An Open Data Warehouse Construction Approach for Social Information Databases

Hideki NISHIMOTO

Key Words : Data Grid, Social Database, Data Management, Data Warehouse, Statistical data, Optimization, Government data, Management System

Abstract

This paper presents an application of the online analytical processing (OLAP) technology to the construction of a statistical database for social information.

The paper also discusses investigations into an effective method to realize such an application. The purposes of this research are to provide an open-style OLAP system that can assure consistency with the resource data and to increase the flexibility of data processing with widely used database management technology and optimization algorithms of data grid systems. Attaining these goals would promote the distribution and application of public data grid systems for social information.

INTRODUCTION

There is increased demand for dynamic analysis applications that approach data from various perspectives by using discovery methods; this demand is due to the greater ease of using the statistical data that is available on networks. On the other hand, there is still a requirement for advanced security functions that, for example, prevent breaches in privacy protection or the intrusion of contaminated data. Therefore, the use of the conventional database management technology is unavoidable.

This paper proposes a mechanism that uses multi-dimensional processing to transform data schema efficiently while preserving statistical information by using the relational database management system (RDBMS).

Social information databases contain such public information as economic indicators, financial results, markets conditions, corporate information, government white papers, laws and judicial precedents, and various other kinds of statistical data. Furthermore, they need to collect those data from the numerous sources of each country and each organization.

Recently, many social statistical data in the public organizations of each country are

being actively displayed on World Wide Web pages thanks to the expansion of the Internet application (Figure 1). Japan and Australia are now linked from the level of each government's national offices to cities, towns and villages, and the latest information can be easily accessed any time. Also, spread sheets can be embedded in a page, so users are able to process data instantly on a desktop after downloading.

We are now able to use the statistical data that is distributed on a network easily. Therefore, there is increased demand for dynamic analysis applications that approach data from various perspectives by using discovery methods, although so far work flow has been analyzed and processed by batch processing and exchanged by such media as magnetic tape. On the other hand, there is still demand for advanced security functions to prevent breaches in privacy protection or the intrusion of contaminated data. To cope with this situation, in this research we use the conventional database system to manage social information data so that we can assure the consistency and security of information. Also, we are going to consider the system base that in each application conducts schema transformation freely and maintains a form for this in individual sites.

In business data application fields, it has become main stream to arrange forms from the user's viewpoint as a warehouse of data in the site. The forms are then analyzed by convenient programs that handle the results in a standard language such as direct SQL after the various kinds of information have been stored in a conventional database system. This method is called OLAP (Online Analytical Processing) in the data warehouse.

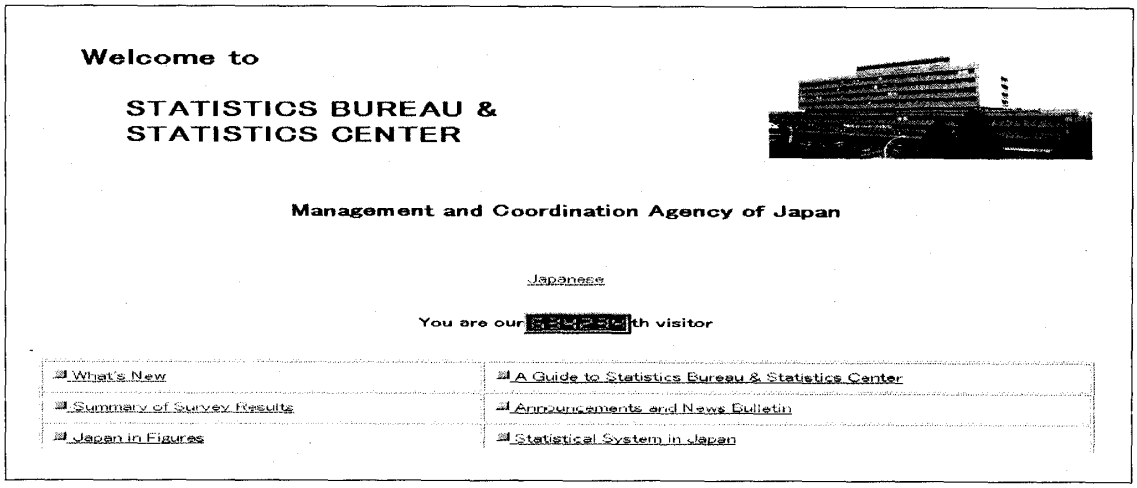
Social information data is the type most appropriate to this management style because it has specific characteristics, it is generally voluminous, the timing of an update is in a bundle, and it gradually exercises an influence over the application.

This paper investigates how to apply the OLAP technology to the construction of a statistical social database. The purposes of this research are to provide an open-style OLAP system that can assure consistency with the resource data and to increase the flexibility of data processing with widely used database management technology and optimization algorithms. Attaining these goals would promote the distribution and application of public social information (Figure 2).

Two characteristics of statistical social data are that the timing of an update is in a bundle and that the data gradually exercises an influence over the application. The period from one update to the next update can be relatively long, such as one year or one month. Also, the update work generally involves more addition processing of a bundle of new data in chronological order than change processing of the existing data. Furthermore, update processes are not so urgent because resource data renewal speed is gradual and the

Figure 1. Example of statistical social information homepage

(a) Statistics Bureau & Statistics Center, Management and Coordination Agency of Japan, URL: http://www.stat.go.jp/



(b) Australian Bureau of Statistics World Wide Web Information Service, URL: http://www.abs.gov.au/

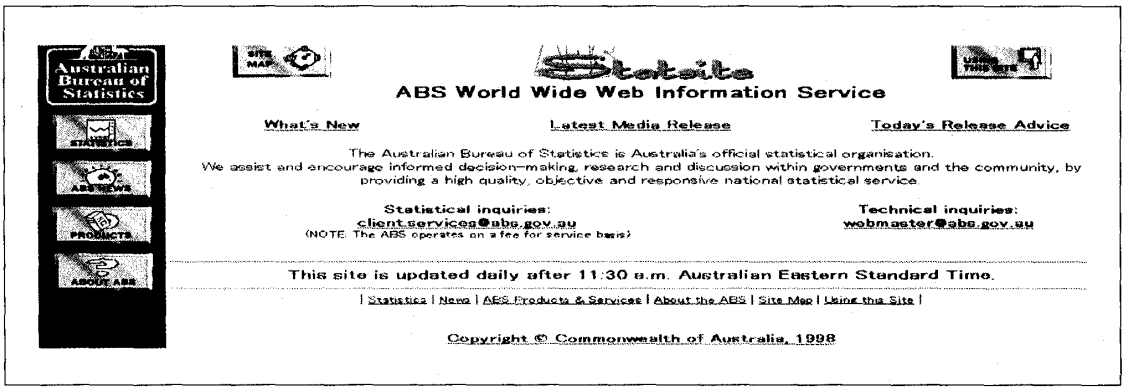
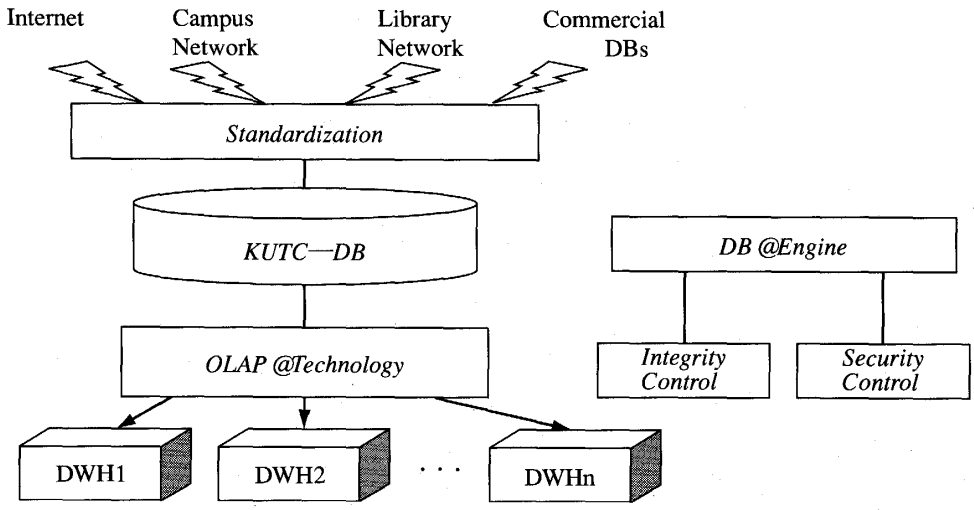


Figure 2. Open-style OLAP System



influence on the application is not such an important problem.

In the implementation of this system, it is not important to process a rapid update transmission in real time as with a nucleus data processing system. Therefore, it is sufficient that it supplements data with a loose renewal surveillance function after once acquiring resources, even when the system adopts an open-style.

Through this research, an open-style OLAP system has been devised that is equipped with a function to maintain consistency with resource data for a period of time even when the data is acquired from other sites.

STATISTICAL SOCIAL DATA AND OLAP

There are fundamentally many ways in which the database schema of statistical social information is given in two dimension tables. Table 1 is a part of the census results of Japan and Australia. Resource I is "B05 BIRTHPLACE (COUNTRIES) BY SEX" and also "B08 LANGUAGE SPOKEN AT HOME BY SEX" from Census of Population and Housing for Australia. Resource II is extracted from "Total average time (in weeks) in

Table 1. Example of statistical social data

Resource I. Census of Population and Housing for Australia (Extracting and Abbreviating)

http://www.statistics.gov.au/websitedbs/D3110124.NSF/

BIRTH

Birthplace Australia	Male 6527013	Female 6700763	Persons 13227776
Canada	12122	13010	25132
Chile	11509	12311	23820
China	53429	57580	111009
Croatia	24990	21991	46981
Egypt	17440	16719	34159
Fiji	17331	19771	37102
Germany	53805	56526	110331
Greece	64417	62103	126520
Hong Kong	33414	35016	68430
Hungary	13590	11673	25263
India	39677	37874	77551
Indonesia	21926	22249	44175
Ireland	26620	24849	51469
Italy	126452	111794	238246

LANG

Language English	Male 6657753	Female 6840341	Persons 6657753
Arabic	83698	78268	83698
Indigenous	21824	22368	21824
Cantonese	91558	98546	91558
Mandarin	42710	44610	42710
Croatian	33710	32992	33710
French	18152	19906	18152
German	46312	50339	46312
Greek	130329	128690	130329
Hungarian	12340	13506	12340
Indonesian	13108	12817	13108

Resource II. "Time use and leisure activities" of the social life basis survey : Whole time (Extracting and Abbreviating) http://www.stat.go.jp/05e41.htm

LIFE

Pref	Population (1000 people) over 10 years old	Sleep	Personal hygiene	Meal	Commuting	Job
Hokkaido	5,005	7.55	1.06	1.34	0.24	3.54
Aomori	1,296	8.03	1.08	1.39	0.24	4.11
:	:	:	:	:	:	:
Ibaraki	2,624	7.54	1.10	1.41	0.32	3.59
Tochigi	1,758	7.51	1.09	1.39	0.29	4.02
Gunma	1,775	7.51	1.11	1.40	0.28	4.03
Saitama	6,053	7.42	1.08	1.42	0.46	3.51
Chiba	5,200	7.39	1.12	1.40	0.44	3.42
Tokyo	10,674	7.44	1.10	1.43	0.38	3.51
Kanagawa	7,410	7.38	1.10	1.40	0.46	3.45
Niigata	2,207	7.56	1.09	1.33	0.26	4.08
:	:	:	:	:	:	:
Nagano	1,950	7.52	1.07	1.45	0.27	4.23
Gifu	1,867	7.47	1.07	1.35	0.30	4.03
Shizuoka	3,316	7.45	1.07	1.36	0.29	4.04
Aichi	6,100	7.46	1.05	1.35	0.33	4.06
Mie	1,633	7.52	1.08	1.36	0.31	3.52
Shiga	1,138	7.46	1.10	1.34	0.34	4.01
Kyoto	2,347	7.49	1.09	1.39	0.33	3.50
Osaka	7,833	7.39	1.08	1.40	0.40	3.46
Hyogo	4,792	7.42	1.08	1.40	0.36	3.42
Nara	1,278	7.43	1.03	1.41	0.41	3.37
:	:	:	:	:	:	:
Kochi	714	7.59	1.12	1.38	0.25	3.58
Fukuoka	4,340	7.44	1.09	1.37	0.31	3.43
Saga	766	7.49	1.07	1.35	0.24	4.05
Nagasaki	1,328	7.49	1.08	1.37	0.29	3.46
Kumamoto	1,616	7.52	1.07	1.41	0.25	3.52
Oita	1,078	7.45	1.07	1.38	0.26	3.56
Miyazaki	1,022	7.52	1.09	1.41	0.23	3.51
Kagoshima	1,552	7.53	1.13	1.43	0.22	3.48
Okinawa	1,077	7.55	1.10	1.38	0.24	3.34

Resource III. "Time use and leisure activities" of the social life basis survey : Spare Time (Extracting and Abbreviating) http://www.stat.go.jp/05e42.htm

SPORTS

Age	Baseball	Basketball	Soccer	Rugby	Judo	Karate /aikido	Sumo
10～14	40.2	48.7	40.7	0.9	3.6	2.0	2.6
15～19	25.9	37.0	24.6	1.8	7.1	1.4	0.8
20～24	22.2	12.1	11.1	0.7	0.8	1.2	0.2
25～29	18.9	5.2	6.7	0.4	0.4	0.7	0.2
30～34	16.7	3.1	5.7	0.3	0.3	0.4	0.3
35～39	18.6	3.5	6.4	0.2	0.4	0.4	0.4
40～44	15.7	2.5	4.4	0.1	0.3	0.4	0.3
45～49	10.2	0.9	1.6	0.0	0.2	0.3	0.2
50～54	6.0	0.4	0.5	0.1	0.2	0.2	0.1
55～59	4.4	0.2	0.2	0.0	0.2	0.2	0.0
60～64	2.2	0.0	0.1	0.0	0.1	0.1	0.0
65～69	1.2	0.1	0.1	0.0	0.0	0.1	0.0
Total	13.8	8.2	7.4	0.3	1.0	0.5	0.4

S & R

Age	Foreign language	Business relation	Industry relation	Medicine care	Education	Civilization	Natural science
10～14	22.6	0.3	0.5	2.6	1.9	5.4	7.4
15～19	25.4	4.4	3.4	6.5	4.1	11.5	12.5
20～24	17.3	9.3	5.9	5.6	4.8	8.2	6.2
25～29	11.6	9.7	4.9	4.9	3.3	3.3	2.2
30～34	10.1	9.9	4.6	5.3	4.2	3.8	2.4
35～39	7.9	8.7	3.8	5.4	5.7	3.3	2.6
40～44	7.1	8.4	3.9	5.3	5.8	3.5	2.8
45～49	6.0	7.4	3.3	4.6	5.0	3.8	2.2
50～54	5.0	6.6	2.9	4.1	4.4	4.0	2.0
55～59	3.8	5.5	2.3	4.4	4.4	3.9	2.0
60～64	3.2	3.4	1.4	4.0	4.0	4.3	2.2
65～69	2.2	1.9	1.0	4.2	4.2	4.6	1.9
Total	9.4	5.9	3.0	4.5	4.1	4.7	3.6

whole life by action, sex, administrative prefecture of Japan." Resource III is extracted from "Percentage of persons engaged in sports by sex, age group" and "Percentage of persons engaged in studies and research by sex, age group" from the survey on "Time use and leisure activities" in the social life basis survey done by the Statistics Bureau and Statistics Center of the Management and Coordination Agency of Japan.

It is convenient to express the table format as a data cube when the data has a multi-dimensional format (Figure 3). Furthermore adding a processing operation of several stages to this permits a matching to each user's analysis site (Table 2). As examples, (a) expresses the relationship between the birthplace and language spoken at home by sex, (b) compares life time of the Kinki area with the Kanto area, and (c) compares time spent in sports and in study/research by age. Accordingly, the table expression of the form that the user requires can be shown as one scene for each of various applications.

It is common to use an original program or a statistical processing utility because there is no standard algorithm to define and create these forms. On the other hand, the format data need to be transformed into the first normal form for storage into a relational database to ensure adequate data sharing, integrity and security.

When the degree of redundancy in such forms is high or when updates are frequent, this format data is required to be further normalized and sometimes divided into tables. To use these format data or tables, it is necessary for the user to write a search sentence

Figure 3. Data cube expression

Language Spoken: *English*, *Arabic*, ⋮, *Other*

Birthplace	Male	Female	Persons
Australia	6527013	6700763	13227776
Canada	12122	13010	25132
Chile	11509	12311	23820
China	53429	57580	111009
Croatia	24990	21991	46981
Egypt	17440	16719	34159
Fiji	17331	19771	37102
Germany	53805	56526	110331
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮
elsewhere	347892	348840	696732
India	305219	311621	616840
visitor	65798	73796	139594

Table 2. Expression and processing of statistical social information (Data Scene)

(a) Relationship between birthplace and language spoken at home by sex (Data Scene I)

Sex	Australia	Canada	:	Speaks English only	Arabic (including Lebanese)	Australian Indigenous Languages	:
Male	6527013	12122	:	6657753	83698	21824	:
Female	6700763	13010	:	6840341	78268	22368	:
Persons	13227776	25132	:	6657753	83698	21824	:

(b) Comparison of life time in the Kinki area with that in the Kanto area (Data Scene II)

Area	Pref	Population (1000 people)	Sleep	Personal Hygiene	Meal	Commuting	Job
Hokkaido	Hokkaido	5,005	7.55	1.06	1.34	0.24	3.54
Kanto	Saitama	6,053	7.42	1.08	1.42	0.46	3.51
	Chiba	5,200	7.39	1.12	1.40	0.44	3.42
	Tokyo	10,674	7.44	1.10	1.43	0.38	3.51
	Kanagawa	7,410	7.38	1.10	1.40	0.46	3.45
Chubu	Shizuoka	3,316	7.45	1.07	1.36	0.29	4.04
	Aichi	6,100	7.46	1.05	1.35	0.33	4.06
Kinki	Osaka	7,833	7.39	1.08	1.40	0.40	3.46
	Hyogo	4,792	7.42	1.08	1.40	0.36	3.42
Kyusyu	Fukuoka	4,340	7.44	1.09	1.37	0.31	3.43

(c) Comparison of sports and study/research time by age (Data Scene III)

Age	Soccer	Rugby	Judo	Foreign language
10～14	40.7	0.9	3.6	22.6
15～19	24.6	1.8	7.1	25.4
20～24	11.1	0.7	0.8	17.3
25～29	6.7	0.4	0.4	11.6

that includes some "join" operations in a standard data operation language such as SQL or prepare a "view table" definition in advance.

Thus, there are many differences in schema constitution between the application and relational database approach. Because of this, it is common for an application that handles statistical data to utilize the standard file offered by the operating system or a popular spreadsheet program. However, it is desirable to store and manage basis data with a widely used database management system in an environment that permits displaying, sharing and securing the data while updating in through a network.

Furthermore, because the two above characteristics of statistical social information, that there is a low update frequency and there is much addition done in a bundle, a mechanism is required to ensure long-term synchronization between the basis data in the database system and the data in the data warehouse.

The online analytical processing (OLAP) concept was born from dynamic complicated data processing becoming necessary at user sites due to the focus of data processing in enterprise applications moving toward end user computing.

In OLAP, the purpose of data processing is not only simple report preparation but utilization and analysis of data to support decision making in such ways as problem discovery and alternate plan presentation. Furthermore, in opposition to the database concept that chiefly aims rapidity for and shared use, an approach to information preservation that emphasizes each client's particular demands is called data warehouse (Codd, Codd & Salley). Data warehouse handles data management from the viewpoints of application and history record rather than the rapid processing of data updates.

The concepts of OLAP and data warehouse can be considered the result of widely used database systems having become alienated from the viewpoints of end users and business because too much emphasis has been put on the uniformity and update consistency of data. There has not yet been a standard proposed for the realization method and operation system of OLAP or data warehouse, and the research is still at a groping stage.

This work proposes a data management method that exploits the characteristic of statistical social data ; the reproduction of data at the user site is maintained in the sequence of the original schema transformation operation in data warehouse while preserving basic data resources in a conventional relational database.

The reproduction data is called "Data Scene," and the sequence of schema transformation operation is called *Sequence Operations for Data Scene Construction* (SODC). Each example in Table 2 is a data scene.

Whenever original resource data are changed, SODC generally needs to be applied to it for integrity control.

DATA-SCENE CONSTRUCTION

Developing an implementation method having a multi-dimensional nature is the most significant problem most in adopting OLAP and data warehouse. The following three methods have been proposed (Ishii 1995).

(1) Implementation in database structure

This implementation method features an expanded relational database that supports multi-value attributes or repeating group attributes. It has difficulty with dynamic schema transformation but is excellent for managing nested data or three-dimensional data.

(2) Implementation in a database view

This method uses complex view support of a relational database and can update data in real time as well as realize complicated schema transformation. However, it is inappropriate in the OLAP environment, which requires a heavy load to maintain the view that is always linked to an outside site, i.e., a personal desktop computer.

(3) Implementation at the application program level

In this method, data republication is created and maintained in the individual application program. Here, it is possible to realize an environment closed to the user site, but the load of the application program is higher.

We have adopted the third method to realize our OLAP and data warehouse for evaluating the flexibility at the application site. At this time, our system proposes the introduction of optimization procedures in OLAP operations to avoid the problem of high load at the application site. For example, not all of the updates of the resources give notice of the site of the data scene, and only the difference data related to the data scene are sufficient (Figure 4).

The support system proposed by this research plays an important role in controlling consistency. It involves a surveillance and notification function, which reflects the updated data of the resource to the site of the data scene and smoothly conforms to SODC with the appropriate timing (Figure 5).

The processing complete data form required for each user is called a data scene. Table 2 shows examples of data scenes. Users construct their own data scene while receiving the support of OLAP in the data warehouse.

Data scene and SODC are able to define data in the following manner. In Figure 6 the data scene of site X is known as Sx, which is the target data set that be continuously given several schema transform operations and is called SODC as in algebra. The notation R is a data resource, Op is application operation, and r expresses the result applied the sequence operations.

It is assumed that the operation set of SODC simply consists of the relational algebra and schema transformation operations that include Transpose and Nest. Each operation is described as follows :

Sel (F | R) : Selection operation on the condition *F* to the relation *R*.

Figure 4. Update of data scenes

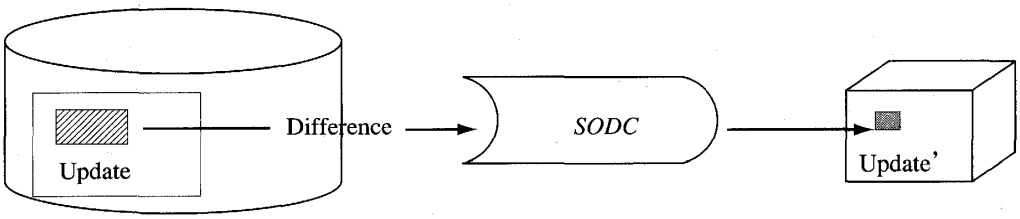


Figure 5. Update surveillance function

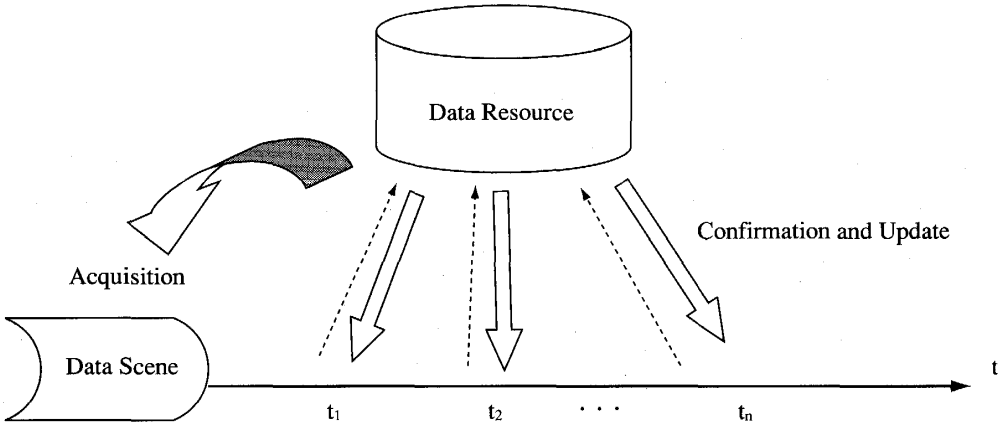
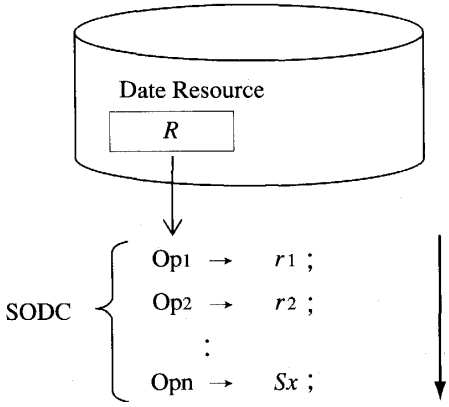


Figure 6. Sequence Operations for Data Scene Construction



Proj (A | R) : Projection operation on the attribute set *A* to the relation *R*.

Join (R, S) : Equi-Join of the relation *R* and the relation *S*.

Gen (A = value, F | R) : Generalization operation that sets the value to the new Attribute set *A*, to tuples that fills condition *F* of the relation *R*.

Nest (A | R) : Nest operation of the attribute set *A* to the relation *R*.

Trans (R) : Transpose operation to the attribute and key value of the relation *R*.

Each SODC shown in the second section is as follows and also illustrated in Figure 7.

Data Scene I)

Trans (BIRTH) → A ;

Trans (LANG) → B ;

Join (A, B) → S_I ;

Data Scene II)

Gen (Area = "Hokkaido", Pref = "Hokkaido" | LIFE) → A ;

Gen (Area="Tohoku", Pref="Aomori"∨"Iwate"∨"Miyagi"∨"Akita"∨"Yamagata"∨"Fukushima" | A) → B ;

Gen (Area="Kanto", Pref="Ibaraki"∨"Tochigi"∨"Gunma" ∨ "Saitama" ∨ "Chiba" ∨ "Tokyo" ∨ "Kanagawa" | B) → C ;

Gen (Area="Chubu", Pref="Niigata"∨"Toyama"∨"Ishikawa"∨"Fukui"∨"Yamanashi"∨"Nagano"∨"Gifu"∨"Shizuoka"∨"Aichi" | C) → D ;

Gen (Area = "kinki", Pref = "Mie" ∨ "Shiga" ∨ "Kyoto" ∨ "Osaka" ∨ "Hyogo" ∨ "Wakayama" | D) → E ;

Gen (Area = "Chugoku", Pref = "Tottori" ∨ "Shimane" ∨ "Okayama" ∨ "Hiroshima" ∨ "Yamaguchi" | E) → F ;

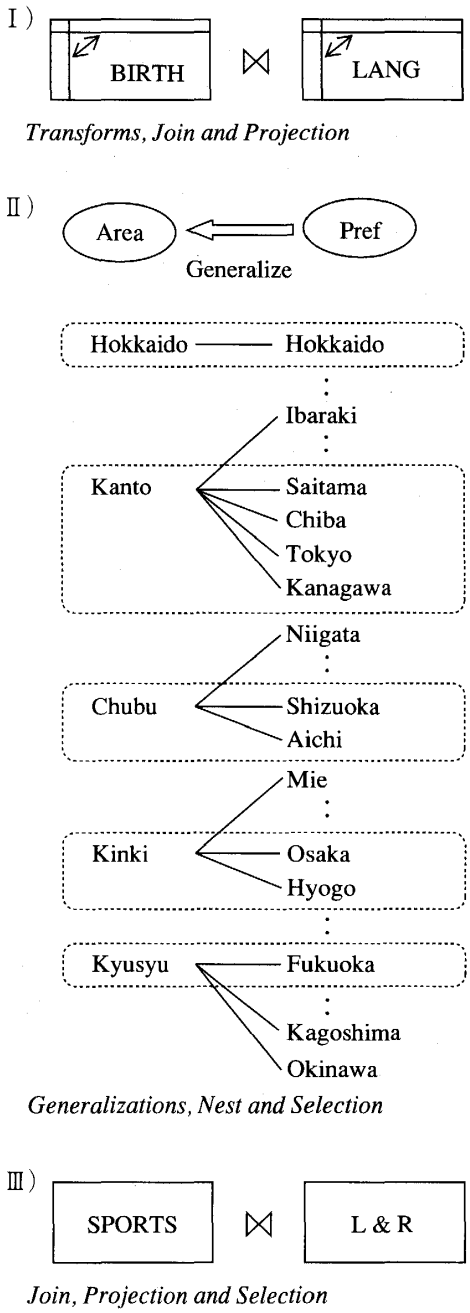
Gen (Area = "Shikoku", Pref = "Tokushima" ∨ "Kagawa"∨"Ehime"∨"Kochi" | F) → G ;

Gen (Area="Kyusyu", Pref="Fukuoka"∨"Saga"∨"Nagasaki"∨"Kumamoto"∨"Oita"∨"Miyazaki"∨"Kagoshima"∨"Okinawa" | G) → H ;

Nest (Pref | H) → I ;

Sel (Population > = 3000 | I) → S_{II} ; Pref shows the complement set of attribute Pref.

Figure 7. Sample of SODC process



Data Scene III)

Join (SPORTS, S & R) → A ;

Proj (Soccer, Rugby, Judo, Foreign Language | A) → B ;

Sel (age<30, B) → S_{III} ;

SODC needs to be re-executed to update the data scene when there is a change in the resource data. When there is a change notification from the resource basis relation, it is important to limit the range where the object spreads to reduce the load of update operations as much as possible and to enhance the efficiency of the SODC execution. The

core work of this research is to prepare useful rules and algorithms for SODC execution. The next section presents transformation examples by using several useful rules when the OLAP computation consists of the above SODC.

INTERNAL PROCESSING FOR OPTIMIZATION

This section presents one idea for an efficient way to construct an individual data scene such as the data warehouse. This idea is to smoothly convert the SODC for each user defines to another effective sequence of operations that are equivalent to those of the original SODC.

Because a user-defined SODC does not necessarily provide good performance, the system evaluates it and remakes the sequence of the operation set to reduce the amounts of calculation and communications at the site of the data scene, as shown in Figure 8. Furthermore, if it is possible to minimize the notification range of the influence update associated with the data resource update, additional economizing can be achieved.

We examined the efficiency of the conversion with the examples by using SQL and basis inside operations.

The following transformation is possible from our existing research and rule learned by experience (Nishimoto & Ura 1989b), (Nishimoto 1996).

This rule can be applied to Data Scene I.

If X and Y are relations which have same attributes set, then *Join* (*Trans* (X) | *Trans* (Y)) converts to *Trans* ($X \cup Y$).

As a result, SODC is able to transform it in the following manner:

BIRTH ∪ LANG → A;

Trans (A) → S_I;

This transformation is possible due to the characteristics of the statistical data, which consists of a category variable, avoids Join and can substitute with Union.

Data Scene II is able to transform it in the following manner from the relation between *Nest* operation and Group By:

Sel (Population>=3000 | LIFE) → A;

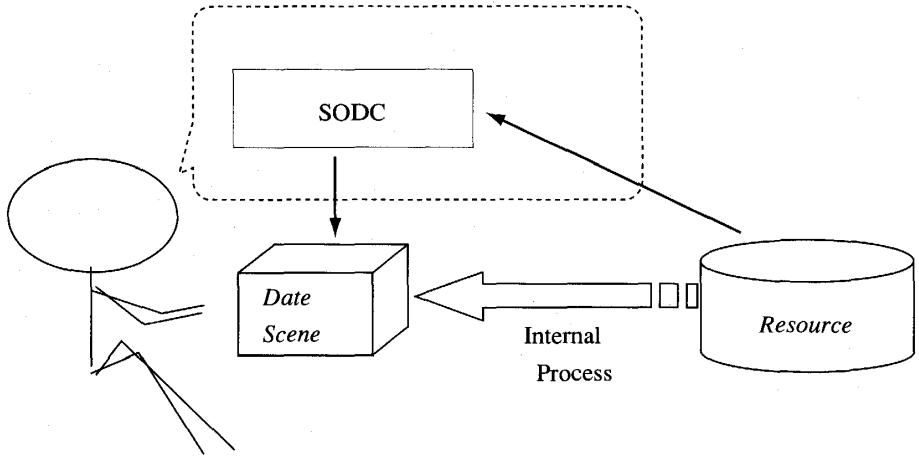
AddAttribute (Area | A);

Update (Area="Hokkaido", Pref="Hokkaido" | LIFE) → A;

Update (Area="Tohoku", Pref="Aomori"∨"Iwate"∨"Miyagi"∨"Akita"∨"Yamagata"∨"Fukushima" | A) → B;

Update (Area="Kanto", Pref="Ibaraki"∨"Tochigi"∨"Gunma"∨"Saitama"∨"Chiba"∨

Figure 8. SODC inside conversion



"Tokyo" ∨"Kanagawa" | B) → C ;

Update (Area="Chubu", Pref="Niigata"∨"Toyama"∨"Ishikawa"∨"Fukui"∨"Yamanashi" ∨"Nagano"∨"Gifu"∨"Shizuoka"∨"Aichi" | C) → D ;

Update (Area="kinki", Pref="Mie"∨"Shiga"∨"Kyoto"∨"Osaka"∨"Hyogo"∨"Wakayama" | D) → E ;

Update (Area = "Chugoku", Pref = "Tottori" ∨ "Shimane" ∨ "Okayama" ∨ "Hiroshima" ∨ "Yamaguchi" | E) → F ;

Update (Area = "Shikoku", Pref = "Tokushima" ∨ "Kagawa" ∨ "Ehime" ∨ "Kochi" | F) → G ;

Update (Area="Kyusyu", Pref="Fukuoka"∨"Saga"∨"Nagasaki"∨"Kumamoto"∨"Oita"∨ "Miyazaki"∨"Kagoshima"∨"Okinawa" | G) → H ;

GroupBy (Pref | H) → S_{II} ;

The operation *AddAttribute* (A | R) adds new attribute *A*, Update (set clause, F | R) is the operation that performs the value change of tuples to meet the condition *F*, and *GroupBy* (A | R) is the operation that groups and connects tuples to a common value of specified attributes.

These operations can be implemented easily by SQL and macro commands. The load of Update and GroupBy of the sequence are reduced by preceding them with Selection.

Data Scene III uses an optimization rule (Ullman 82) of a general relation database and is able to transform in the following manner :

Sel (age<30, SPROTS) → A ;

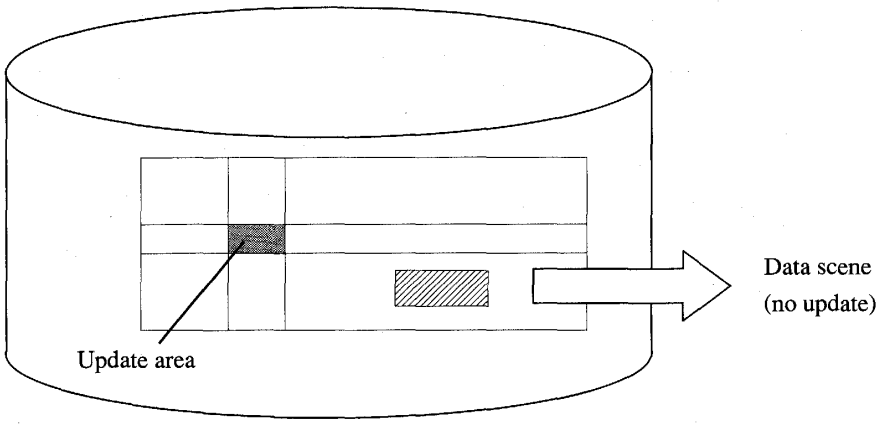
Proj (Soccer, Rugby, Judo | A) → B ;

Sel (age<30, S & R) → C ;

Proj (Foreign Language | C) → D ;

Join (B, D) → S_{III} ;

Figure 9. Judgment of update spread



The effect of this transformation reduced the input tuples number to the Join operation by preceding selection it with and projection.

The revision and correction of daily, weekly or yearly data are generally given as the data update because the presentation form of statistical and other data is decided in advance. Along with Update of resource R, the update is necessarily spread over the data scene Sx, but it can avoid the influence of the update spread when it follows in the sequence (Figure 9).

Assume that A_D is the attribute set of the data scene and A_U is the attribute set that appears setting clause in the update operation on the resource relation. If $A_D \cap A_U = \phi$ is satisfied, then there is no update spread to the data scene.

Furthermore, assume that F_u is the condition of the update operation on the resource relation. If $Sel\ (F_U \mid S_x) = \phi$ is satisfied, then there is no update spread to the data scene. Moreover, we are able to discover this condition early by making the introductory remarks Selection and Projection as in the example of Data Scene III.

We can instantly judge whether an update of the data scene is necessary by checking the update log of the server and SODC. Although the two above rules illustrate the simplest case, we can expand them to a realistic case that accompanies schema transformation of the data grid system of social information.

CONCLUSION

It is an important task to arrange an environment where it is possible to exchange and process statistical social information freely while ensuring economy and security by using a conventional database management system.

The purpose of this research is to improve the use of statistical social data by adopting

OLAP technology. This paper presents a method that is a first step in this direction and describes various efficiency problems. We intend to create an optimization algorithm that better fits a more generalized SODC and to advance a concrete implementation.

REFERENCES

Codd, E. F., Codd, S. B. & Salley, C. T. 1993, 'Providing OLAP to User Analysis : An IT Mandate', Arborsoft-paper (*http://dev.hyperion.com/resource_library/white_papers/providing_olap_to_user_analysts_2.cfm*).

Ishii, Y. 1995, *The Data Warehouse,* Management Science Research Center of Japan.

Jaeschke, G.& Schek, H.-J. 1982, 'Remarks on the Algebra of Non First Normal Form Relations', *Proceedings of ACM Principles of Database System,* : 124-138.

Jarvelin, K. & Niemi, T. 1995, 'An NF2 Relational Interface for Document Retrieval, Restructuring and Aggregation', *Proceedings of ACM SIGIR' 95* : 102-110.

MCAJ 1998, The home page of Statistics Bureau & Statistics Center, Management and Coordination Agency of Japan, *URL:http://www.stat.go.jp/.*

Niemi, T. & Jarvelin, K. 1995, 'A Straightforward NF2 Relational Interface with Applications in Information Retrieval', *Information Processing and Management,* 31 (2) : 215-231.

Nishimoto, H. & Ura, S. 1989a, 'Query Processing toward Complex View Support for a Library Database System', *Designing and Using Human-Computer Interfaces and Knowledge Based Systems,* Salvendy, G & Smith, M. J. (Eds.), Elsevier Science Publishers, Netherlands : 203-210.

Nishimoto, H. & Ura, S. 1989b, 'Complex View Support for a Library Database System', *Information Processing and Management,* 25 (5) : 515-525.

Nishimoto, H. 1996 'Open type OLAP system for Social Database Construction', *Kansai University Inf. Processing Center Forum,* 11 : 39-45.

Nishimoto, H. 1997, 'Schema Transformation and Effects of Social Statistical Databases', *Ryukoku Journal of Economic Studies,* 36 (3). 1-15

Nishimoto, H. 2001, "User logging system and analysis method" (In Japanese), *Human Interface Symposium 2001.*

Nishimoto, H. 2005, "Extraction System of Historical Facts from Japanese Documents", *The 3rd ACS/IEEE International Conference on Computer Systems and Applications, 2005.*

Ullman, J. D. 1982, *Principles of Database,* Computer Science Press.

ABS 2005, The home page of Australian Bureau of Statistics World Wide Web Information Service, *URL:http://www.abs.gov.au/.*

(Received May 2, 2005)