Integrated Statistical Education System with Interactive Components on the Internet

SAKURAI, Naoko[1] INOUE, Tatsuki[2]

[1)] *Tokyo University of Information Science, Japan (sakurai@rsch.tuis.ac.jp)*

[2)] *Waseda University, Japan (inoue@wiaps.waseda.ac.jp)*

Abstract

This paper presents the outline of our interactive project about statistics education on the Internet introducing a learning system of statistics, an online book reading system of valuable statistical writing, a searching system of statistical products and a dynamic link module system of statistics which have currently been under development. Statistics is a very suitable discipline for cross-sectional division such as exact social science. Statistics can be alive with real life data and the analysis of them leading to decision making of our social situation. Nowadays the ability for statistical data analysis is strongly needed by both the academic and business world where enormous number of digitized data (for example "POS database": point of sale) has been accumulated day by day through computer network. In addition, the large demand for statistical analysis of categorical data in psychology among personnel affairs. Statistics literacy will be one of the most important items, same as computer literacy soon. We dedicate our system to current networking and the data accumulating society.

1. Introduction

The education of statistics in Japan had brought great success of quality control in manufacturing. Though Japanese industrial products had been popular during those days, the ability has gradually been lower because of lack of statistical sense among working people. Nowadays the huge accumulated database made by the widely spread computer network has been increased to analyze with statistics reaching to decision making among various departments. Statistical analysis methodology has already been the special tool indispensable to our everyday life. Moreover, it has been exact common sense for governmental offices, public agencies, some banks, newspaper or broadcasting companies and any private companies offering open source as "database" on their web site. For example, Yamada[3] showed that the number of sites of private companies updating statistical data regularly has more than doubled from September 1998 to May 2000. This tendency has to be expected to continue. The Japanese ministry has announced the disclosure of various points of public data leading private companies to also disclose theirs. The government is also planning to publish microscopic data resources which have strongly been in demand by research people in universities or econometric organizations. Database offering will also have brought the needs for statistical analysis.

It is already quite common for U.S. and European people to search exact data from Internet web sites where economical or social survey microscopic data is available and have already been in public use for statistical analysis. For them the ability of searching, collecting and analyzing the exact data on the Internet web is more important which is causing them "digital divide". Now, in

Japan, we will meet the same situation in the very near future. The environmental situation of computer networking infrastructure has rapidly become very common for us where every document for learning is easy to find on the Internet web. As the networking condition of Japanese around universities, the business world office environment and the individual home has greatly been changing, a surprising environment for every rapid extension of broad banded network has brought us large scaled potential to lots of fields especially in education. This is excellent for the university learning condition of statistics in particular.

These two situations mentioned before, such as “abundant database” and “broad-banded computer network”, has brought the strong needs for the ability of statistical data analysis. This is also a social need. The third inevitable point is that Statistics is very suitable for the online education style with which we will be able to expect new potential comparative to prior style of only in the square classroom. The traditional way of Japanese university classes is almost composed of only explanation of mathematical aspects and few offer practical data exercises which are one of the most important features of intuitive understanding of statistics. Internet technology will make it possible to develop educational resources against new media making use of multimedia, dynamic charts and hotlinks to data-keeping sites. At the same time lots of web sites related to statistical education has been easier to develop by many faculties cooperatively. We are now in the age of global needs for statistics and statistical analysis will not disappear as long as “hot data” is alive. In the near future we will meet the age of statistics/analysis literate.

In this paper, we will introduce faculties’ projects for the development of the integrated educational web site about statistics for its dissemination covering merits/demerits and future planning.

2. Interactive Text on the Web

The first project we will show here is “ITLS” (Interactive Text for Learning Statistics), which consists of a lot of interactive materials for statistical education. There was an earlier situation of ITLS whose birth age was still on the way of computer networking development over various places, such as universities, the business office environment, public agencies and individual homes. When it started in 1999, mass education in the university classroom was still common, although the computer network environment at special private universities had been improved to a small extent. At that time there were not so many people expected all educational materials to be on the Internet. It was really reasonable to utilize both methods of developing online resources and general classroom lectures.

Along with the rapid growth of computer network around private universities, faculties of those universities have started developing cooperative online text for more dynamic and effective education of statistics. One of our expectations for using materials on the Internet through PC is that students would be actively able to participate, going ahead along the learning process ideally composed by faculties. In the developing process, faculties have standardized the total design; this is also an important point. Merits are not only to students; self-study business people and all teaching staff who could get the lecture notes as their sub text are also in favor. ITLS is totally a well-composed online text including dynamic charts and improved slides using Java Applets, expanding the search engine.

The contents of ITLS are as follows:

- Statistics and Excel
- Data Analysis –Descriptive-
- Data Analysis –Inferential-
- Basic Theory of Regression Analysis

- Analysis of Time-Series and Seasonal Adjustment
- Lectures on Multivariate Analysis
- Statistical Analysis by Statistica
- Search Engines and Links

Here is some brief explanation of each component above. “Statistics and Excel” teaches the fundamentals of data analysis procedures on the PC through the operation of spreadsheet software, for example Excel. “Data Analysis” introduces the basic concepts of statistical data analysis including practical data operation. “Statistical Analysis by Statistica” tells about introduction of one such existing software for statistics. “Search Engines and Links” shows a brief outline of the Internet, how to meet in the econometric or social database (statistical information) and how to download data.

We serve the functional covering of ITLS for better understanding and promotion of practical data analysis technologies in online text as follows; “keyword searching system and its supplementary explanation,” “browsing/downloading slides for teachers,” “introduction of related reference web sites,” “kinds of database for exercise training,” “question and answer system” and “questionnaire sheet.” These cannot be reached by general paper textbooks. Explanatory contents are structured by the mixture of simple and easy words, graphical tools, colors and colored images, hyperlinks and dynamic charts, which help users to get the statistical sense intuitively, comparative to the conventional text whose pages are mostly completed by mathematical formula. If the user would like to see a more mathematical theory, they can do that through another project “EBSA” being explained in the next section. For PC beginners, kind conceptual panels are served which consists of basic practices in data entry, file saving, web browsing/searching, how to setup email design and how to create web pages. These also are helpful to the course of information processing for beginners. All contents are currently offered as follows; over 300 slide panels for the lecture, 120 explanations of keyword and 17 dynamic charts with Java Applets. Open mind of posting the cooperative developed material sources on the Internet would release the freedom of choice for statistical education on user, time or place. Teachers can get the merits of sharing educational materials which should regularly be updated. Breakthrough of debating for statistical education on the Internet will come out soon, appreciating these virtual exchange of opinions from all over the world.

As this project can be utilized not only by university students or business people but also the teaching side staff for every department related to statistics, it is virtually impossible to cover all aspects only by the project development done by our project members. It will need a wider variety of materials. Therefore, faculties have published project’ concepts and contents at every chance of domestic/international academic conferences for the purpose of cooperation and sharing development materials. Sakurai[4] introduced this system at an early chance of an academic meeting and its activity has brought an opportunity for international cooperation. Doing international cooperation among researchers outside of Japan is easier than before because of the rapid progress of computer network. Here is an example of an international multimedia database system including animation, sound and Java Applets which is under preparation by Hagen University in Germany. This database exhibits language-wise Applets as German, English and Spanish. Our project team is planning to publish an English version, in addition to the German version of the Java Applet, which has been developed under his project at Hagen University. We are currently mutually trading our Japanese version, created by Professor H.J. Mittag, and their German version, translating each as needed. [1] and [2]

Although one important problem still exists in network response speed with large images, clear sound or voice data which we call “multimedia material”, in the near future I hope those kinds of problems will disappear due to progressive networking technology. Not only in the university environment and business office environment, but also in the ordinary home the network condition will improve due to

building a huge backbone infrastructure. During this network age, the plan for both offering a CD version of a project and building some mirror sites will be served in order do accommodate the expected common statistics literacy age. Figures1-4 are some examples of ITLS contents.

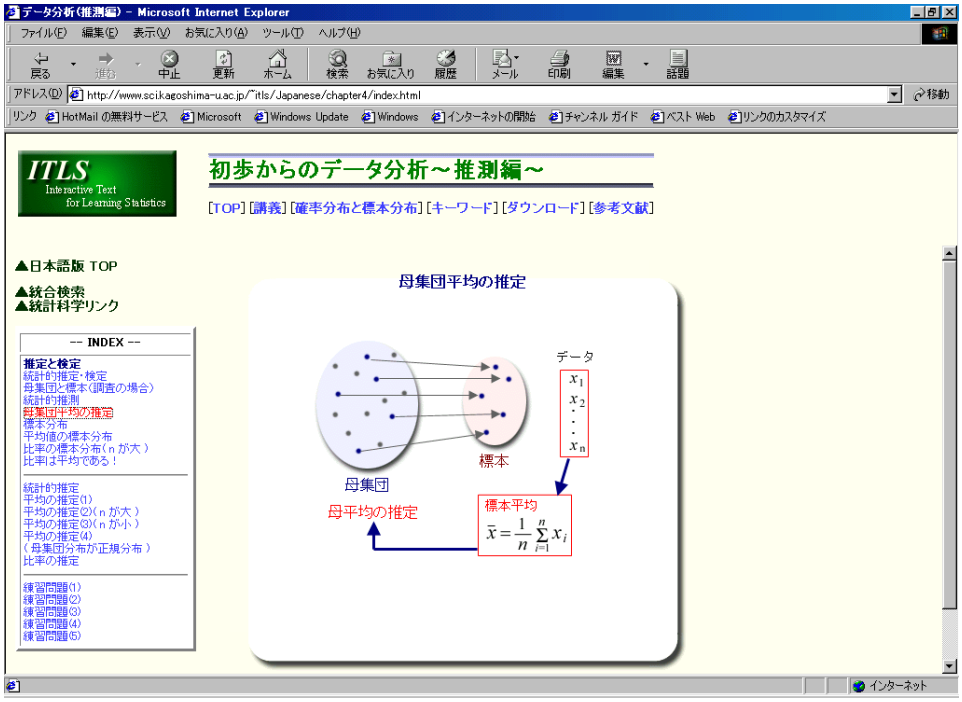

Figure 1 Descriptive & Inferential Statistics

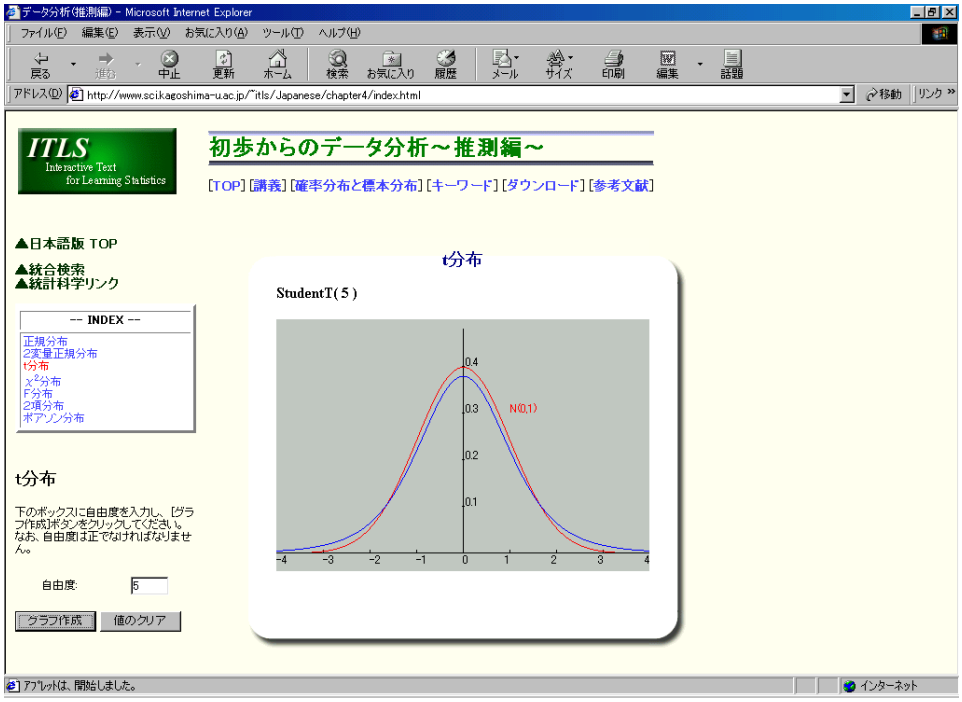

Figure 2 Inferential Statistics

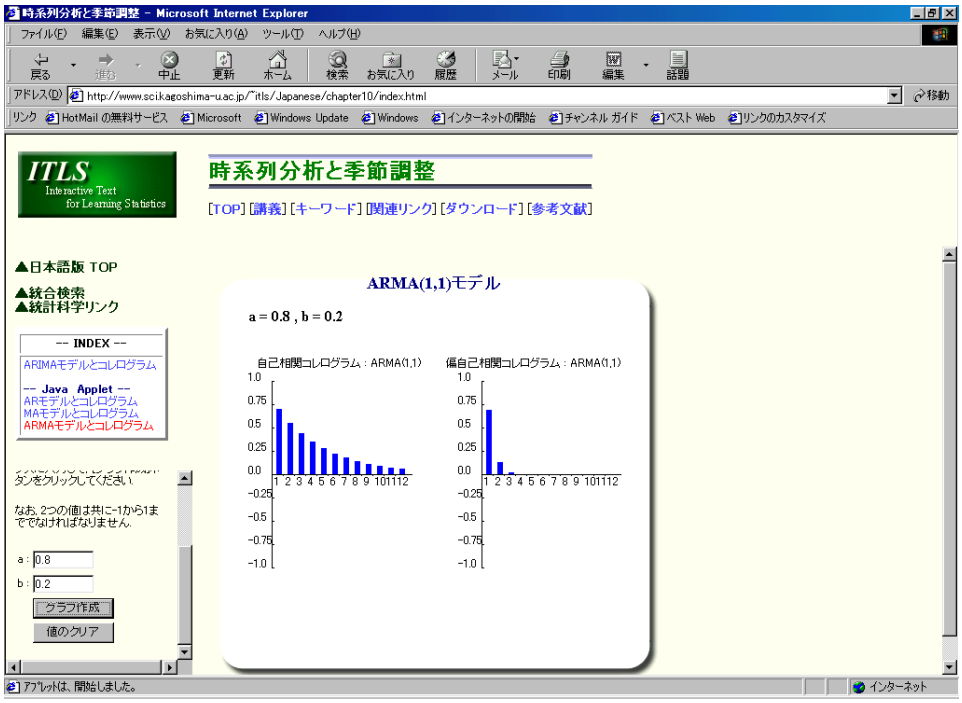

Figure 3 ARMA Model

Figure 4 Linkage with Excel

3. Online Book Reading System

We will introduce our other project named “EBSA (Electronic Book for Statistical Analysis)”. Figure 5 shows one page of EBSA with which every user who will access this site can do online reading of valuable books which are formally used as technical books for statistical science in the theoretical area or as instructional manuals for many application fields. Those books are currently out of print or have no reprint plan. We feel pity for this condition now, because those books have such high quality for their universality and usability in their scientific statistical knowledge of those ages. EBSA is one solution against this large problem. Although the start point of EBSA is May/2000, over 20 books are now available for online reading to extent. The problem for republishing is cleared by receiving official approval from copyright holders (author or the bereaved family) and first publishing companies.

All pages of original books have been scanned and transformed into PDF file formats, as every user can browse and print out selected parts. In addition, by collection and integration of all index words over every book, it has been served with a powerful searching system of a glossary database with which users will be able to refer to the entire applicable pages in each group. When a user enters a key word of some statistical science, the engine searches against the specified word across all the registered books, or the books the user specifies. As not every user is in the same circumstance considering their computer network; image files which causes heaviness on the disk and/or sound files, can sometimes be excluded for each users’ comfortable access. Users can browse or print out only the pages they want. The system also has text-based index pages so that it could often be caught and retrieved by the ordinary search engines. EBSA will enable the user to meet the original books on the Internet and increase the opportunities of learning statistical science. The extended kinds of books for “QC (quality control)” have been under development. Figures 5-6 shows some parts of the EBSA system and Figure 9 shows the integrated searching page for ITLS and EBSA.

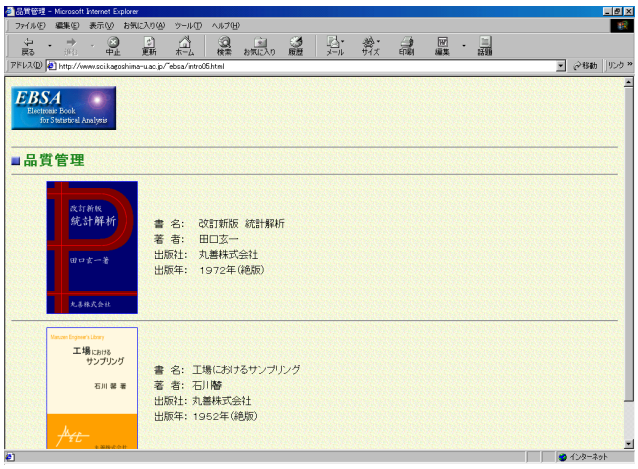

Figure 5 Book List of EBSA

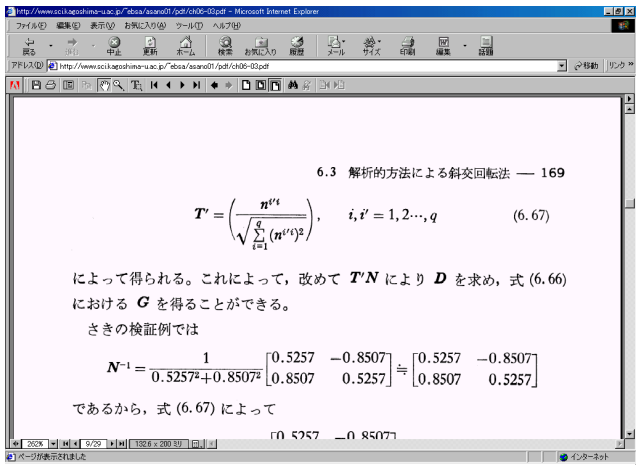

Figure 6 Content Page of EBSA

4. Other Functions

Another important content on the point of going ahead constructing the statistical education system on the Internet web is the existence of scientific statistical analysis software libraries. It is helpful for learners to analyze practical data by PC along with the theoretical learning of statistics. In fact, there are

some reports about its effectiveness. The “DLLSA (Dynamic Link Library for Statistical Analysis)” project had already been started earlier, as an independent one. DLLSA is composed by subroutine libraries, which can be called by any software on the Windows operating system or other through the Internet. Users can call only necessary parts of DLLSA libraries from their application software in their computer. DLLSA system is independent and this is a very excellent point for WWW use.
DLLSA has four main departments as follows.

- Data Management
- Calculation of Statistical Number
- Basic Statistical Quantity
- Multivariate Analysis

Here is further information about one example, Multivariate Analysis.

- Principal Component Analysis
- Metric Multidimensional Scaling
- Hierarchical Cluster Analysis
- Latent Class Analysis
- Corresponding Analysis
- Discriminate Analysis

In addition, since this system has usable interface with Microsoft Excel, the user can do statistical analysis by operating DLLSA as if they would manage statistical analysis under the excel environment. An example page of DLLSA in English is showed on Figure 7.

Let me introduce another trial of Inoue and Sakurai. We are now building a www site of serving the search engine for statistical contents particularly in the environmental department. Now, people on the earth, are under severe condition of environment, for example air, temperature, water or agricultural products. Although we commonly make use of the general purpose searching engine site when we want to find some information, we often have to meet lots of unexpected documents. Therefore we got an idea that if there is a private search engine well-designed to environmental statistical components, it will be much appreciated. The engine will show the accurate item among a large number of scattered documents against the users’ asking with keywords. This searching system is composed by the following three systems, such as collecting documents, generating index and searching interface.(See Figure 8) The accuracy and speed of search depend on the structure of the Japanese Dictionary. It is different for us to manage the Japanese words in the searching process comparative to non-Japanese words. Well made Japanese dictionaries will throw back the good searching result with CGI software. We have a plan for linkage of this site in environmental statistics to our integrated online learning system of statistics. It is pleasing for us that visitors to online the learning site also pays a visit to our searching engine site and meet various documents about environmental statistics which may guide them to the front end of research and activities in those departments. We will do an analysis of the keyword input, in due time.

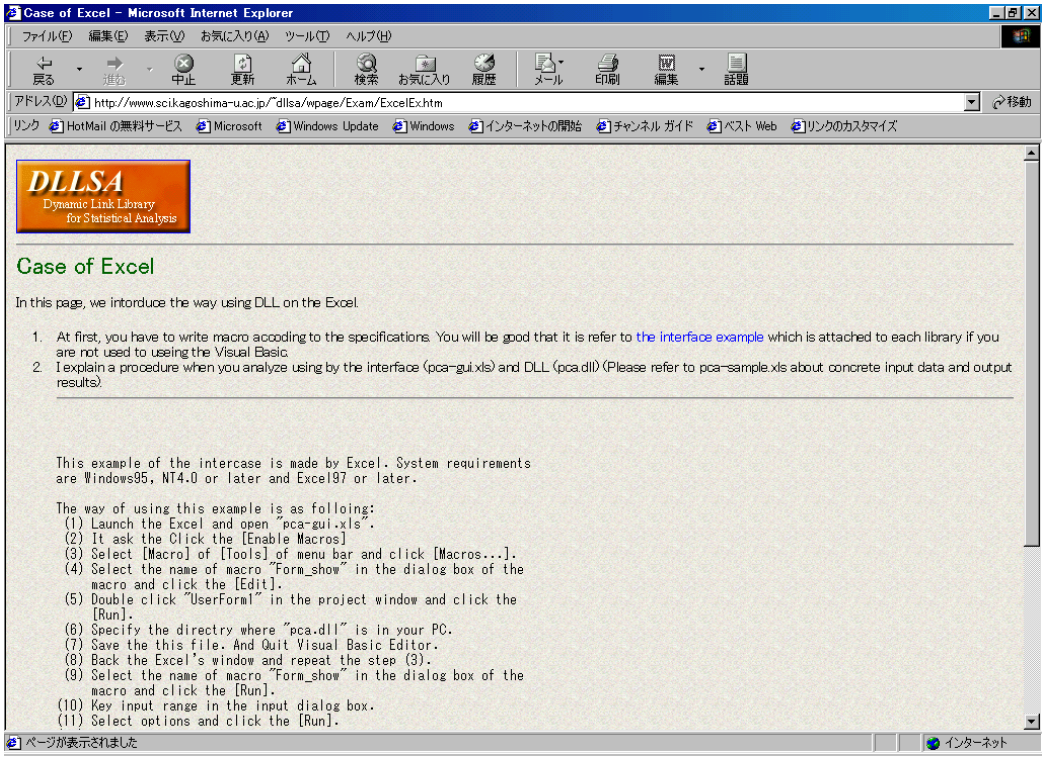

Figure 7 Linkage with MS-Excel

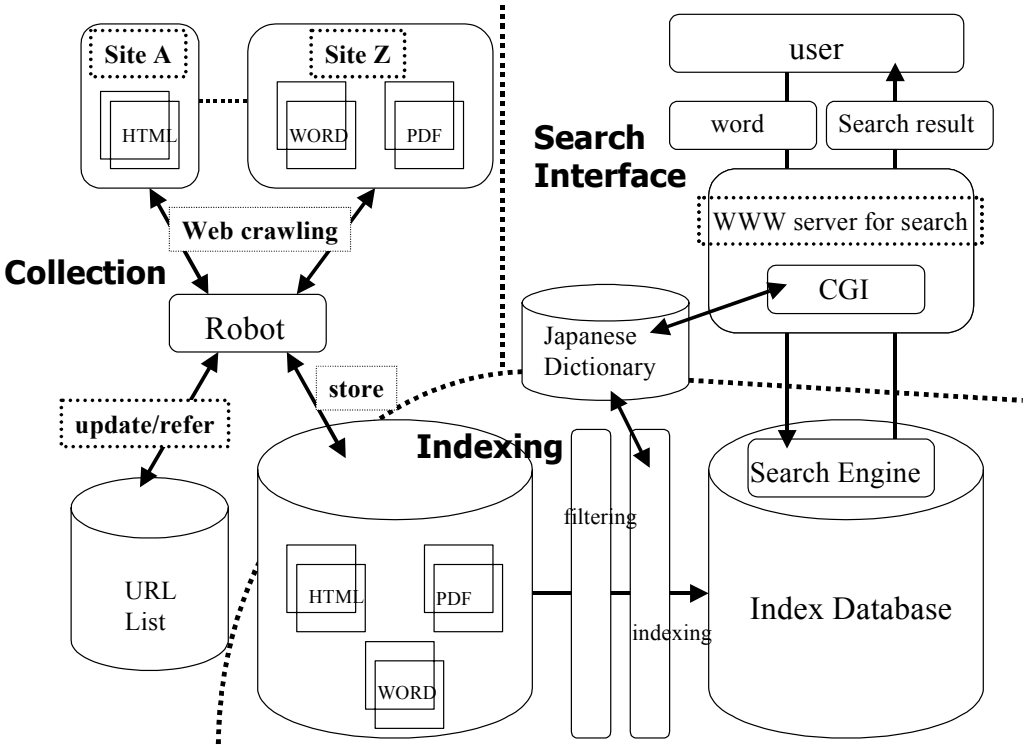

Figure 8 Image of Searching System

5. Conclusion

This paper has covered an introductory outline of the integrated web site including learning services such as online text for statistics education, electronic book reading systems, searching systems for great books and software libraries all which are currently under development. Almost all social science sections have strong needs for statistical data analysis skills with which many people can have the ability of extracting the essence from each data. Statistics is basically inter-disciplinary study applied to positive analysis in various departments. Looking forward to the expecting rapid and huge progress of Information Technology, it is important for us to keep up with those new technical changes, which will bring us another side of educational opportunities. As the data come from the users' answer to the questionnaire sheet, it is expected to be analyzed precisely. For beginners' basic and practical education, the system is highly required to the whole integration of interactive textbook contents, basic software libraries, widely available analytic software libraries, educational database applied to various kinds of fields, search engine and interactive question and answer systems. Text contents include dynamic resources developed by Java applets or other new technology suitable for WWW. Not only text but other media, such as images, computer graphics or sounds will be included in order to enable users to understand and get statistical sense practically and intuitively. Contents should be newer with current knowledge or technology and the additional function setting for the users' convenience. For example a nice interface for popular spread sheet software is welcome to the majority of computer use.

Although each web site introduced here is under construction for integration, our cooperation will be helpful towards educational significance. Some trial for international joint research with German faculties who had already published CD of statistical education with which students can learn statistics and econometrics dynamically, the same as on the Internet web site. If the inter-university computer network system is more improved, sharing digitized educational data sources will much be needed over various educational departments. We hope offering Japanese web content for statistical education could be of some help to dissemination of statistics. The integration of our independent project will have brought unexpected change and effective product to the statistical education field. The website mentioned here is operated on two servers. Each URL is as follows.

http://stst.eco.toyo.ac.jp/~stat/ (Toyo University)
http://www.sci.kagoshima-u.ac.jp/~stat/ (Kagoshima University)

The increase of mirror sites is our urgent duty.

At the end I introduce an example of integration over each project. Figure 9 shows the top page of the integrated searching system, which is among our independent projects. Since contents are accumulated second by second through network, it is indispensable for us to find out the exact information quickly. The project has offered the integrated searching engine service on the web. All users are able to meet his/her wanted pages under each system as soon as possible with the keyword/keyword-list using this engine. While we have been keeping up with updating the contents, it is needed for calculating the effectiveness of the project.

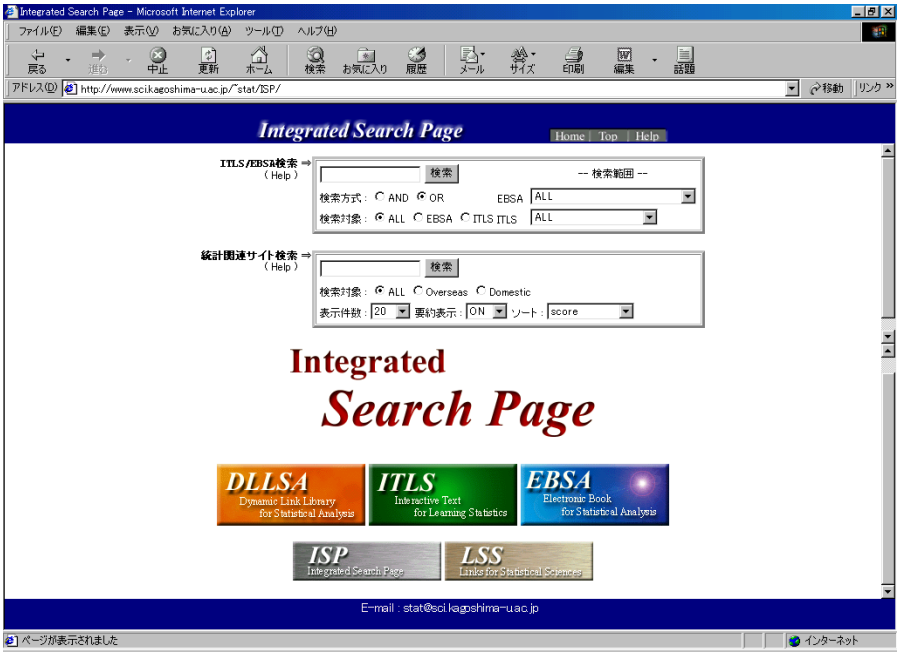

Figure 9 Integration of the projects

Acknowledgements

We thank professor emeritus Chooichiro Asano of Kyushu University for giving us important leadership on the point of making online learning systems of statistics who has many years' experience in teaching statistics and developing software for statistical science.

References

[1] Mittag, H. J. (2000) : Multimedia and Multimedia Databases for Teaching Statistics, CD-ROM version of Proceedings of ICME-9, Makuhari, JAPAN.

[2] Mittag, H. J. http://www.fernuni-hagen.de/STATISTIK/

[3] Yamada, S. (2000) : Report on statistical web sites made by private companies(in Japanese). Proceedings of the 68th annual meeting of Japan Statistical Society 228-229

[4] Sakurai, N. et al.(1999) : Developing a Textbook for Learning Statistics on the Internet, Bulletin of the International Statistical Institute, 52nd Session Contributed Papers, **3**, pp481-482, 1999