

## **A study on key factors for improving medical service based on big data**

Chong Hyung Lee

Professor, Department of Hospital Management, Konyang University, 158,  
Gwanjeodong-ro, Seo-gu, Daejeon,35365, South Korea  
chlee@konyang.ac.kr

### **Abstract**

**Background/Objectives:** This study identifies the key factors that enhance medical service by analyzing keywords of 761 papers registered in the Research Information Sharing Service(RISS) until 2019 in South Korea.

**Methods/Statistical analysis:** Web crawling was performed based on Python to collect academic papers whose title or keywords include “medical service” among papers registered in RISS. Also, text mining, social network analysis, centrality analysis, and topic modeling were applied as big data analysis methods, and R 4.0.2, UCINET 6.7, and NetDraw 2.172 were used as software for analysis.

**Findings:** Among the 3,175 keywords obtained as a result of text mining, keywords such as “Medical service”, “Customer satisfaction”, “Medical service quality”, “Reuse intention”, “Healthcare service”, “Medical service value”, “Medical tourism”, “Emergency medical service”, “Medical service use”, “Loyalty”, “Reliability”, and “Telemedicine” were obtained at a high frequency of occurrence. A social network was derived based on 40 core keywords obtained from text mining, and there was a strong connection between “Customer satisfaction - Medical service quality”, “Medical service - Customer satisfaction”, “Customer satisfaction - Reuse intention”, “Medical value quality - Reuse intention”, “Customer satisfaction - Medical service value”. Five topics such as “Reuse and recommendation intention”, “Telemedicine & u-Healthcare”, “Medical service accessibility”, “Satisfaction management”, and “Medical tourism” were obtained as critical topics representing the obtained keywords. The topic “Reuse and recommendation intention” with the highest proportion among the five topics indicates that there is a high quality of the relationship between the consumer and service provider when the consumers are delighted and confident in the service provider’s service.

**Improvements/Applications:** Based on the big data analysis method, the key factors of medical service improvement and the relationship between factors were presented. It can be used as fundamental data in attempts to enhance medical services.

**Keywords:** Medical service, Web crawling, Text mining, Social network, Centrality analysis, Topic modeling.

## 1. Introduction

South Korea launched the national health insurance system in 1989. Since the demand for medical care has radically increased, many medical institutions have been established. The competition among medical institutions became fierce, and many medical institutions faced difficulties in managing their business as the investment and costs for human resources, medical equipment, facility and environment, and administration increased. To overcome the difficulties, medical institutions attempted to adopt various advanced medical equipment and corporate management strategies for providing superior service to the patients. The patients were treated as customers, quality of additional services, diagnosis, and treatment improved, and certifications were obtained from the domestic and overseas medical institution to gain customer trust and satisfaction.

In South Korea, medical institutions are divided among clinics, maternity clinics, and hospital-level medical institutions. Clinic-level medical institutions include primary care, dentist, and oriental medicine clinics. Also, hospital-level medical institutions include hospitals, dentist, oriental medicine, nursing hospitals, and general hospitals[1]. Medical institutions are divided into national, public, corporate, union, private, and military hospitals. Out of a total of 71,102 medical institutions, 63,681 are clinic-level institutions and 65,056 are private institutions[2].

Many kinds of research for improving medical service have been performed over the past 30 years. The researches are divided into healthcare, public health, emergency medical, and oriental medical services, and the content of the researches include policy establishment and assessment, medical costs, medical service quality, facility and environment, service accessibility, service convenience, customer satisfaction, job satisfaction by employees, attracting overseas patients, and information system development. Also, to identify trends in medical services research, studies were conducted to collect and analyze papers registered in the electronic database[3,4]. When analyzing 486 research papers for medical service marketing, it was shown that research on medical service quality, satisfaction and revisit, and reuse were actively performed[5]. Recently, research papers that applied the text mining of big data and social network analysis methods by utilizing web crawling have been published[6,7]. Web crawling is a method of systematically collecting data from the web, and text mining is a method to find helpful information by applying the natural language processing to unstructured data[7,8]. Additionally, social network analysis is conducted to identify

the relationships among the collected information[9]. Throughout web crawling, text mining, and social network analysis, it has been found that meeting the needs of patients visiting small and medium-size hospitals requires hospital facilities, safety, transportation, networks with advanced hospitals, and excellent medical expertise[6].

This study aims to find key factors for improving medical service by utilizing big data analysis such as web crawling, text mining, social network analysis, and topic modeling. For this purpose, the papers including 'medical service' in titles or keywords of papers were collected, and core keywords with a high frequency of appearance among keywords of the collected papers were obtained as key factors. In addition, it presents the connections and the strength of connections among core keywords, and derives topics representing the obtained keywords.

## 2. Research method

### 2.1. Data collection

Web crawling was performed based on Python's BeautifulSoup and Selenium libraries to collect papers whose title or keywords included "medical service" among domestic academic papers registered in RISS provided by the Korea Education and Research Information Service (KERIS). As a result, paper title, author name, publication year, keywords, and abstracts were collected from 1,060 papers.

### 2.2. Data refining and analysis

Among the collected 1,060 papers, 761 papers were selected as final research objects after excluding 239 papers duplicated and 47 papers about overseas medical service. As shown in Table 1, the earliest publication year among the selected papers was 1990, and 22 papers were published before 2000, 221 from 2000 to 2009, and 518 from 2010 to 2019, inferring that studies on medical service were actively conducted in recent years. When the keywords were both in English and Korean, only Korean keywords were included, and for those 279 papers whose keywords were in English, the words were translated into Korean. In addition, several keywords with the same meaning were unified into a single term. For instance, the words "Quality of medical service", "Excellence of medical service", and "Medical service quality" were used in the same sense, and they were refined to "Medical service quality".

**Table 1: Number of Selected Papers by the Publication Year**

Year	1990– 1994	1995– 1999	2000– 2004	2005– 2009	2010– 2014	2015– 2019	Total
Numbe	3	19	49	172	278	240	761

r							
---	--	--	--	--	--	--	--

To perform the study, text mining, social network analysis, centrality analysis, and topic modeling analysis were applied, and R 4.0.2, UCINET 6.7, and NetDraw 2.172 were used.

### 3. Results and Discussion

#### 3.1. Keyword frequency analysis

After refining the keywords, text mining was used to extract 1,710 types of keywords, and including those that were repeated, the total occurrence frequency turned out to be 3,175. Forty words had the occurrence frequency of seven or higher, and the most appeared keyword excluding “Medical service (220 times, 6.9%)” was “Customer satisfaction (194 times, 6.1%)”. In addition, there were keywords such as “Medical service quality”, “Reuse intention”, “Healthcare service”, “Medical service value”, “Medical tourism”, “Emergency medical service”, “Medical service use”, “Loyalty”, and “Reliability” in the order of occurrence frequency, and the result was shown in Table 2.

**Table 2: Occurrence Frequency by Core Keyword**

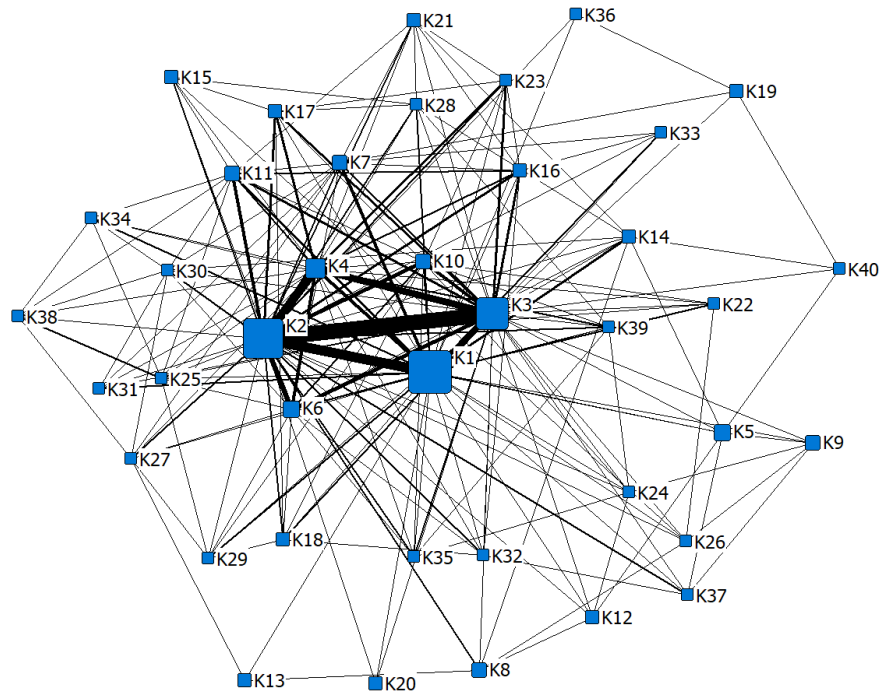
No.	Keyword	Freq uenc y	No.	Keyword	Freq uenc y
K1	Medical service	220	K2 1	Oriental medical services	13
K2	Customer satisfaction	194	K2 2	Nursing services	11
K3	Medical service quality	142	K2 3	Relationship quality	11
K4	Reuse intention	65	K2 4	Medical service accessibility	11
K5	Healthcare service	37	K2 5	Customer response	10
K6	Medical service value	35	K2 6	Elderly	10
K7	Medical tourism	29	K2 7	Medical service encounter	9
K8	Emergency medical	27	K2	Reuse	9

	service		8		
K9	Medical service use	25	K2 9	Job satisfaction	9
K1 0	Loyalty	24	K3 0	Behavioral intention	9
K1 1	Reliability	24	K3 1	National image	8
K1 2	Telemedicine	20	K3 2	Waiting time	8
K1 3	Service design	19	K3 3	Medical institution	8
K1 4	Medical consumer	19	K3 4	Medical service quality improvement	8
K1 5	Dental medical service	16	K3 5	Hospital reputation	7
K1 6	Hospital image	15	K3 6	Hospital medical service	7
K1 7	Recommendation intention	15	K3 7	Outpatient	7
K1 8	Customer orientation	14	K3 8	Medical service failure	7
K1 9	Medical service industry	14	K3 9	Human service	7
K2 0	Public health services	13	K4 0	Efficiency	7

### 3.2. Result of core keywords analysis via with the social network analysis method

Social network analysis was performed to visually understand the connection relationship between the 40 core keywords with a high frequency of appearance. In social networks, square nodes represent keywords, and links connect related nodes. To identify the connection, a co-occurrence matrix was created reflecting the number of times that two keywords appeared in the same paper, and the social network as in Figure 1 was drawn based on the NetDraw program. Nodes that are largely displayed in Figure 1 mean that they appear frequently. Furthermore, bold links

represent the frequent appearance of the two keywords in the same paper, inferring that the connection relationship is strong. Therefore, the nodes such as “Medical service”, “Customer satisfaction”, “Medical service quality”, “Reuse intention”, and “Healthcare service” appear much bigger. The connection relationship is strong in the order of “Customer satisfaction(K2) – Medical service quality(K3)”, “Medical service(K1) – Customer satisfaction(K2)”, “Customer satisfaction(K2) – Reuse intention(K4)”, “Medical service quality(K3) – Reuse intention(K4)”, and “Customer satisfaction(K2) – Medical service value(K6)”.



**Figure 1. Social Network Based on 40 Core Keywords**

### 3.3. Result of centrality analysis

For identifying the significance of core keywords in the social network, centrality analysis was performed, and among the methods, degree centrality and closeness centrality were used. Degree centrality is a method that identifies the number of links between a certain node and other nodes and evaluates a higher number of links as a higher degree of centrality. Looking at the centrality analysis result as shown in Table 3, the degree centrality was high in the order of “Customer satisfaction”, “Medical service quality”, “Medical service”, “Reuse intention”, “Medical service value”, and “Loyalty”. The ranks of keywords such as “Customer satisfaction”, “Medical service quality”, “Medical service value”, “Loyalty”, “Reliability”, “Hospital image”, “Recommendation intention”, “Relationship quality” was higher than the results obtained from the frequency analysis as shown in Table 2. Therefore, these keywords could be interpreted as more important than as considered in terms of occurrence frequency.

Closeness centrality is a method that evaluates the route from a certain node to all other

nodes and determines the node that has a shorter route as a more important node. Looking at the closeness centrality of the top six keywords, the rank was the same as the top six words obtained from the degree centrality except that the rank of “Medical service”, and “Medical service quality” was reversed. Looking at other keywords, the ranks of “Reliability”, “Medical consumer”, “Hospital image”, “Customer response”, “Behavioral intention”, and “Recommendation intention” were higher than those from the frequency analysis.

**Table 3: Result of Centrality Analysis**

No	Keyword	Degree	Keyword	Closenes s
1	Customer satisfaction	350	Customer satisfaction	0.886
2	Medical service quality	260	Medical service	0.867
3	Medical service	202	Medical service quality	0.867
4	Reuse Intention	160	Reuse intention	0.684
5	Medical service value	87	Medical service value	0.629
6	Loyalty	56	Loyalty	0.609
7	Reliability	56	Medical Tourism	0.6
8	Hospital image	45	Reliability	0.6
9	Recommendation intention	40	Medical Consumer	0.574
10	Medical Tourism	37	Hospital Image	0.574
11	Relationship quality	35	Customer Response	0.574
12	Hospital reputation	23	Behavioral intention	0.574
13	Human service	22	Recommendation intention	0.565
14	Customer Orientation	21	Reuse	0.565
15	Reuse	21	Human service	0.565
16	Medical consumer	20	Healthcare service	0.557
17	Behavioral intention	19	Customer Orientation	0.557
18	Oriental Medical services	17	Oriental Medical services	0.557
19	Dental Medical service	16	Relationship quality	0.557
20	Customer response	16	Medical service accessibility	0.557
21	Medical service Encounter	16	Medical service Encounter	0.557

### 3.4. Result of topic modeling analysis

In order to find out topics about medical service, a topic modeling analysis based on latent Dirichlet allocation(LDA) was performed on the entire keywords except for the word medical

service. An LDA-based topic modeling analysis is a method that finds out the keywords that represent each topic based on the topic distribution in each document and the keyword distribution in each topic[9-11]. The package of R, topicmodels, was used. The optimal number of topics was determined by repeatedly increasing the number of topics by one until the keywords in each topic explained the corresponding topic well. As a result, the number of topics was five. The ratio of each topic and the top ten keywords included in each topic were shown in Table 4. Among the total topics, the topic with the highest ratio was “Reuse and Recommendation intention”(30.33%), followed by “Telemedicine & u-Healthcare”(18.52%), “Medical service accessibility”(17.37%), “Satisfaction management”(17.26%), and “Medical tourism”(16.49%).

**Table 4: Topic Modeling Analysis Result**

No	Topics	Top 10 Keywords	Proportion
1	Medical service accessibility	Medical service use, Healthcare service, Customer satisfaction, Oriental medical services, Medical service accessibility, Emergency medical service, Telemedicine, Private health insurance, Elderly, Health inequality	17.37%
2	Telemedicine & u-Healthcare	Medical service quality, Healthcare service, Customer satisfaction, Public health service, Telemedicine, Medical consumer, Nursing services, Efficiency, u-Healthcare, Elderly	18.52%
3	Satisfaction management	Emergency medical service, Service design, Medical service quality, Customer satisfaction, Customer orientation, Job satisfaction, Organizational commitment, Business performance, Healthcare accreditation, Patient experience	17.26%
4	Reuse and recommendation intention	Customer satisfaction, Medical service quality, Reuse intention, Medical service value, Reliability, Loyalty, Medical tourism, Hospital reputation, Recommendation intention, Relationship quality	30.33%
5	Medical tourism	Customer satisfaction, Medical service quality, Medical service industry, Medical tourism,	16.49%



		National image, Dental medical service, Health screening, Hospital medical service, Diagnosis related group, Patient safety	
--	--	---	--

The topic “Medical service accessibility” indicates that it is important to enhance the accessibility of various medical services to address health inequality experienced by medically vulnerable people such as the elders who live alone, disabled, teenage family-heads, single-parent family, homeless patients, basic livelihood security recipients, low-income class, and migrant workers[12,13]. Also, the topic “Telemedicine & u-Healthcare” shows that the information of the patient’s health status is delivered to a medical institution online, and the patient can benefit from telemedicine services[14]. In addition, the topic “Satisfaction management” implies that factors such as “Healthcare accreditation”, “Service design”, “Internal employees”, “Organizational commitment” and “Job satisfaction” contribute to enhance the quality of medical service and increase customer satisfaction and business performance of medical institutions[15,16]. The topic “Reuse and recommendation intention” indicates that there is a high quality of the relationship between the consumer and service provider when the consumer is highly satisfied and confident in the service provider’s service. It also shows that the high medical service quality and relationship quality enhances loyalty as well as the reputation of the hospitals and has a positive impact on the “Reuse and recommendation intention”[17]. The topic “Medical tourism” combines medical service with tourism for patient treatment and indicates that medical service quality is an important factor for patients in deciding where to visit for medical tourism. It also affects the national image[18,19].

#### 4. Conclusions

In order to identify key factors for improving medical service, this study collected academic papers published in South Korea until 2019 and analyzed the keywords consisted of the word “Medical service” in a title or a keyword. Keywords from a total of 1,060 papers in RISS were collected using a Python-based web crawling, and the keywords of 761 papers that met the purpose for this study were selected. As a result of text mining analysis, the keywords such as “Medical service”, “Customer satisfaction”, “Medical service quality”, “Reuse intention”, “Healthcare services”, “Medical service value”, “Medical tourism”, and “Emergency medical service” were obtained with a high level of occurrence frequency. The connection relationships between the 40 core keywords were strong in the order of “Customer satisfaction - Medical service quality”, “Medical service - Customer satisfaction”, “Customer satisfaction - Reuse intention”, “Medical service quality - Reuse intention”, “Customer satisfaction - Medical service value”. The result of

centrality analysis showed that “Customer satisfaction”, “Medical service quality”, “Medical service”, “Reuse intention”, and “Medical service value” were identified as key factors. Also, five topics representing the obtained 3175 keywords were obtained: “Medical service accessibility”, “Telemedicine & u-Healthcare”, “Satisfaction management”, “Reuse and recommendation intention”, and “Medical tourism”. Among these topics, the topic “Reuse and Recommendation Intention” was considered the most important topic.

The limitation of this study was that among domestically published papers related to the medical service enhancement, those that did not include “medical service” in the title or keywords were excluded from the study objects. However, the core keywords for medical service enhancement are derived, and the connection relationships between the keywords are visually presented. In addition, the presentation of topics representing the keywords for medical services can be used as fundamental data in attempts to enhance medical services.

## 5. Acknowledgment

This paper was supported by the Konyang University Research Fund in 2020.

## 6. References

1. Korean Law Information Center [Internet]. c1997. Medical law; 2009 Jan 30 [cited 2021 Jan 10]. Available from: <https://law.go.kr/LSW/nwRvsLsInfoR.do?lsiSeq=91333>
2. National Health Insurance Service [Internet]. c2013. National health insurance statistical yearbook; 2019 Oct 1 [cited 2021 Jan 13]. Available from: <https://www.nhis.or.kr/nhis/together/wbhaec06300m01.do?mode=download&articleNo=132759&attachNo=60437>
3. Jang KS, Chung KH, Kim YH. A literature review of research on medical service design in Korea. *Journal of Korean Nursing Administration Academic Society*. 2018 Jan;24(1):85-96. DOI: <http://doi.org/10.11111/jkana.2018.24.1.85>.
4. Kim W, Kim JM, Shin J, Kim TH, Lee SG. Analysis of research trends in healthcare service marketing. *Korean Journal of Hospital Management*. 2019 Mar;24(1):21-35.
5. Kim M, Choi M, Youm Y. Semantic network analysis of online news and social media text related to comprehensive nursing care service. *Journal of Korean Academy of Nursing*. 2017 Dec;47(6):806-16. DOI: 10.4040/jkan.2017.47.6.806.
6. Hwang SW. Extracting of interest issues related to patient medical services for small and medium hospital by SNS big data text mining and social networking. *Korean Journal of Hospital Management*. 2018 Dec;23(4), 26-39. Available from: <https://www.koreascience.or.kr/article/JAKO201810760744047.pdf>

7. Mitchell R. Web scraping with Python. 2nd ed. California: Oreilly; 2015. p. 119.
8. Bali R, Sarkar D. R machine learning by example, 1st ed. Birmingham: Packt Publishing; 2016. p. 287.
9. Hanneman RA, Mark R. Introduction to social network methods. CA: University of California, Riverside. 2005. Available from: [http://faculty.ucr.edu/~hanneman/nettext/C18\\_Statistics.html](http://faculty.ucr.edu/~hanneman/nettext/C18_Statistics.html)
10. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. *Journal of Machine Learning Research*. 2003 Mar;3:993-1022.
11. Yoon JE, Suh CJ. Research trend analysis on smart healthcare by using topic modeling and ego network analysis. *Journal of Digital Contents Society*, 2018 May;19(5):981-93. DOI: 10.9728/dcs.2018.19.5.981.
12. Hur L, Lee YS. Activities to improve the accessibility to clinical social work for patients from vulnerable/disadvantaged/marginalized social groups. *Quality Improvement in Health Care*. 2013 Dec;19(2):80–8. DOI: 10.14371/QIH.2013.19.2.082.
13. Park YY, Park JH, Park YH, Lee KS. Assessment on the spatial accessibility of medical institutions providing national gastric cancer screening service using a geographic information system. *Korean Journal of Health Service Management*. 2019 Mar;13(1):15–30. DOI: 10.12811/kshsm.2019.13.1.015.
14. Lee J. The recent trends in telemedicine in the era of Covid-19 and policy recommendations for the balanced growth of healthcare Service Industry in Korea. *Journal of the Convergence on Culture Technology*. 2020 Nov;6(4):591–8. DOI: 10.17703/JCCT.2020.6.4.591.
15. Lee YH, Lim JD. Relationship between changes of patient safety & medical service quality and changes of management activity after medical institution accreditation : mental hospitals and geriatric hospitals. *Journal of Korea Contents Association*. 2015 Jan;15(1):286-299. DOI:10.5392/JKCA.2015.15.01.286.
16. Noh EM. Convergence study of servicescape and human services on the perceived values and customer satisfaction. *Journal of the Korea Convergence Society*. 2018 Jan;9(1):325–30. DOI:10.15207/JKCS.2018.9.1.325.
17. Crosby LA, Evans KR, Cowles D. Relationship quality in services selling : an interpersonal influence perspective. *Journal of Marketing*. 1990 Jul;54(3):68-81. DOI:10.1177/002224299005400306.
18. Zhang J, Lee HY. A study of factors influencing Chinese customer intention to select an international medical tourism destination. *Korea Journal of Business Administration*. 2016 Jun;28(6): 1703-23. UCI:G704-000789.2015.28.6.010.
19. Kim YJ, Kim J. Effects of expected medical service and country image on medical tourism

intention. International Business Review. 2018 Sep;22(3):187-214.  
DOI:10.21739/IBR.2018.09.22.3.187.