Web Service and Structure of University Data

- Development of Japanese Higher Education Database -

Masaaki Ida
Department of Research and Development
National Institution for Academic Degrees and University Evaluation
Tokyo, Japan
ida@niad.ac.jp

***Abstract*—This paper describes the state of research and development of web services and data structure of university survey data, especially, Japanese higher education survey data. Our research and development are aimed for university comparative analysis on the data with consideration on general university information structure. Institutional data of university, college, or college of technology are substantially important for data analysis or knowledge discovery in the higher education management field. However, university institutional data are not necessarily standardized and compiled, so it is not easy to integrate their information for various reporting and data analysis. In the past decade, investigation of the integrated university data sets has been done to deal with various kinds of university institutional information including university survey or school basic survey data in Japan, by providing the structured university data via web services. This paper describes the state of research and development of web services and data structure of university survey data, which are utilized for understanding the university characteristics. We describe: (i) a proposal of the XML schema for Japanese university data (ii) development of various Web APIs (XML, JSON) of university database for survey cards.**

Keywords- database; school basic survey; web service; data structure.

I. INTRODUCTION

A. University Insititutional Data

Development of education-related databases is substantially important for data analysis and knowledge discovery in education field worldwide [1]. Institutional data of universities are not easy to analyze because they are not necessarily standardized and integrated in each university itself or national education-related agencies. However, some advanced education-related database systems are progressively developing.

In the United States, *Integrated Postsecondary Education Data System* (*IPEDS*) [2] of National Center for Education Statistics has been developed to collect and analyze basic institutional information about universities and colleges in U.S. IPEDS standardizes and accumulates the information nationwide. This system comprehensively holds basic institutional data, such as institutional characteristic, degree completion, enrollment, human resource, finance, student financial aid, graduation rate, and so on. Moreover, this system is equipped with facilitated data analysis tools to conduct university comparative analysis.

College Portrait [3] is also higher education database that is "a source of basic, comparable information about public colleges and institutions presented in a user-friendly format".

In European area, for example, *Unistat* [4] system is developed in order to search, review, and compare subjects at UK universities and colleges. It "is the official website to help you make an informed choice when deciding which UK university or college to apply to. It includes the results of the latest National Student Survey".

In Asia, Korean government started their university evaluation system which consists of the university information disclosure system, *Korea Academyinfo* [5], conducted by the Korean Council for University Education, so that higher educational information of Korean universities is published on their web site.

These databases are well-organized and comprehensive systems with easy web-based operation on their web pages. However, in order to cooperate (mash-up) with other information systems, e.g., in-house database developed in individual higher education institutions, or external web services (Google Chart API), more improved systems should be equipped with various web service functions and standardized data sets.

B. University Basic Survey in Japan

In Japan, Ministry of Education, Culture, Sports, Science and Technology collects basic information about higher education institutions in Japan [6]. This basic survey data include the yearly information of higher education institutions, such as number of faculties or staffs, number of enrolled students by grade (undergraduate, graduate, foreign student), number of graduates by subsequent course, number of those who were employed after graduation by industry and by occupation, faculties, facilities, and financial data. Fig. 1 shows an example of university basic survey sheets.

Its online submission system via internet has been developed, which is equipped with authentication and encryption functions. Persons of universities in charge of data

 ISBN: 978-1-61208-280-6

submission must fill the data into electronic files (PDF files) and submit them through internet to the ministry's data collection server.

The data in the files are saved as XML data as duplication data of the submitted PDF file data. However, the XML data structure is designed only for data submission purpose. And it is not designed for data analysis purpose. These submitted survey data are compiled and published by the ministry as summarized statistics data tables. Parts of the survey data of some universities are published on their web sites. However, under the present circumstances sufficient amounts of data, detailed and standardized data required to conduct intercollegiate comparative analysis are not necessarily obtained. It is difficult to examine detailed situation of higher education institutions from various perspectives.

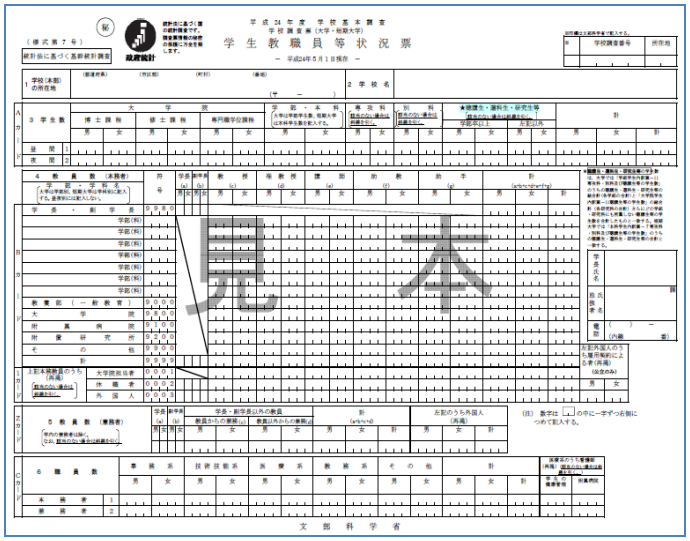

Fig. 1. Survey sheet (number of faculty and student) [7].

C. eXtensible Business Reporting Language

The *eXtensible Business Reporting Language* (*XBRL*) [8-10] is one of the computer languages based on XML, which is a standard for the electronic exchange of data between businesses on the internet. XBRL utilizes some XML technologies such as XML Schema and XLink standards. Based on XML, tags are applied to items of financial business data so that financial data can be processed efficiently by computer software. XBRL is implemented in a wide range of scenes such as tax payment system and financial data transfer system in stock exchange.

XBRL consists of an *XBRL Instance*, containing primarily the business facts being reported, and a set of *Taxonomies*, defining metadata about these facts, such as what the facts mean and how they relate to one another:

- *Instance* holds the following information: business facts, contexts (date and time information, scenario), units, footnote, and references.
- *Taxonomies* are the reporting-area specific hierarchical dictionaries. The XBRL specification defines five different kinds of linkbases (*Label* linkbase, *Reference* linkbase, *Definition* linkbase, *Calculation* linkbase, and *Presentation* linkbase). Taxonomies consist of hierarchical structure.

Different taxonomies are required for different purposes in various application fields, therefore, we extend or modify taxonomies for university information.

In the following sections of this paper, Section II presents structured data and university information database system. Section III presents construction of university survey database and schema of university survey data. Section IV presents web service and application of university survey data.

II. UNIVERSITY INFORMATION DATABASE SYSTEM

A. Structured Data Sets and Taxonomies

In this section, we describe the general information structure of university information and the XBRL extension for university information as shown in Fig. 2.

As shown in the lower right side of the Fig. 2, various data and databases are developed in each field such as enrollment, finance, personnel data, and so on, which possess and provide their data in various manners cooperating with other databases. These databases provide specified and designated information as HTML files or XML files via Web API [11].

In the upper left side of the figure, taxonomy hierarchical structure expresses the generalized information structure with specified taxonomy levels, three level structure, such as national or international standard level, institution type level, and individual institution level. The structured taxonomy sets and their data stored in various concrete databases are linked mutually, so that rigid and various definitions of data and hierarchical structures are possible, and flexibility for the changes of database equipment and time transition of data definition is guaranteed.

Their data are transmitted via web services, which are composed of REST type Web APIs (JSONP) with XML and JSON data transferred into the Business Intelligence system as shown in the right-hand side of the figure. The Business Intelligence system produces integrated university reports and the results of comparative analysis of higher education institutions combined with outer databases, reporting system, and analysis systems (mash up).

 ISBN: 978-1-61208-280-6

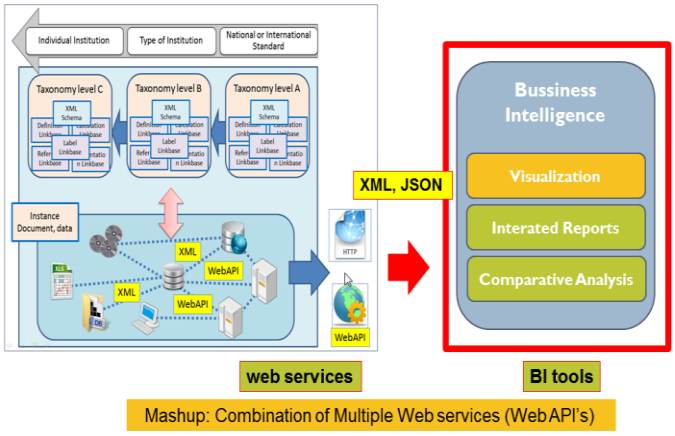

Fig. 2. University information system and taxonomy hierarchical structure.

B. *Extension of Taxonomies*

We conducted taxonomy design for university financial statements in Japan [12], i.e., we proposed taxonomy for *Balance Sheet* and *Income Statement* of university, which is an extension from ordinal XBRL taxonomy. The structure of proposed taxonomy for university financial statement consists of five different kinds of linkbases or files. In the taxonomy, the definitions of the vocabulary (Japanese and English) are assembled in the definition file folder for general company or in the extended university finance file folder. The presentation orders or format structure of the financial elements in the financial statement are defined in the university finance presentation file folder.

Adding to the extension for university finance data, we proposed further extension for university institution data [13], which aims for expressing various indexes of institutional situation, e.g.,

Ratio of education expenses to student
= (*Education expenses*) / (*Number of student*).

In order to treat various indexes related to higher education institutions for institutional reports, we extended the general hierarchy of terminologies, e.g., taxonomy extension such as *number of undergraduate male* and *under graduate female* extended to ordinal taxonomy.

III. UNIVERSITY SURVEY DATA

A. *Construction of University Survey Database*

In this section, we focus on the data of *University Basic Survey* and its database. We have examined the data structure of university survey of Japan. Based on the examination we are considering database system as an infrastructure for applying them to analyze various aspects of universities. As a pilot system, we tried to collect sample survey data of U.S. and store them into the developed database so far.

Up to now for Japanese university survey data, we designed and developed useful tools that upload and transform the XML data in university basic survey, and store them into a relational database. The data flow and process for the survey are shown in the Fig. 3.

University Survey Formats (sheets) are consist of detailed university information card, such as "Institutional structure, faculty member and staff (7_A (number of students), 7_B1 (number of academic staffs), 7_C, 7_Z)", "Student (8_D2 (number of students of each department), 8_3, 8_E, 8_G, 8_7, 8_R)", "Graduate student (9_H4, 9_5, 9_I, 9_S, 9_8)", "College (10_J6, 10_9, 10_K)", "Foreign student (11)", "Facility (20)", "Financial data (22A, 22B)", "Employment (30)", and so on. The detailed card information, such as card 7_A, card 22A, are described in reference [7].

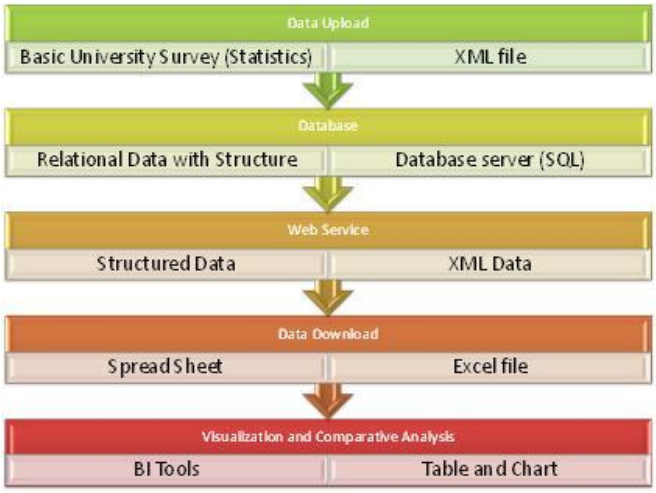

Fig. 3. Data flow and process for Japanese university basic survey.

B. *Schema of University Survey Data*

In this paper, we propose XML schema of various university information in University Survey Formats. The XML schema represents the University Survey information of several universities into a standard and integrated university formats.

Fig. 4 shows the overview of XML Schema structure of university information. Each dashed-green rectangle indicates university level schema, and dashed-blue rectangle indicates department level schema. Fig. 4 is a part of whole structure of the XML Schema of university survey information.

 ISBN: 978-1-61208-280-6

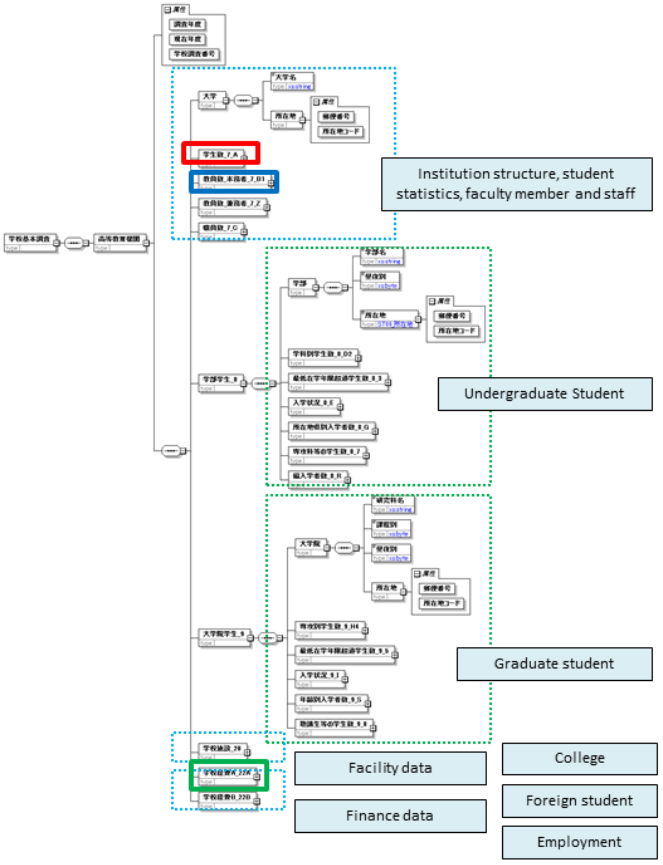

Fig. 4. XML Schema for Japanese university basic survey.

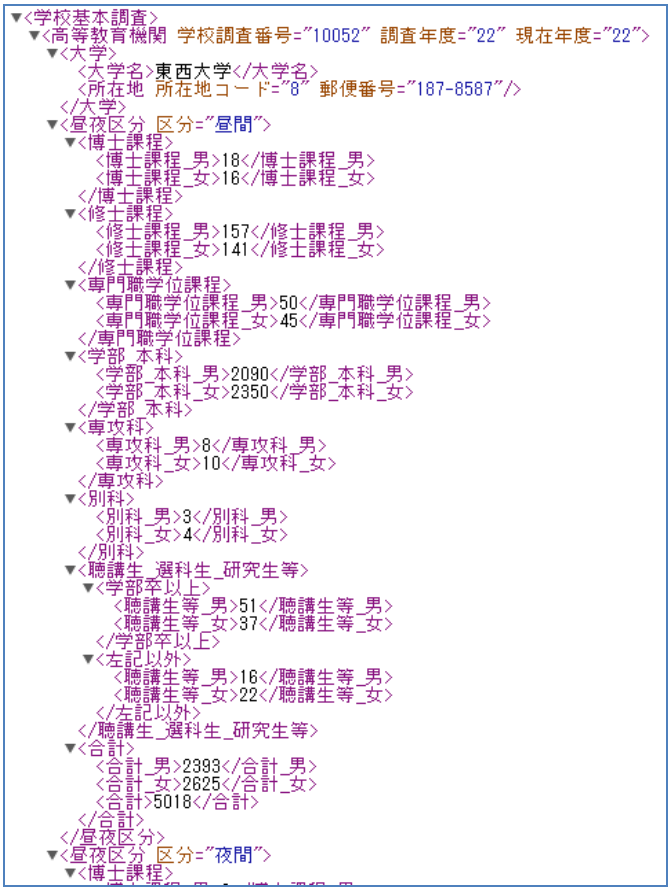

Fig. 5. Output example of web service (Number of student; Format 7).

IV. WEB SERVICE OF UNIVERSITY SURVEY DATA

A. *Web API of University Survey Database*

In this paper, we develop two kinds of Web APIs (XML, JSON) of university database for survey cards.

Several web APIs were considered which are suitable for data analysis and data dissemination. This type of web services cause independency of application modules which can be easily redesigned and reformed.

The following are examples of RESTful web service (Web API) [14] retrieved by survey year and institution number of Japanese universities and so on.

Fig. 5 shows an example of output of web service on the number of student form the university survey (survey format 7), that is corresponding to red rectangle part in Fig. 4. Elements in Japanese mean “university name”, “address”, “number of undergraduate student”, “number of graduate student (master)”, “number of graduate student (doctor)”, and so on.

Fig. 6 shows an example of output of the web service concerning university financial data from the university survey (survey format 22) , that is corresponding to green rectangle part in Fig. 4. Elements in Japanese mean “university name”, “address”, “faculty salary”, “staff salary”, “education expense”, “management expense”, and so on (unit: 1,000 Yen).

Fig. 6. Output example of web service (Finance; Format 22).

B. Web API and Reporting

Receiving the web service data form database system at the client side, this system generates spreadsheet or PDF files, which are *reporting sheets* subject to the conventional formats of university survey booklets.

When the XML data from web services are obtained, that is corresponding to blue rectangle part in Fig. 4. The data are simply transmitted into spreadsheet files with data bindings as shown in Fig. 7. The elements in Japanese mean the number of "professor male", "professor female", "associate professor male", "associate professor female", and so on for each university faculty. The Excel sheets only possess the relationships between items of XML returned by web service and the columns of sheets. Therefore, in case that some data on University Database are modified, we don't need to adjust the spreadsheet structure, and the data in each sheet would be automatically changed.

University DB system –> Web API (REST) –> XML –> Data binding –> Excel Sheets

大学	学部	教授		准教授		講師		助教	
		男	女	男	女	男	女	男	
Name	FacultyNan	ProfessorM	ProfessorF	AssociateF	AssociateF	LecturerMa	LecturerFe	AssistantF	Assis
A医科大学	医学部	20	50	40	20	40	20	40	
A医科大学	歯学部	34	53	45	60	56	57	58	
A医科大学	看護学部	62	63	64	32	66	67	68	
A医科大学	薬学部	15	15	15	15	15	15	15	
B大学	文学部	46	9	26	11	8	1	1	
B大学	教育学部	21	5	32	10	11	8	1	
B大学	法学部	46	42	22	3	10	30	30	
B大学	経済学部	32	45	52	3	23	30	30	
B大学	理学部	29	23	23	23	0	30	30	
B大学	工学部	43	45	10	10	43	52	30	
B大学	生物理工学部	30	30	30	30	11	9	26	
B大学	農学部	30	30	30	30	23	5	32	
B大学	環境学部	32	30	30	30	0	30	22	
B大学	経済学部	62	52	20	50	40	20	1	
C工業大学	理学部	46	9	34	53	45	60	66	
C工業大学	工学部	20	5	62	63	64	32	15	
C工業大学	理工学部	32	30	15	15	15	15	8	
C工業大学	システム工学部	10	65	46	9	20	50	40	
C工業大学	生物理工学部	64	32	21	5	34	53	45	
C工業大学	法経学部	15	15	46	42	62	63	20	
D大学	文学部	26	11	32	45	15	15	34	
D大学	教育学部	32	10	29	20	50	40	20	
D大学	法学部	22	3	43	34	53	45	60	
D大学	経済学部	52	3	23	62	63	64	32	
D大学	理学部	23	23	0	15	15	15	15	
D大学	工学部	10	10	43	46	9	26	11	
D大学	農学部	30	30	11	21	5	32	10	
D大学	医学部	30	30	23	46	42	22	3	
D大学	薬学部	30	30	0	32	45	52	3	
D大学	国際教養学部	30	40	20	29	23	23	23	

Fig. 7. Reporting results: Binding of spreadsheet and web service.

C. Web API and Comparative Analysis

We can utilize these web services in case of comparative analysis. In this section, we show some examples of comparative analysis using JSON data derived from the web services and mash up with outer web APIs. Simplified and smart analyses can be executed as shown in the following examples. Outer web APIs are included in Google API (Google Chart, Map, …).

DB system –> Web API (JSONP) –> JSON –> "Mash Up" -> JavaScript (jQuery) –> Visualization (Chart, Graph)

1) Comparative Analysis on Ration of Expenditure

Fig. 8 shows an example of comparative analysis of three major national universities on expenditure ratio. The data from the web API is JSON data of JSONP call back function, so that this analysis is programmed by JavaScript combined with "*Google Chart API*". Mash-up programing is not difficult for beginner of programming because of the typical combination of web services for university comparison and analysis.

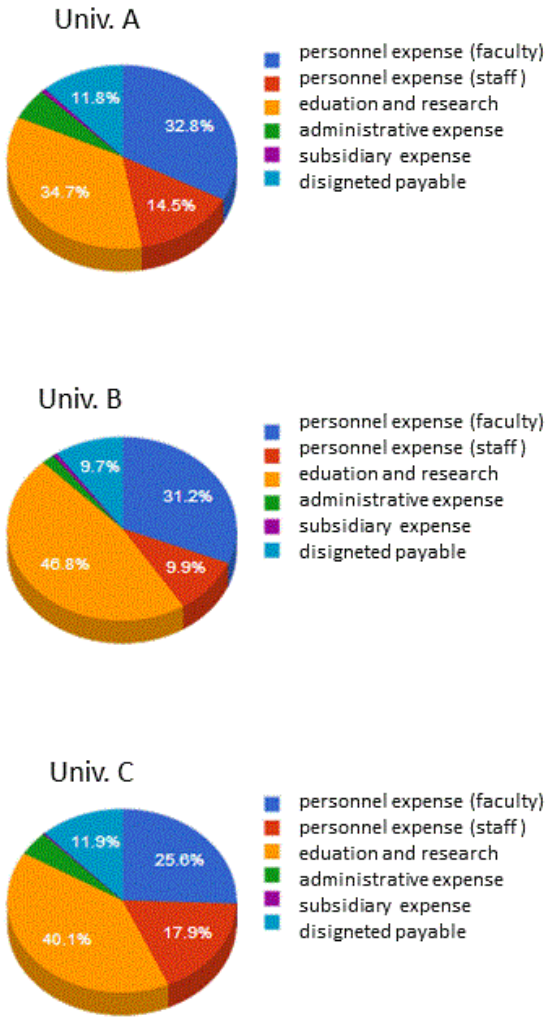

Fig. 8. Example of university survey data analysis: Annual expenditure of three national universities in Japan.

2) Comparative Analysis on Faculty Salary

Fig. 9 shows an example of university survey analysis on total faculty salary on various sections (faculty, university hospital, research institution, total amount) on four major national universities in Japan. This visualization is programmed by JavaScript combined with some JavaScript libraries such as jQuery.

 ISBN: 978-1-61208-280-6

V. CONCLUSION

Applications of education related information are substantially important for data analysis and knowledge discovery in education field. This paper described the state of research for web service and data structure of university survey data, which is utilized for analyses of university characteristics. In this paper, we (i) proposed the XML schema for Japanese university data, and (ii) developed various Web APIs (XML, JSON) of university database for survey cards. In order to handle of more general university data such as the data between some countries, we have to coordinate differences between those data for effective comparisons. We hope that our proposal will play an important role as an infrastructure for data analysis and knowledge discovery in higher education field.

REFERENCES

[1] C. Romero, S. Ventura, M. Pechenizkiy and R. Baker (eds.), Handbook of Educational Data Mining, CRC Press, 2010.

[2] Integrated Postsecondary Education Data System, IPEDS, http://nces.ed.gov/ipeds/

[3] College Portrait, http://www.collegeportraits.org/

[4] Unistat, http://unistats.direct.gov.uk/

[5] Korea Academyinfo, http://www.academyinfo.go.kr/

[6] Ministry of Education, Culture, Sports, Science and Technology, basic information of higher education institutions in Japan, http://www.mext.go.jp/b_menu/toukei/

[7] Example of survey sheet (sheet of "faculty and student"), http://www.mext.go.jp/component/b_menu/other/__icsFiles/afieldfile/2012/03/30/1318957_3.pdf

[8] eXtensible Business Reporting Language, http://www.xbrl.org/

[9] R. Debreceny, C. Felden, B. Ochocki, M. Piechocki, et al., XBRL for Interactive Data: Engineering the Information Value Chain, Springer, 2009.

[10] C.Hoffmann, and L.A. Watson, XBRL, Wiley, 2010.

[11] M. Ida, "Web Service and Visualization for Higher Education Information Providing Service," Proc. of ICSESS2010, pp.415—418, 2010.

[12] M. Ida, "XBRL Extension for Knowledge Discovery in Higher Education," the 8th International Conference on Fuzzy Systems and Knowledge Discovery, pp.2177—2180, 2011.

[13] M. Ida, "XBRL Financial Database for Higher Education Institutions," the 14th International Conference on Advanced Communication Technology, pp. 398—401, 2012.

[14] L. Richardson and S. Ruby, Restful Web Services: Oreilly & Associates Inc, 2007.

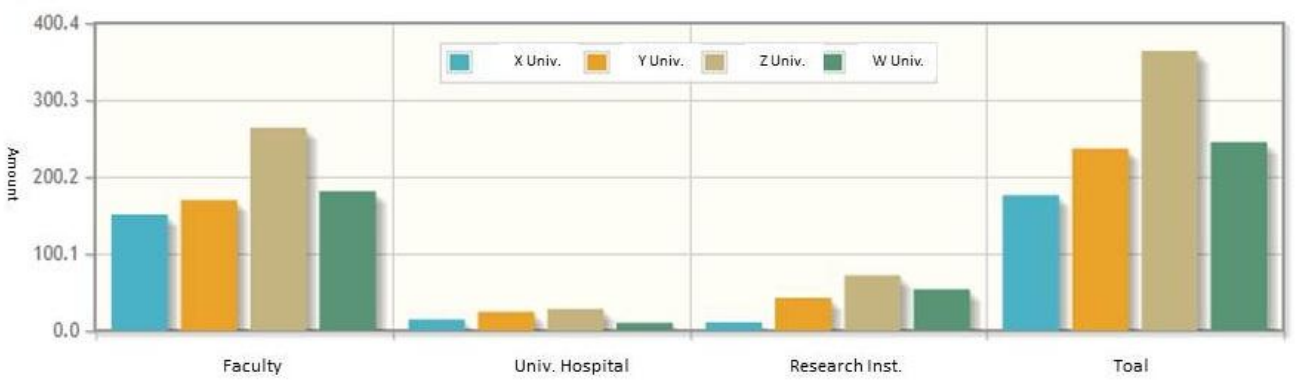

Fig. 9. Example of university survey analysis for financial information: Total salary on sections (10^9 Yen)