

Study on Improving Oriental Medicine Statistical System for Multidimensional Statistical Data

Sang-Jun Yea, Chul Kim, Jin-Hyun Kim, Hyun-Chul Jang, Sang-Kyun Kim, Mi-Young Song

Information Research Center, Korea Institute of Oriental Medicine, Yuseong-gu Daejeon, 305-811, Korea

ABSTRACT

Oriental medicine statistics are essential in research planning, research evaluation, and policy decision based on objective data. However, integrated administration of such statistics is not presently possible in the oriental medicine field, which has been slow in incorporating information communication technology. In an effort to address this problem, the Korea Institute of Oriental Medicine (KIOM) developed an oriental medicine statistical system in 2009, and the system has been offered in the traditional medicine information portal of OASIS. However, according to a 2010 survey targeting OASIS users, those surveys reported that needs for a system where various statistical data can be extracted via an interactive approach to multidimensional data. As a result of an analysis of the functions of the existing system, it was found that it is necessary to array and arithmetically analyze Stats Value, Drill Up & Drill Down, and Pivot. To this end, the existing DB schema should be redesigned.

Based on our analysis result, we redesigned the database into a structure that is applicable to the reverse pivot algorithm. We used J2EE/JSP and a Flex framework to design and develop an oriental medicine statistical system that can provide multidimensional statistical data. Considering that the improved oriental medicine statistical system is planned to be offered by OASIS of KIOM, utilization and value of oriental medicine statistical data are expected to be enhanced.

Keywords: Statistical System, OASIS, Multi-dimension, Oriental Medicine

1. INTRODUCTION

Thanks to intensive capital investment and breakthroughs in information communication technology, the importance of information communication as social infrastructure has increased to unprecedented levels. Collection, control, analysis, and provision of massive amounts of information, which were not possible in the past due to lack of information communication infrastructure, have become possible, and areas relevant to statistics have shown the greatest progress. With nationwide statistical surveys, a huge amount of money and manpower has been invested to research and analyze statistical data, and the data are used as objective grounds when important policies are decided [1]. However, the insufficient information communication infrastructure has presented a serious obstacle to carry out statistical surveys and it takes a long time to generate and analyze statistical data. Accordingly, utilization of statistical data has been significantly restricted [8].

Compared to other fields of study, application of information analysis and utilization has been very poor in the field of oriental medicine, which has been relatively slow in integrating information communication technology, to the extent that establishment of an oriental medicine research paper database was a starting point of actual informationization [2]. Oriental medicine statistics are essential to research planning and

evaluation and realistic policy decision making, and are necessary information for performing scientific research on par with that of western medicine. However, oriental medicine statistics were not controlled in an integrated manner till the mid-2000s, accordingly causing difficulties in their use, with the result that necessary statistics have had to be extracted from various yearbooks, whitepapers, and open sources issued by the government [8].

Two tasks were conducted in order to resolve this problem. First, in 2009, the Association of Korean Oriental Medicine (AKOM), Pusan National University Graduate School of Korean Medicine, and KIOM formed a Committee on Issuing a Korean Oriental Medicine Yearbook. And they gathered oriental medicine statistics, which were dispersed across all fields of study, to issue a yearbook of Korean oriental medicine covering five areas, including administration, research, and service. In addition, since developing a system that enables a user to search oriental medicine statistics in an integrated manner and check data in 2009, KIOM has offered the system via OASIS. Currently, the oriental medicine statistical system provides 124 types of oriental medicine statistics belonging to 43 categories [3].

According to the survey targeting OASIS users in 2010, the users requested improved functions of the oriental medicine statistical system [3],[4]. Specifically, they said that, although they want to extract various meaningful information by utilizing statistical data, it is not possible with the present system. They mentioned that the system needs the function of

* Corresponding author: E-mail : tomita@kiom.re.kr

Manuscript received Jul 27, 2011 ; accepted Sep.02, 2011

OLAP (Online Analytical Processing), which would allow them to extract various statistical data through an interactive approach to multidimensional statistical data. This study was intended to develop a system that is capable of providing multidimensional statistical data in an interactive manner by upgrading the existing oriental medicine statistical system.

2. RELATED STUDIES

In a study focused on how to improve accessing the oriental medicine statistics, Han et al. (2007) emphasized that the oriental medicine field faces a clear shortage of basic statistical data, as the field has only recently been recognized as an independent industrial area. He additionally stressed the need for generating and offering comprehensively statistics that can properly reflect the status of the oriental medicine field [5].

In a study on building a Web-based OLAP system for educational statistical information, Lee et al. (2007) explained that in order to meet customer demand for utilizing statistics directly on an education and human resource system via the Internet, a Data Warehouse (DW) should be introduced to establish the Web OLAP system. This would make a significant contribution to shortening statistics-related working hours and increasing utilization [6].

In a study on realizing a Pivot table-generation algorithm for analyzing OLAP on a relational database, Shin et al. (2007) provided a Pivot algorithm that transforms a vertical table structure expressed by the general relational database expresses into a multidimensional horizontal table structure available on OLAP and realized a system to verify the validity of the algorithm [7].

In a study on designing and realizing a statistical system for Chinese medicine, Yea et al. (2009) drew common items by analyzing the main statistical contents of oriental medicine, designed a database and realized the system by using a summary table STDP. However, the study had a limitation in realizing a system where a user can extract statistics in an interactive manner and utilize them in a various manner [8].

3. ANALYSIS ON THE EXISTING SYSTEM

The oriental medicine statistical system is offered by OASIS, KIOM's traditional medicine information portal [3]. As seen in [Figure1], a user can find statistics that he or she wants by searching the name of the statistics, keywords, descriptions, and directory method-using browsing. Upon clicking on a statistic name, more detailed statistical data can be available on the Stats Viewer and, as seen in [Figure1], some statistical items can be hidden through item filtering. Configuration of the OLAP system's functions varies depending on the purpose of usage, but the functions generally offered include Pivot & Filtering of statistical items, Drill Up & Drill Down of statistical items, change in the statistical data array, and arithmetic analysis of the statistical data [6]. In the oriental medicine statistical system, only Stats Item Filtering, from among these OLAP functions, is realized. Furthermore as the graph on the top of the Stats Viewer has a synchronized

relationship with the table, any change in the Filtering function is reflected in the graph.



Fig. 1. The statistical searching system of OASIS

The oriental medicine statistical system applied Summary Table Dynamic Production (STDP) to design the database schema in order to save various types of statistics [8]. However, the Summary Table STDP algorithm, which stores statistical data by using the vertical table structure, has a limit in configuring statistical items, as seen in [Figure2]. In other words, many statistical items (such as item1, item2) can exist in a row of the table but only one item (such as item A) should exist in a column. Therefore, in order for the oriental medicine statistical system to offer statistics, the table structure of statistical data should be modified to satisfy the limit. These limitations lead to disadvantages such as loss of original statistics information and impossibility of offering various statistics.

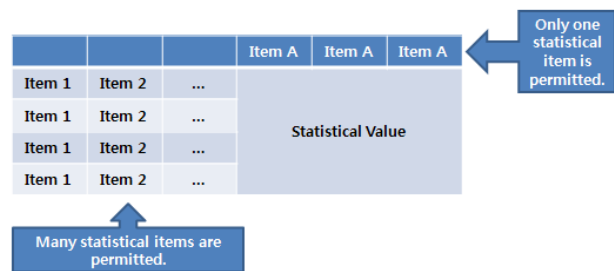


Fig. 2. The data structure limitation of existing system

4. SYSTEM REALIZATION

This study is designed to solve functional problems of the oriental medicine statistical system available on OASIS and design problems of the database schema, thus realizing an oriental medicine Stats Viewer where multidimensional statistical data is approached in an interactive manner and then extracted and utilized. To this end, this study is conducted as described below.

4.1 System Design

For storing the multidimensional oriental medicine statistics, as seen in [Table 1], the database schema is composed of CSTATISTICS saving Meta Data of oriental medicine Stats, CDIMENSION saving information on Stats Items, STATISTICS table, CFACT saving information on Stats Value, DIMENSION table, FACT table and DATA (n) table saving Stats Data. In the database schema diagram, idx columns on each table are connected by foreign keys, as seen in [Figure3], and in DIMENSION and FACT tables, rowcol and seq column, display locations of each Stats Item and each Stats Value of the original statistics. As for the DATA (n) table, n refers to the idx value of the STATISTICS table and the table is composed of columns with Stats Items (number of m) and Stats Values (number of l).

Table 1. DB schema explanation

Table	Description
CSTATISTICS	Classification information for the oriental medicinal statistics
STATISTICS	Metadata for the oriental medicinal statistics
CDIMENSION	Classification information for the statistical item
DIMENSION	Statistical item composition information for the individual oriental statistics
CFACT	Classification information for the statistical value
FACT	Statistical value composition information for the individual oriental statistics
DATA(n)	Analyzed oriental statistical data



Fig. 3. DB schema diagram

The oriental medicine statistical system is designed with a Web Application structure of J2EE/JSP Framework and, as seen in [Figure4]. Its functional modules of a Statistics Analysis Engine, OLAP Engine, Admin Viewer, and User Viewer. In the Statistics Analysis Engine, Excel file statistics are generated into a DATA (n) table by using Stats Items and

Stats Values designated by an administrator. During this process, the reverse pivot algorithm is applied from the horizontal table to the vertical table, consequently analyzing statistical data and generating a table. The OLAP Engine consists of sub-modules including the Pivot Module, Graph Module, Drill Module, Filter Module, Sort Module, Math Module, and Export Module. It allows users can extract and utilize multidimensional statistics in an interactive manner by using a *.csv file generated by using the DATA(n) table. Once a user reconfigures the multidimensional statistics in an interactive manner, Pivot, Drill, and Filter Module relocate Stats Items and calculate Stats Values internally in the *.csv files and display them in various table types. Math Module calculates partial and whole aggregates, frequency, and average of statistics and offers them to a user. Export Engine supplies a function of saving the user-configured statistics in Excel file.

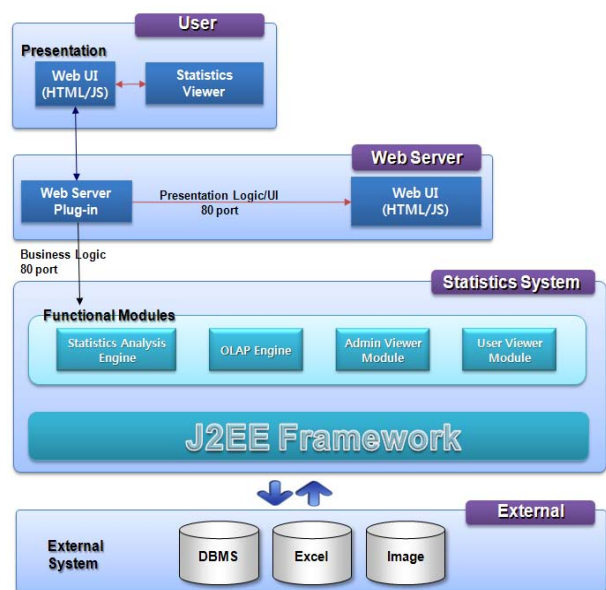


Fig. 4. Oriental statistical system diagram

4.2 System Development

The oriental medicine statistical system is developed in Windows XP by using Oracle 11g, Eclipse v3.3, Flex Builder v3.0, and Tomcat v5.5. The system is made up of an administrator function to control Stats Classification, Item, and Value and register the oriental medicine statistics and a user function to configure the multidimensional statistics in various ways and support utilization of statistical data.

First, the administrator viewer has a function to administer Stats Classification & Stats Item/Value on the left side and a function to register and administer Meta Data of the oriental medicine statistics on the right side, as seen in [Figure5]. The Stats Classification that an administrator administers is shared with a user, and it is possible to register statistical information within the limit of the Stats Item and Stats Value an administrator registers. When an administrator uploads a Stats Excel file and allocates Stats Items and Stats Values, the Statistics Analysis Engine analyzes the statistics and the OLAP Engine generates available data.

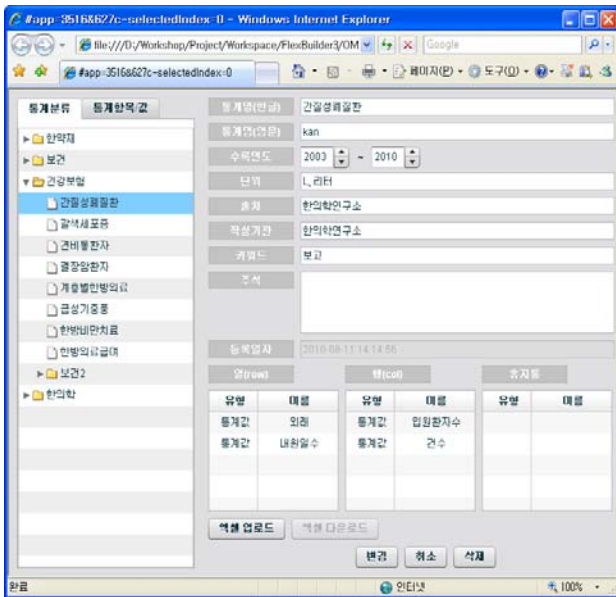


Fig. 5. Administrator viewer screen

As for User Viewer, there is a menu on the top, as seen in [Figure6]. The configuration at the bottom can be changed into table, graph, and table + graph. Stats Item's Pivot can be changed into the configuration that a user wants, based on the drag and drop method. Item-Filtering and Drill Up/Down are possible by clicking the Stats Items. If it is not possible to check the graph at a glance due to vast statistical data, the zoom-in function can be used to modify the size. In the drop-down box located at the bottom, Stats Values can be changed into sum, frequency, and average, thereby making it possible to check statistical data.

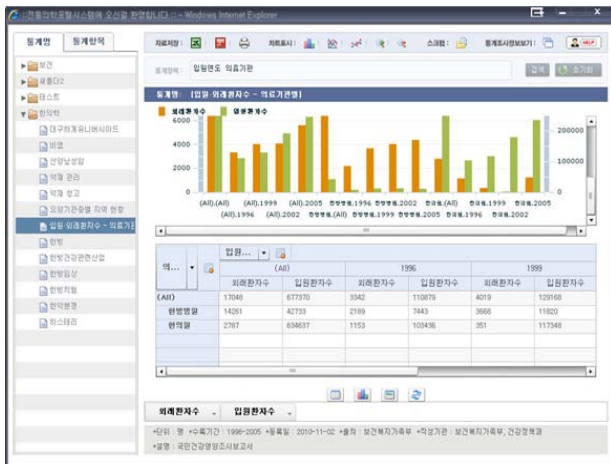


Fig. 6. User viewer screen

5. SYSTEM EVALUATION

In a study on suggestion and verification of evaluation factors to select OLAP tool, Kim et al. (2001) drew 9 categorized 34 items to evaluate the OLAP tools. From it we selected 12 factors in 3 categories, which are related OLAP user function. We then evaluated these factors of the existing

system and the improved system. We get the evaluation results, as seen in [Table2]. From the evaluation results, we knew that searching, filtering and qualification functions which is part of 'client tools' category have been realized in the existing system, but all functions of 'client tools' category are realized in the improved system. Also we knew that some functions of 'multidimensional view' and 'aggregation' categories are realized in the improved system, so our study aim which develops a system that is capable of providing multidimensional statistical data in an interactive manner by upgrading the existing oriental medicine statistical system are satisfied. Because 4 factors in 'multidimensional view' category means the various ways to facilitate the interactive manner of high dimensional statistics, the improved system is much better than the existing system in user interactivities. Especially the first factor (Drill Up/Down, Drill Access, Pivot) is the core of user interactivity, it provides user to handle statistical data with as one likes. Such as user can change the position and visibility of dimensions and fold or unfold of factors.

Table 2. Evaluation results

Category	Factors	Existing System	Improved System
Aggregation	Aggregation Function	x	o
	Aggregation Management	x	x
Multidimensional View	Drill Up/Down, Drill Access, Pivot	x	o
	Add Layer	x	x
	Add Dimension	x	o
	Extensibility	o	o
Client Tools	Alert Message	x	x
	Search	o	o
	Filter	o	o
	Qualification	o	o
	Sorting	x	o
	Retrieve	x	o

6. CONCLUSION

The oriental medicine community is limited to extracting statistical data from various yearbooks, white papers, and open sources issued by the government due to a lack of infrastructure to provide oriental medicine statistics in an integrated manner. In order to address this limitation, the oriental medicine statistical system currently available on OASIS, KIOM's traditional medicine information portal established in 2009, provides 124 statistical data belonging to 43 categories. Notwithstanding, the system is incomplete in terms of the functions by which a user can utilize statistical data in various manners and as a result, there is demand for improving the system's functions. Accordingly, this study is designed to apply an OLAP function by which multidimensional statistics can be utilized to analyze problems of the oriental medicine statistical system, from the perspective of functionality and a database.

To this end, this study redesigned the database schema and developed administrator functions and user functions. With respect to the existing system developed HTML-based functions, there is a fundamental limit that a user cannot use the functions in an interactive manner. However, this study improved the system by changing the HTML-based functions into FLEX-based functions, thus making it possible for an administrator and user to enjoy RIA's (Rich Internet Application) advantages such as rapid response and rich functions. As DB schema designed via the application of STDP is not applicable to the OLAP Engine, the DB schema was redesigned so that the reverse pivot algorithm, which can generate a vertical table, is available for statistical data. The improved oriental medicine statistical system realizes the OLAP functions, including stats item's Pivot, Filtering, Drill Up/Down, and graph creation, and further allows various methods of utilization and analysis of multidimensional statistics. Consequently, it is expected that the value of oriental medicine statistical data built on OASIS will increase.

From this study, we have improved the existing oriental medicine statistical system to provide more ease. To satisfy the aim, we adopted OLAP functions to handle high dimensional data in statistics. We anticipate that the researcher in oriental medicine save more times to analyze the statistics or trends in the fields to use improved system than to use existing one.

In the future, if oriental medicine statistical data is accumulated to the extent that the limit of a relational DB is exceeded, a DW will be constructed. Additionally, if the functions of the oriental medicine statistical viewer developed by this study are improved to include an extensive Decision Support System (DSS), it will be possible to generate objective data required for deciding on oriental medicine-related policy and to offer basic data for conducting scientific research on par with that of western medicine.

REFERENCES

- [1] <http://www.census.go.kr/hcensus/index.jsp>.
- [2] C. Kim, M.Y. Song, S.K. Kim, S.J. Yea, H.C. Jang, D.M. Um, "A Study of Information Strategy Plan Through Korean Information Center", KIOM Journal, vol. 14, no. 3, 2008, pp. 127-135.
- [3] <http://oasis.kiom.re.kr>.
- [4] S.J. Yea, H.Y. Kim, C. Kim, J.H. Kim, S.K. Kim, H.C. Jang, M.Y. Song, "A Study on the User Needs Analysis to Improve OASIS", KIOM Journal, vol. 16, no. 2, 2009, pp. 109-118.
- [5] C.Y. Han, D.Y. Kim, Y.J. Oh, H.S. Moon, "A Study on the Development of Statistics in Korean Medicine", KIOM Journal, vol. 13, no. 1, 2007, pp. 125-128.
- [6] K.J. Lee, J.M. Park, S.K. Kang, S.C. Hwang, "Construction of Web based OLAP system for Education Statistics Information", Proc. KCC2007, 2007, pp. 110-113.
- [7] S.H. Shin, J.H. Kim, Y.S. Moon, S.W. Kim, "Implementation of Constructing Pivot Tables for OLAP Analysis from Relational Database", Proc. KIISE2007, 2007, pp. 164-169.
- [8] S.J. Yea, C. Kim, H.C. Jang, S.K. Kim, J.H. Kim, M.Y. Song, "Design and Implementation of Statistics System for Traditional Korean Medicine", KCC Journal, vol. 9, no. 12, 2009, pp. 332-338.
- [9] K.W. Kim, Y.M. Seo, "The Suggestion and Verification of Evaluation Factors to Select OLAP Tool", The Journal of Management Study, vol. 30, no. 3, 2001, pp. 929-956.



Yea Sang-Jun

He received the B.S., M.S in computer science from KAIST, Korea in 2002, 2004 respectively. From 2008, he has been with the Information Research Center, Korea Institute of Oriental Medicine. His main research interests include computer vision and oriental

medicine contents.



Kim Chul

He received the B.S., M.S in industrial engineering from KAIST, Korea in 1998, 2000 respectively and also received Ph.D. in oriental medicine informatics from Wonkwang university, Korea in 2009. From 2006, he has been with the Information Research Center, Korea Institute of Oriental Medicine. His main research interests include RFID/USN and oriental medicine informatics.



Jang Hyun-Chul

He received the B.S., M.S in computer engineering from Chungnam national university, Korea in 1996, 1999 respectively and also received Ph.D. in computer engineering from Chungnam national university, Korea in 2007. From 2008, he has been with the Information

Research Center, Korea Institute of Oriental Medicine. His main research interests include bioinformatics and semantic web.



Kim Sang-Kyun

He received the B.S., M.S in computer engineering from Chungnam national university, Korea in 1999, 2001 respectively and also received Ph.D. in computer engineering from Chungnam national university, Korea in 2008. From 2007, he has been with the Information

Research Center, Korea Institute of Oriental Medicine. His main research interests include semantic web and database.

**Kim Jin-Hyun**

She received the B.S. in mathematical education and oriental medicine from Ehwa woman's university and Daegu haany university, Korea in 2001, 2009 respectively. From 2009, she has been with the Information Research Center, Korea Institute of Oriental Medicine.

Her main research interests include oriental medicine contents and terminological standards.

**Song Mi-Young**

She received the B.S in chemistry from Sookmyung woman's university, Korea in 1984 and also received M.S, Ph.D. in chemistry from KAIST, Korea in 1981, 1991 respectively. From 2004, she has been with the Information Research Center, Korea Institute of Oriental

Medicine. Her main research interests include oriental medicine contents and bioinformatics standards.