A Study on Change in Perception of Community Service and Demand Prediction based on Big Data

[1]Chun-Ok Jang

[1]*Assistant Professor., Department of Social Welfare, Honam University, Korea*
spring@honam.ac.kr

Abstract

The Community Social Service Investment project started as a state subsidy project in 2007 and has grown very rapidly in quantitative terms in a short period of time. It is a bottom-up project that discovers the welfare needs of people and plans and provides services suitable for them. The purpose of this study is to analyze using big data to determine the social response to local community service investment projects. For this, data was collected and analyzed by crawling with a specific keyword of community service investment project on Google and Naver sites. As for the analysis contents, monthly search volume, related keywords, monthly search volume, search rate by age, and gender search rate were conducted. As a result, 10 items were found as related keywords in Google, and 3 items were found in Naver. The overall results of Google and Naver sites were slightly different, but they increased and decreased at almost the same time. Therefore, it can be seen that the community service investment project continues to attract users' interest.

1. INTRODUCTION

In 2007, the Ministry of Health and Welfare implemented Community Social Service investment for the purpose of developing social services in consideration of local conditions and demands and creating high-quality jobs creative ideas of local communities. Community service started as a government subsidy project in 2007 and has grown very rapidly in terms of quantity within just a few years. Based on this growth, the local community service investment project was also converted from the initial individual subsidy system to the comprehensive subsidy system in 2012 [1].

The local community service investment project is a bottom-up project in which the local government discovers the welfare needs of the local community and plans and provides appropriate services, breaking away from the central government-led welfare service provision method [2-4]. This is to expand the autonomy of business operations and local financial management according to local conditions through the total amount of subsidies reflecting local demand. As local community service is guaranteed and activated in the future, in terms of current suppliers, various target groups based on local needs and business development, community service model development and spread, target selection criteria coordination, and standard manual

development and dissemination From the user's perspective, it is urgent to find a more active information provision method to enhance user choice, to develop new business to expand the target audience of various classes, and to expand user standards to form a self-sustaining market [5].

Therefore, this paper intends to confirm the degree of interest in community service investment projects in order to find a more active information provision method to enhance user choice, develop new businesses to expand users of various classes, and expand user standards to form a self-sustaining market. Basic data for future modeling development by crawling the Internet or SNS using big data to find out how much people are interested in the research method to find out the change in perception of the local community service investment project by year want to use it as.

2. RESEARCH CONTENTS AND METHODS

2.1 Big Data

The concept of big data was first presented by Cox and Ellsworth at the Proceeding of the 8th conference on Visualization (VIS) conference of ACM in 1997 [6]. Cox marked the start of the big data concept by raising the issue in his thesis 'Application-controlled demand paging for out-of-core visualization' that 'we will face a situation that can no longer be contained in data storage'.

Characteristics of big data include velocity, large volume of data, and variety. It uses data about almost every field in which we live, such as entertainment and daily life. In addition, in 2012, Gartner selected big data as the keyword of the year for IT-related 2013 by 9 out of 10 global market research institutes such as IDC and EMC. It was newly defined as 3VC [7]. The reason why big data is attracting attention is that it is expected to have a profound impact on our lives in line with the rapid increase in the size of the market for big data [8].

Recently, the scope of the definition of big data has been expanded, and huge data that includes not only the existing definition of large-capacity structured data, but also unstructured information that means information such as raw and unstandardized everyday language or conversations. This means a set of and should [9]. Although this unstructured, everyday data is not processed, it has a value that can be fully utilized in the era of big data. If daily information is used for analysis of consumer needs and response trends for marketing techniques, even atypical daily information that previously seemed ineffective will have sufficient value.

2.2 Big Data-Based Keyword Extraction

Extracting data from big data and providing keywords is the most basic and it can be thought of as the most basic method when a company establishes a strategy.

For example, as shown in Table 1 below, you can compare and check the differences in data by comparing and analyzing the content of a total of five journalists. Based on these results, it is possible to provide data to users who want to compare and analyze research trends that are important for each journal.

Table 1. Keywords by Rank

Journal	Keywords by Rank	Remarks
Journal-1	children, students, depression, family, stress, adolescents, anxiety, mothers, attachment, mother, expression, middle, male, female, women, patients, disability, disabilities, parents	-
Journal-2	students, mutism, depression, child, adolescents, dyslexia, family, junior, elderlies, neuroticism, women, parents, soldiers, disability, Vasovagal, Syncope, anxiety, adjustment, middle, self-expression, college, NCOs(Non-commissioned officers), girls, schizophrenics	-

Journal-3	children, mood, HSCT(Hematopoietic stem cell transplantation), diversity, trauma, support, anxiety, SPs(Standardized students, patients, family, PTSD, mothers, individuals, Veterans, child, adults, survivors	-
Journal-4	children, students, depression, patients, DV(domestic violence), girls, burnout, family, attachment, PDA(panic disorder with agoraphobia), adolescents, development, women, disorders, couples, GAD(generalized anxiety disorder), orphans, boys, ADHD, anxiety, psychosis, body, individual	-
Journal-5	BPD(borderline personality disorder), POST-TRAUMATIC, veterans, Parkinson, children, stress, trauma, patients, Schizophrenia, behaviours, colleagues, workforce, depression, psychosis, phobia	-

2.3 How to Extract Keywords

Methods for extracting keywords include Term Frequency, Term Frequency-Inverse Document Frequency (TF-IDF), Pointwise Mutual Information (PMI), and Jaccard Coefficient. It is important to use a suitable method for each situation [10].

■ **Term Frequency(TF).** It is a method to check the frequency of occurrence of a specific word in a specific document. The TF calculation method proceeds in the same way as the method of calculating the number of occurrences for each word in each document in DTM and Bag of Words (BoW), which are count-based word expression methods. It measures how often a given word (*term t*) appears in a document, and it is more likely to occur in long documents than in short documents. Therefore, to solve this problem, as in Equation (1), the frequency of occurrence of the word *t (term t)* in the document *d(i)* is divided by the total number of words in the document *d(i)* to standardize it.

$$\text{TF}(t,\ d) = (\text{frequency of occurrence of word } t \text{ in document } d) / (\text{total number of words in document } d) \quad \textbf{(1)}$$

This method extracts keywords using the frequency of words, and if a sentence such as "I am from Korea" is extracted, 'am' - 3, 'I' - 3, 'from' - 1, 'Korea' are the result values. It can be extracted in the form ' - 1.

■ **Term Frequency-Inverse Document Frequency(TF-IDF).** IDF (Inverse Document Frequency) measures the importance of a word *t* in a corpus. In addition, when calculating TF, all words are given the same importance (weight), but they frequently appear in documents, such as stop words such as articles (a, the) and be verbs (is, am, are), but are of little importance. There are also words that are not spoken. To solve this problem, in Equation (2), IDF lowers the importance (weight) for words that appear in common in multiple documents in the corpus, and on the contrary, the importance is increased.

$$\text{IDF}(t,\ D) = log(\text{Total number of documents in corpus} / \text{number of documents containing word } t) \quad \textbf{(2)}$$

TF-IDF score is multiplied by the above TF score and IDF score as in Equation (3).

$$\text{TF-IDF }(t,\ d,\ D) = \text{TF }(t,\ d) \text{ x IDF }(t,\ D) \quad \textbf{(3)}$$

For example, if there are documents *A* and *B*, it can be judged that the same word mentioned in *A* is more important, even if the same word in documents *A* and *B*, because a word that is frequently mentioned in document *A* can be an important word in document *A*.

■ **Point-wise Mutual Information(PMI).** The PMI method extracts keywords using the frequency of words, and if a sentence such as "I am from Korea" is extracted, 'am' - 3, 'I' - 3, 'from' - 1, it can be extracted in the

form 'Korea' - 1. For example, assuming that the probability that event *A* will occur is P(*A*), the probability that event *B* will occur is P(*B*), and PMI = P(*A*∩*B*) / (P(*A*)*P(*B*)) can express. The above statement can be seen as a method of calculating the probability that *A* and *B* will occur simultaneously among the probability that events *A* and *B* will occur. For example, in document 1, word *A*: 1,000, in document 2, word *B*: 1,500, in document 3, word *C*: 2,000, total number of documents: 10,000, the probability of *P* (document *A*) is = 1,000/10,000 = Calculate the probability of 0.1, *P*(Document *B*) = 1,500/10,000 = 0.15, Probability of *P*(Document *C*) = 2,000/10,000 = 0.2, and then find the probability of coming out of the document at the same time. The number of documents *A* and *B* is 300: P(*A*∩*B*) = 300/10,000 = 0.03, the number of documents *A* and *C* is 700: P(*A*∩*C*) = 700/10,000 = 0.07, this content if the PMI value is calculated based on it can be seen that the correlation and relevance of document *C* is higher.

■ **Jaccard Coefficient.** If there are two sets, the Jaccard Coefficient methodology is a method of extracting the result value based on the degree of overlap between the sets. The Jaccard coefficient can be expressed by the following equation (4).

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}.$$

(4)

For example, there are documents *A* and *B* to be compared.
Document *A*: I am a Jang who is a student.
Document *B*: I am from S. Korea and Jang. And there is a document *C* that you want to compare documents.
Document *C*: You are a *Jang* who is a student.
If we compare documents *A* and *B* based on document *C*, then Score(document *C*, document *A*) = 5 / (8+8-5) = 0.4545...
To explain how it is counted, 5: 5 overlapping words in document *A* and document *C*, (8+8-5) ==> 8 word count in document *A* + 8 word count in document *C* - count as 5 overlapping words becomes this
Same as above, Score(Document *C*, Document *B*) = 1 / (8+6-1) = 0.076...
In other words, it can be seen that document *C* and document *A* are more similar documents.

3. DATA ANALYSIS

Crawling refers to collecting and classifying data. It mainly means collecting, classifying, and storing web pages on the internet, and the main purpose of crawling is to classify where data is stored and where it is located. The main target of crawling is data that exists in various forms. It can be classified into structured, semi-structured, and unstructured data according to the data generation style, but it can be classified into data generated by companies and users according to the entity that produces the data [7].

3.1 Methods of Data Analysis

We crawl and utilize resources on the web to determine whether there is a change in awareness and interest in community service investment projects. In this paper, in order to determine the degree of interest in the "community service investment project", we analyze the Key Words generated for 12 months from October 2021 to September 2022 on Google.com and Naver.com sites.

As an analysis method, we check the total number of searches on Google and the number of web documents for the first time the word "community service investment project" appears and for 12 months. In addition, changes in awareness and interest are checked based on related keywords, changes in monthly search volume, and monthly search rate ratios.

3.2   Results of Data Analysis

The appearance date of the Keyword "Community Service Investment Project" on Google.com started in April 2009, and the total number of web documents was confirmed to be 10,200,000. From September 2019 to August 2022, the relevant keywords are local community service investment project 2022, local community service investment project case, community service investment project guidelines, community service investment project cyber education, community service investment project you can view forms, best practices for local community service investment projects, monitoring of local community service investment projects, application for local community service investment projects, status of local community service investment projects, and inspection of local community service investment projects. Here, the keyword with the highest monthly search volume was "Community Service Investment Project 2022", which showed a monthly search volume of 140, and the number of related search terms for "Community Service Investment Project" was confirmed 10 times per month.

If you look at the contents of Figure 1, it shows the number of Keyword searches per month.

Figure 1. Keyword searches per month

As shown in Figure 2 the prediction of monthly keyword searches up to April 2023 when interpolation was applied using the number of monthly keyword searches for 12 months from September 2021 to August 2022. If you look at the contents, you can see that the search volume increases at the beginning of the year when the "community service investment project" starts.
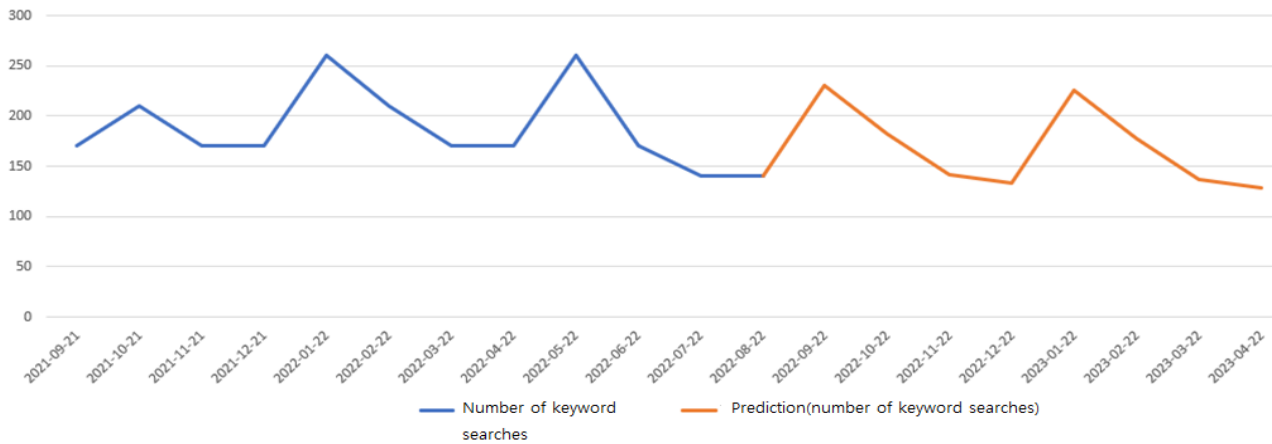

Figure 2. Changes predicted by April 2023 using the number of monthly keyword searches from September 2021 to August 2022

Contents by search result type are checked, and it can be seen in Figure 3 that the web documents are 70%, videos 10%, images 10%, and others 10%.
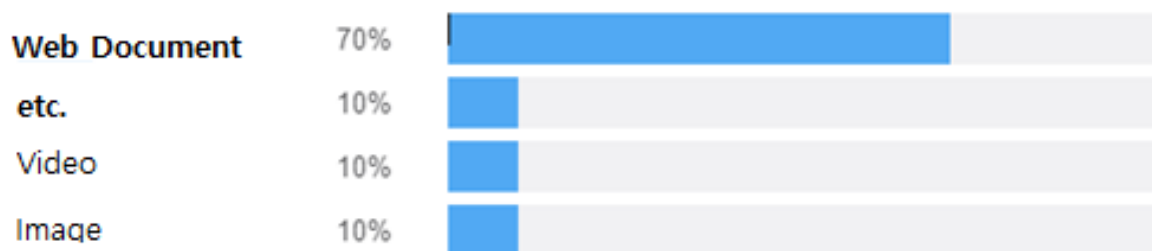

Figure 3. Distribution of content by search result type

Also, look at search domain distribution, socialservice.or.kr (social service e-voucher), korea.kr (Korea policy briefing), sjss.or.kr (Sejong City Community Service Support Group), sejong.go.kr (Sejong Special Self-Governing City), welfare.seoul.kr (Seoul City Community Service Support C), gbss.or.kr (Gyeongsangbuk-Do Community Service Support Group), and bucheon.go.kr (Bucheon City Hall). Here, the average monthly clicks of socialservice.or.kr (social service e-voucher) was the highest at 30, and korea.kr (Korea Policy Briefing)'s "2021 local community service investment project guide data shows the average monthly clicks." was the second highest with 20 cases. In addition, sjss.or.kr (Sejong City Community Service Support Group), sejong.go.kr (Sejong Special Self-Governing City), and welfare.seoul.kr (Sejong City Community Service Support Group) averaged 10 clicks per month.

The appearance date of the Keyword "Community Service Investment Project" in Naver.com started in January 2016, and the total monthly search volume was 840, 480 from PC and 360 from mobile. Related monthly content publications were 17,200 blogs, 6,540 cafes, and 23,900 views.

Table 2 shows the number of searches per month for three related keywords, and shows the cumulative blog publication volume for 12 months.

Table 2. Related keywords

Related keywords	Average monthly searches	Cumulative blog publication volume
2009 Community Service Investment Project Guide	**20cases**	605,000 cases
Community Service Investment Project (Busan)	**30cases**	360,000 cases
Application for community service investment business service	**50cases**	663,000 cases

Figure 4 shows the number of monthly keyword searches on the Naver site. In January 2022, the number of searches was 2,420, the largest of the year, and in July 2022, the number of searches was 1,260, the second most searches. As a result of confirming the cause, it was found that January was the period with the most interest as the user recruitment period, and July was the additional user recruitment period.
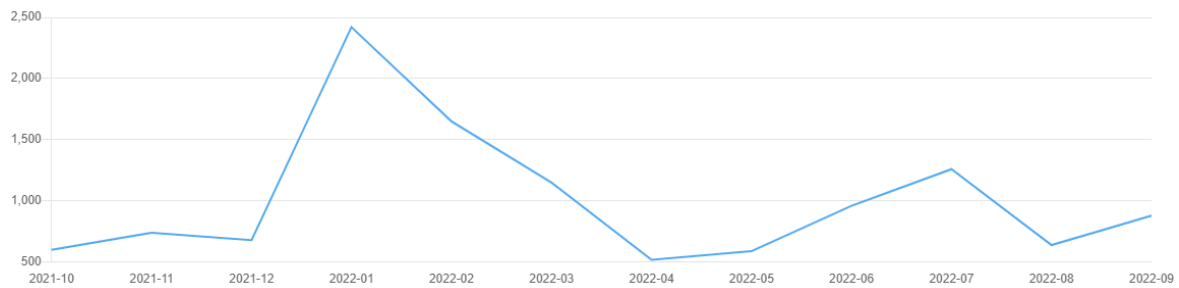
Figure 4. Keyword searches per month

As shown in Figure 5, when looking at the search rate by age, those in their 30s (34.5%) were the highest, followed by those in their 40s (31.3%), 20s (18.2%), 50s (15.2%), and teens (0.9%) and so on.

Looking at the next gender search rate, females were 82.3% and males were 17.7%, indicating that they were more interested in the future.
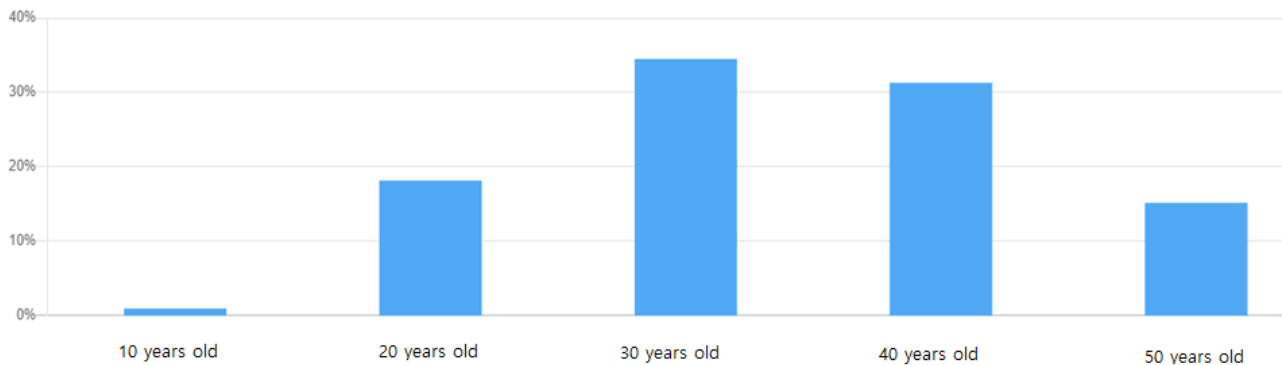

Figure 5. Search rate for keywords by age

4. CONCLUSION

Community Social Service Investment was started as a government subsidy project in 2007 and it has grown very rapidly in quantitative terms in a short period of time. In addition, in order to break away from the central government-led welfare service provision method and to check how much users are interested in local governments to discover the welfare needs of the local community and provide continuous services, Google and *Naver sites* use the keyword that was verified by crawling the community service investment project. As the analysis contents, monthly search volume, related keywords, monthly search volume, search ratio by age, and gender search ratio were conducted. The overall results of Google and *Naver sites* were slightly different, but they increased and decreased at almost the same time. "This is a Google.com crawl of monthly keywords for 12 months from September 2021 to August 2022. Looking at the predicted content by applying the interpolation method to the content distribution by type of data from here, around January 20th, the most It can be seen that there are a lot of searches, which is believed to be due to the start of the social service business. Looking at the number of monthly keyword searches on the Naver.com site, January 2022 was the most with 2,420 searches, and in July 2022, it was the second most searched with 1,260 searches. As a result of confirming the cause, it was confirmed that January was the period with the highest interest as the user recruitment period, and July was confirmed as the additional user recruitment period. In terms of the gender search rate, women showed more interest at 82.3% and men at 17.7%. This is considered to be because women

tend to participate in projects related to social services rather than men. Therefore, as a result of this thesis, it was found that the community service investment project is continuously receiving interest from users.

ACKNOWLEDGEMENT

REFERENCES

[1] H.S. Woon, "A New Social Change and Responding Improvement of Community Social Service," *Crisisonomy*, Vol. 10, No. 10, pp. 147-171, 2014.

[2] Y.H. Chon and M.H. Lee, "A Study of Local Governments' Evaluation of the Marketization of Social Service: Focusing on the Registration System of Community Social Service Investment," *Health and Social welfare review*, Vol. 38, No. 3, pp. 521–554, Sep. 2018. DOI: https://doi.org/10.15709/HSWR.2018.38.3.521

[3] Chun-Ok Jang, "A Study on the Factors Influencing the Satisfaction of Community Service: Focused on G metropolitan city," *The International Journal of Advanced Culture Technology,* Vol.7 No.1, pp. 1-7, 2019. DOI: https://doi.org/10.17703/IJACT.2019.7.1.1

[4] Chun-Ok Jang, "Effect of Social Service Quality on Service Satisfaction and Service Repurchase - Focusing on Social Service Investment Project," *The International Journal of Advanced Culture Technology,* Vol.9, No.4, pp. 1-7, 2021. DOI: ttps://doi.org/10.17703/I JACT.2021.9.4.213

[5] W.-S. Lee and S.-U. Hong, "A study on the Efficiency of Community Service Investment Projects Data Envelopment Analysis, DEA-Centered on 8 districts in Daegu," *Industrial Promotion Institute,* Vol. 7, Issue 2, pp. 1-13, 2022.

[6] M. Cox and D. Ellsworth, "Application-controlled demand paging for out-of-core visualization," in Proceedings of Visualization '97, Phoenix AZ, October, November, 1997. DOI:10.1109/VISUAL.1997.663888

[7] M. Beyer and D. Laney, The Importance of 'Big Data': A Definition, Connecticut 2012, CT: T. Gartner Chavez, C. O'Hara and V. Vaidya, Data driven, Seoul: Mc Graw Hill, 2019.

[8] H.-J. An, "A Business Performance Study of Data Quality Management for Big Data Adoption: Focused on Corporate Data Quality Management Process (Doctoral Dissertation)," Seoul: Kookmin University, 2015.

[9] T. Bell, Big Data: An opportunity in search of a metaphor, 2011.

[10] https://wikidocs.net/31698.