising awareness among specialists and patients alike about the importance of medical information in the fight against cancer demonstrated that there is a general awareness of the issue. It was recognized that further efforts must be made to create awareness among specialist organizations of the value and necessity of setting the global health agenda for the sake of scientific development. Approaches must also be developed that enable countries and regions at different levels of development to share data in a comparable manner.

ROAD TO 7TH ASIA CANCER FORUM

The discussions at the 6th Asia Forum identified a number of key issues that need to be tackled if a comprehensive cancer network is to be achieved. Knowledge gaps exist between the current status of cancer research and treatment in front-runner countries, such as Japan, and the perception of issues in developing and emerging countries. It was recognized that the issue of obstacles to sharing common challenges is one that requires further discussion and analysis. The Asia Cancer Forum will continue to examine means for sharing information in a meaningful and comparable manner. In particular, the role of IT in opening up cancer issues for global health consideration will be focused on in future meetings, with input being sought from policy-makers in government and from the private sector, including pharmaceutical companies. The 7th Asia Cancer Forum is scheduled to be held on 3 November 2010, with invited speakers from the Asian region and major pharmaceuticals coming together to discuss the way forward for a comprehensive cancer network in Asia. With the participation of representatives of academia, government and industry at the 7th Asia Cancer Forum, it is anticipated that the technical issues, specifically relating to knowledge and know-how gaps between front-runner and developing countries, will be further discussed, with a view to crystallizing a future path for a comprehensive cancer information network in the Asian region.

Conflict of interest statement

None declared.

Preliminary Linguistic Analysis of Large Number of Medical Incident Reports 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

2.1 Overview of Medical Incident Reports

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 [4].

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 188 words, on the other hand, the one by the Project to Collect Medical Near-Miss/Adverse Event Information is 80 words[2]. 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.

2.2 Data Sets

We used free composition format written in Japanese relating to medical agents from 2007 to 2011 by Osaka City University. The number of documents is 18,340. Each case study is in a free composition format, with the abstract, background, and solution being approximately 188 words 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. METHODOLOGY OF NATURAL LANGUAGE PROCESSING AND NETWORKD ANALYSIS

3.1 *Methodologies for 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 "background/causes," "details," and "solutions," using the tfidf method. 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 of the words using the Jaccard coefficient.

Also, we show the networks among each document which are determined by the similarities between documents based on the tfidf method. 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.

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 [5]. In this research we used MeCab, one of the most common engines for conducting morphological analysis [6].

There is the possibility that words obtained using MeCab are too finely classified to conduct the analysis of links. Therefore, we connected words using the following methods and used them as new words.

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 [7]. 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 (*0.1* in this research) then we treated those two words as one word.

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 research, 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[8]. *Tf-idf* is calculated as follows:

$$tf-idf(t,d) = tf(t,d) \times idf(t) \quad (1)$$

$$tf(t,d) = n(t) / \sum_{k \in T} n(k) \quad (2)$$

$$idf(t) = \log |D| / \{d : d \in t\} \quad (3)$$

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 from the degree of similarities between words in documents. Here, the simplest co-occurrence index for finding links between the two word A and B is the number of co-occurrence $|A\cap B|$ for two words. 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 [9].

$$Jaccard: |A\cap B| / |A\cup B| \quad (4)$$

A link is established between the two words in the event that these indices exceed the threshold value.

4. PRELIMINARY ANALYSIS RESULTS

Table1: Top 10 Characteristic Word in Incident Reports (*TF*; Term Frequency)

(TF)	Background	Details	Solutions
1	Patient	Report	Check
2	Check	Patient	Time
3	Drug	Check	Patient
4	Nurse	Attending Doctor	Direction
5	Direction	Monitor	Drug
6	Internal use	Duty Doctor	Explanation
7	Infusion	Doctor	Nurse
8	Pill	Direction	Thoroughness
9	Room	Nursing	Drug maker
10	Administration	Explain	Doctor

Table2: Top 10 Characteristic Word in Incident Reports (*TF-IDF*)

tf-idf	Background	Details	Solutions
1	Drug	Patient	Check
2	Patient	Report	Time
3	Check	Attending Doctor	Patient
4	Internal medicine	Check	Direction
5	Direction	Doctor	Drug
6	Nurse	Monitor	Explanation
7	Administration	Apologizing	Nurse
8	Infusion	Nursing	Thoroughness
9	Root	Duty Doctor	Before
10	Channel	Direction	After

The top ten characteristic words that appear in the incident report such as "background/causes," "details," and "solutions" with *tf* (term frequency) are shown in Table 1. The top ten characteristic words that appear in the incident report with *tf-idf* (Eq.(1)) are shown in Table 2. Under the category of "Background," the words "Patient" "Check," "Drug," "Nurse," and "Direction" 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 "Background," the words "Drug" "Internal medicine," "Infusion," and "Channel" rank high in Table 2. Therefore, accidents related to medicines are important for analyzing the reports. Also, the words "Check," "Direction," and "Explanation" rank high under the category of solutions. In addition, the words related to the medical process such as "before" and "after" are high rank.

One word such as "lack," "confirmation," or "drugs" alone cannot express the tendency of the accident. In this research, co-occurrence networks of words were created by connecting the words that co-occurred with each other at a high frequency. The degree of co-occurrence is calculated using the Jaccard coefficient[3] shown in Section 3.4.

Figure 1 shows the networks of characteristic words created using the accident reports related to incorrect drugs. Each node represents a word, and an edge represents the intensity of the co-occurrence between the words. First of all, in viewing the network for "background/causes" (see Figure 1 (a)) it is clear that the network is created around the word "Check," and one can see that the cause of many accidents is the fact that the "Check" on "drugs," "Patient," by "Nurse". Connecting the words that co-occur frequently allows us to understand what tends to become inadequate. In the network of "accident details" (see Figure 1 (b)) many different words appear at once, indicating the presence of diverse accident details. Viewing the network for "solutions" (see Figure 1 (c)), as with the network for "background/causes," it is created around the word "confirmation."

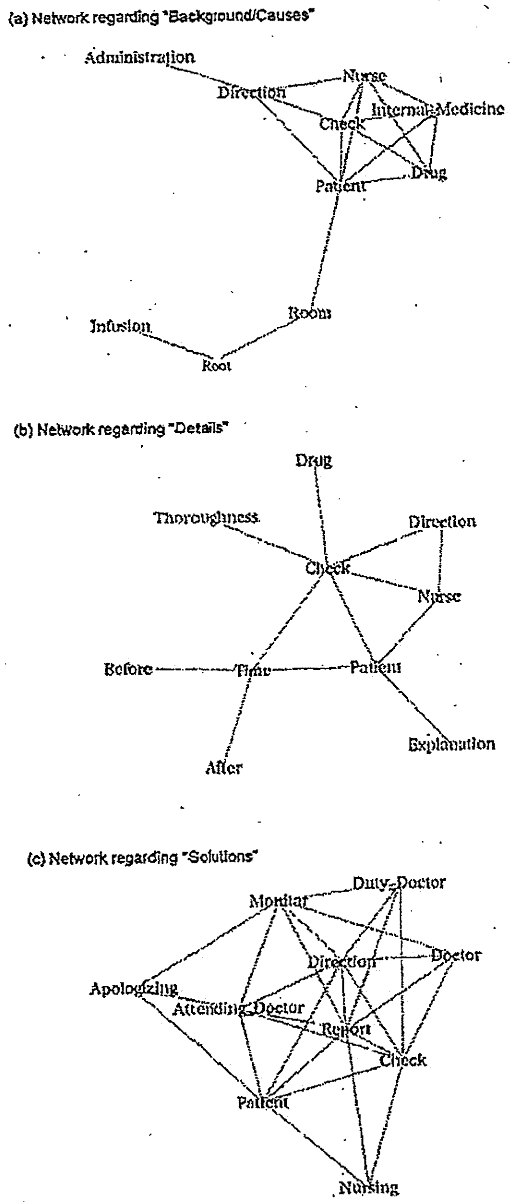

Figure 1: Co-occurrence Network of the Words in Incident Reports

5. 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 "check," 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 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.

6. REFERENCES

[1] Heinrich H W. Industrial accident prevention: A safety management approach. McGraw-Hill Customer Service. 1931.

[2] Japan Council for Quality Health Care (JCQHC), http://www.med-safe.jp/contents/english/index.html

[3] International Classification for Patient Safety (ICPS), WHO, http://www.who.int/patientsafety/implementation/taxonomy/en/

[4] International Classification for Patient Safety (ICPS) Documents, WHO, http://www.who.int/patientsafety/implementation/taxonomy/icps_statement_of_purpose.pdf

[5] Manning C D, and Schutze H. Foundations of statistical natural language processing, The MIT Press, London. 2002

[6] MeCab: Yet Another Part-of-Speech and Morphological Analyzer, http://mecab.sourceforge.net/

[7] Matsuo Y, and Ishizuka M. Keyword Extraction from a Document using Word Co-occurrence Statistical Information. JSAI 17(3):213-227. 2002

[8] Salton G. Introduction to Modern Information Retrieval Mcgraw-Hill College. 1983

[9] Rasmussen E. Clustering Algorithms, Information Retrieval: Data Structures and Algorithms. William B Frakes and Ricardo Baeza-Yates. 1992

[10] Newman M E J. Fast algorithm for detecting community structure in networks. Physical Review E 69, 066133. 2004