PRAXITELE:The New Generation of Data Analysis Tools

Karakos Alexandros

Democritus University of Thrace- Department of Electrical and Computer Engineering Section of software and application development
University Campus GR671 00 Xanthi, Greece

Abstract   The purpose of this paper is to present a new approach into data analysis area and to investigate on the next generation of data analysis tools. Multivariate Data Analysis refers to any statistical technique used to analyze data that arises from more than one variable. When available information is stored in database tables containing records and variables (rows and columns), Multivariate Analysis can be used to process the information in a meaningful fashion. Ajax is a new implementation of established web development technologies to gain interactivity between users and servers through multiple client sided server connections in the background. Ajax is a kind of that leads Web programming into the direction of Web 2.0, the new generation of using and developing the internet. With Ajax, Web applications finally start feeling like desktop applications to the Internet users. The Ajax programming techniques in the field of data analysis can to investigate on the next generation of data analysis tools. As example, to demonstrate our approach to the new generation of data analysis tools, we present PRAXITELE, an online service for applications of data analysis. You can find PRAXITELE at: http://peramos.ee.duth.gr/dataAnalysis2/dataAnalysis/praxitelis/English/index.html This web service provides to the users a friendly environment, where they can manage their data and apply on line different methods of Multivariate Data Analysis.

Keywords   Data Analysis, Internet, Ajax, Web 2.0, Web Services

1. Introduction

Multivariate Data Analysis refers to any statistical technique used to analyze data that arises from more than one variable. This essentially models reality where each situation, product, or decision involves more than a single variable. The information age has resulted in masses of data in every field. Despite the quantum of data available, the ability to obtain a clear picture of what is going on and make intelligent decisions is a challenge. When available information is stored in database tables containing records and variables (rows and columns), Multivariate Analysis can be used to process the information in a meaningful fashion[1].

The important growth in the use of Qualitative Data Analysis Software (QDAS) in the qualitative methodological design process is changing how researchers approach analysis. Qualitative researchers are progressively expanding the adoption of QDAS as a tool in the interpretation and analysis stages, and the increasing application has been cited as a major contribution to the rigor and credibility of qualitative research. There has been little systematic discussion, though, of various QDAS functions relevant to educational research. Moreover, software use has also raised concerns that the tools increasingly drive methodological practices. Qualitative data collection, analysis, and reporting require consistent, diligent attention in order to ensure a rigorous study.

Data analysis tools and data analysis software are typically used to sort through enterprise data in order to identify patterns and establish relationships. Similar to "data mining," data analysis techniques are useful for virtually any business to gain greater insight into the trends within their business, their industry, and their customer base.

Computer Assisted/Aided Qualitative Data AnalysiS (CAQDAS) is the use of computer software to aid qualitative research such as transcription analysis, coding and text interpretation, recursive abstraction, content analysis, discourse analysis, grounded theory methodology, etc[2]. CAQDAS is used in psychology, marketing research, ethnography, and other social sciences. A CAQDAS Networking project program[3], established in 1994, should have:

- Content searching tools
- Coding tools
- Linking tools
- Mapping or networking tools
- Query tools
- Writing and annotation tools

Software which falls under the CAQDAS 'umbrella' includes a wide range of packages but its general principles are concerned with taking a qualitative approach to qualitative data. A qualitative approach is one where there is a need to interpret data through the identification and possibly coding of themes, concepts, processes, contexts,

---

etc., in order to build explanations or theories or to test or enlarge on a theory. Many approaches (including, for example, participant observation, action research, grounded theory, conversation analysis etc.) broadly make use of qualitative perspectives, though each may emphasise particular processes or differ in their sequence. Qualitative researchers may also use quantitative data, for instance, relating to the descriptive features of research participants – sex, age and so on – to help in the organisation of qualitative data. This assists in cross-referencing and comparing qualitative ideas across groups and subsets. This approach remains distinct from 'content analysis' methodology, in which it is the statistics of word or phrase frequencies and their occurrence relative to other words or phrases across a textual dataset that are the basis of the analytic work[4].

PRAXITELE is an easy to use Internet-based data analysis software package. The software runs inside Internet Explorer, Netscape and all others browsers. This means that anyone with a web browser, who is given authorization to use your information, can access it and perform analysis[5]. PRAXITELE can have unlimited users and it requires no training for the average novice user. This, in turn, enables rapid deployment of important information analysis. For the various experience levels among users, it offers different interfaces with diverse functions. Reports can be presented in various formats, from either a simple, two-dimensional, black and white table to a more in-depth, colored graph. These reports can then be exported to Word, Excel, PDF and others, with the click of a button.

For more information about PRAXITELE visit the new web site at:

http://peramos.ee.duth.gr/dataAnalysis2/dataAnalysis/praxitelis/English/index.html

2. Web Services

The World Wide Web consortiums (W3C) define a Web Service as follows:

A Web service is a software application identified by a URI, whose interfaces and binding are capable of being defined, described and discovered by XML artifacts and supports direct interactions with other software applications using XML based messages via Internet-based protocols[6].

Web Service is a software system designed to support interoperable machine to machine interaction over a network (Figure 1). Web services are frequently just API's that can be accessed over a network, and executed on a remote system hosting the requested services[7].

Web Service take many of the ideas and principles of the Web and apply them to computer/computer interactions. Like the World Wide Web, Web services communicate using a set of foundation protocols that share a common architecture and are meant to be realized in a variety of independently developed and deployed systems. Like the World Wide Web, Web services protocols owe much to the text-based heritage of the Internet and are designed to layer

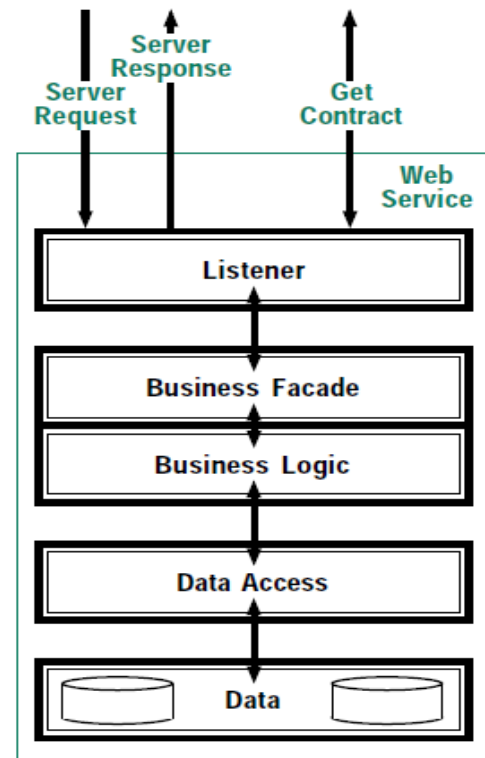as cleanly as possible without undue dependencies within the protocol stack[8].

**Figure 1.** Generic Web Services architecture (*Kirtland,* 2001)

An important area in which Web services differ from the World Wide Web is scope. HTTP and HTML were designed around "read-mostly" interactive browsing of content that is often static, or at least highly cacheable. In contrast, the Web services architecture is designed for highly dynamic program-to-program interactions. In the Web services architecture, many kinds of distributed systems may be implemented. Examples include synchronous and asynchronous messaging systems, distributed computational clusters, mobile-networked systems, grid systems, and peer-to-peer environments. The broad spectrum of requirements in program-to-program interactions forces the Web services protocol stack to be much more general purpose than the first Web protocols[8].

A Web Service is a unit of application logic providing data and services to other applications. Applications access Web Services via ubiquitous Web protocols and data formats such as HTTP, XML and SOAP, with no need to worry about how each Web Service is implemented[9].

2.1. Dynamic Web Applications

A new technique for developing dynamic web applications has emerged recently, called AJAX (sort of Asynchronous JavaScript Technology and XML), which is based on well-established technologies. While in a traditional web application the entire page is reloaded for each request, which results in the user waiting for server response, with the AJAX technique the server

communication happens in the background, and the response returned to the user's browser contains only a small amount of data, which is used to update the page without reloading[10].
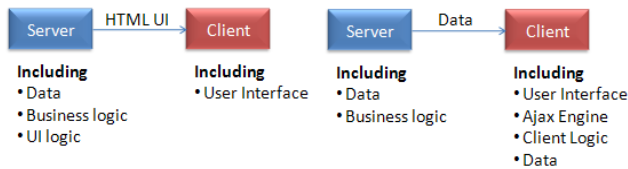
Figure 2. The architecture of a HTML page server and an AJAX server

The AJAX model of communication brings the desktop type functionality to the web and raises the user's expectations from the web[11]. This pushes developers around the world to adopt the AJAX approach and include it in their tools and frameworks[12].

With AJAX, software developers write JavaScript to handle the user interface (UI) events of an HTML component (e.g. when a character is typed in a text box). The JavaScript handler can perform dynamic data validation by sending an XMLHttpRequest to the server[13]. A JavaScript callback function is assigned to process the server response and render the validation results on the web page asynchronously. This approach reduces frequent page reloads in web applications and largely improves their interactivity[14].

Despite the fact that the contribution of JavaScript and XML has been momentous in the development of modern web pages, the new features introduced in web designing consequently presented several impediments in the total quality of services. The contemporary internet user has been able to navigate to pages with diverse content, though every time he wanted to interact with the page content, the page had to be reloaded to take the user's request into consideration. As a result, the user had to abide with unwanted delays, a quite frustrating reality particularly in case the interactive content finally proved to be ineffectual. Especially without persistent connections, each HTTP file request yields separate TCP connection to the server which accordingly requires its own TCP handshake, a time consuming process especially on high-latency links. AJAX came as a viable solution to these problems allowing the update of web pages "on the fly" eliminating such delays [15].

The most obvious feature of AJAX is the transmutation of Internet applications into desktop applications. The client can view or even interact with a webpage without a need to communicate with a server but only when to retrieve data. As a result most AJAX applications can be run mainly on the client. The information flow in AJAX starts from the client's interaction which is translated into an event. The required data located in the server are been retrieved and finally transferred to the client for presentation.

The main difference between the information flow using the HTML UI and AJAX is detected in the very first step, since in AJAX communication with the server commences only on event of an interaction with data which are stored in the server and not with data which are been stored in the client.

In Figure 2, the differences in the architecture of a HTML page server and an AJAX server are been illustrated and in Figure 3, is illustrated the classic web application model (synchronous) and the Ajax web application model (asynchronous)[16].
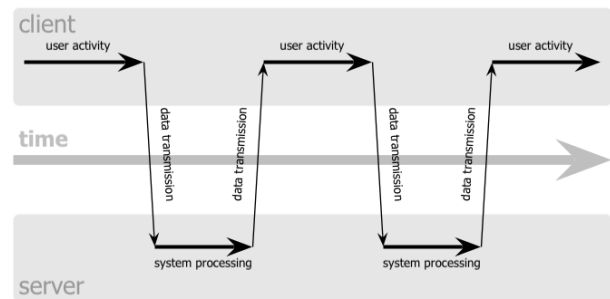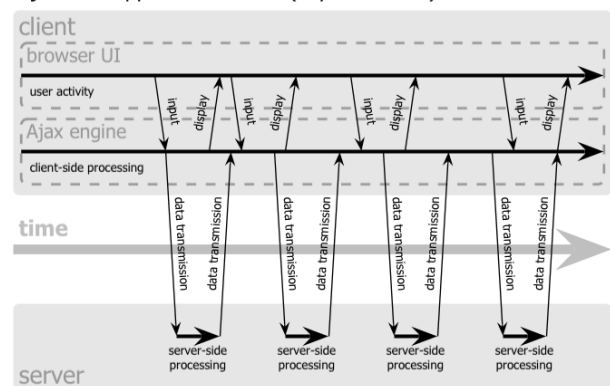
Figure 3. The classic web application model (synchronous) and the Ajax web application model (asynchronous)

3. Data Analysis Software and Tools

The arrival of the first mainframe software and then their PC-based successors has revolutionized multivariate statistical analysis. No longer are analyses limited in size or complexity and the full advantages of these techniques can be utilized. However, availability of this software was quite limited and even the available software often only performed basic analyses. The emergence of internet tools made these techniques available to everyone.

A typology suggested by Miles and Weitzman concern the handling and management of qualitative (and at that time, mostly textual) data. In[17] Weitzman and Miles provide a critical, in-depth look at 24 separate applications.

Actually, there are mainly three categories of multivariate statistical packages[18]:

Complete Multivariate Packages, contain all of the multivariate techniques and more as **Minitab**, perhaps the most "educator-friendly" software package, **SAS** (http://www.sas.com/) the most widely known statistical packages

(along with SPSS), Statgraphics (http://www.statgraphics. com/), Statistica(http://www.statsoft.com/) etc.

Basic Multivariate Packages, contain the most common multivariate techniques, but usually not all of them as **Stata** (http://www.stata.com/) the more oriented toward life sciences and **ViSta** (http://forrest.psych.unc.edu/research/index.html) developed by Forrest Young.

Specialized Statistical Methods, address a particular statistical analysis (e.g., SEM or time series) in more detail than the more comprehensive packages. These programs play an important role in augmenting the more comprehensive packages and providing access to emerging techniques before they are incorporated into the more general programs as LISREL (http://www.ssicentral.com/) the most widely used structural equation modelling (SEM) program that now includes statistical methods for complex survey data.

Organizations and individuals have taken advantage of the Internet's ease of access and availability to provide and disseminate information on many topics of interest to statisticians and educators. These resource sites are an invaluable source of material that can be used at all levels of statistical instruction, and range from the simplest concepts to the most complex.

Listed below are a sampling of the sites which provide a wide range of coverage both in types of information and the perspectives taken.

StatLib, the preeminent Internet resource for statistical information

UCLA Statistical Archive, maintained by the Department of Statistics at UCLA,

The World-Wide Web Virtual Library: Statistics maintained by the Department of Statistics at the University of Florida.

A new service, the **StatFactory** (http://statfactory.co.uk/) present in beta version, **StatWorker**, a web based statistical package used advanced web technologies (ASP) to make the user interface very intuitive and interactive. Active Server Pages (ASP), also known as *Classic ASP* or *ASP Classic*, was Microsoft's first server-side script engine for dynamically-generated web pages. To start using StatWorker, this Online Statistical Analysis Tools, it must register to get 3 months free access.

Today there are additional platforms for web application development that are more common on other operating systems. Both JavaServer Pages and PHP are more commonly found on web servers running non-Microsoft operating systems.

The purpose of this paper is to present a new approach into data analysis area and to investigate on the next generation of data analysis tools because there is not programs and packages in data analysis area available for free public access by World-Wide Web (Explorer, Mozilla or other browsers). This web service provides to the users a friendly environment, where they can manage their data and apply on line different methods of Multivariate Data Analysis.

4. Praxitele

In a distributed development environment, the display and analysis of project data are complicated by heterogeneous environments[19].

The applications combine sources files, executables files, graphics, HTML and Ajax scripts. Ajax is a web development technique for building responsive web applications. With Ajax, Web applications finally start feeling like desktop applications to the Internet users.

The programming of each such application consists of several parts:

1. Development of HTML files with the Ajax technologies, enabling the transfer of input data and parameters of programmed procedure. As a rule, it is through a frame on client's display.

2. Development of files, which, except that it solves required computational or graphical problem, reads the data from the input HTML frame, and prepares the output field.

3. Then, other HTML procedure transfers the output to the output page or window visible on client's display. The graphs are in .png format.

The free and in real time web service for data analysis on the Internet, without the need of purchasing the software is inadequate. The aim of this paper is the construction and presentation of a web page, where the user can apply input and manipulation techniques of data and various methods of data analysis, out of charge and no matter the operating system of his PC.

Therefore an attempt has been made to separate the components of the application which can be executed independently and locally at the client-side, from those which need to be processed from the server. To do this the programming language PHP was selected, to write the server-side scripts, and AJAX technology for the writing of the client-side scripts.

The above method of implementation was decided as very fast and effective and therefore got selected as the basic platform of the new implementation of the application. Great attention was given to the graphic presentation of the results. The application is available in English and Greek language widening the range of users of the application.

JpGraph library, was used to develop the graphics, which is also written in PHP. The library is freeware and is available on the internet at http://www.aditus.nu/jpgraph/. In the manual of JpGraph library there are numerous examples and detailed list with the classes and methods supported by the library. JpGraph is a continuously developing library so it can utilize the updates of PHP language.

To create the environment of the server the WampServer package was used, which includes Apache, PHP and MySQL database. It also includes PHPMyAdmin and SQ Lite Manager for easier management of the database. WampServer is an open source project, free to use (GPL licence).

The last distribution used was WampServer 2.0i which includes PHP 5.3.13, Apache 2.2.22 and MySQL 5.5.24.

WAMP package can be obtained free of charge at http://www.wampServer.com   where it also offers technical support from other users who use it.

4.1. The Application

PRAXITELE, is a web-based application, written using JavaScript, DHTML and PHP and facilitates team collaboration and data sharing.

There are three reasons for adopting this model rather than a peer-to-peer or Java-based system.

Firstly, the service should be available to as many people as possible and a pre-requisite for a Java-compatible browser may limit the potential user base. Furthermore, an increasing number of industrial companies are restricting access to the Internet and specifically not allowing peer-to-peer operations for security reasons.

Secondly, many of the routines which are published on the server have significant processing overheads, and would be severely restricted in performance under Windows and Macintosh environments and under the interpreted Java language.

Finally, the server is at the first stage of a projected development to a fully integrated and distributed data analysis resource.

Basic capabilities of the application:

● upload or create data in text delimited (comma, tab, semicolon or other defined by the user delimited) or CSV format. A **comma-separated values (CSV)** file is a common and simple file format, CSV files are often used for moving tabular data between two different computer programs, for example between a database program and a spreadsheet program.

● manipulate and organize data in data tables and storage it locally on the user PC

● produce, customize, print and export high quality statistical plots

● use statistical procedures including descriptive statistics, and multivariate data analysis

● the orientation in the system is provided by the menu of themes and sub-themes

The capture by print screen technique of the main page is shown on Figure 4.

The user can either select its own sequence or he can follow the recommended sequence (which actually corresponds to the order of menu bars).

A typical page is shown on Figure 5, with a table of data and the menu bars. On Figure 6, we demonstrates the check of values of the table given and on Figure 7, we demonstrates the dendrogram of classification (cluster analysis) of a data set.
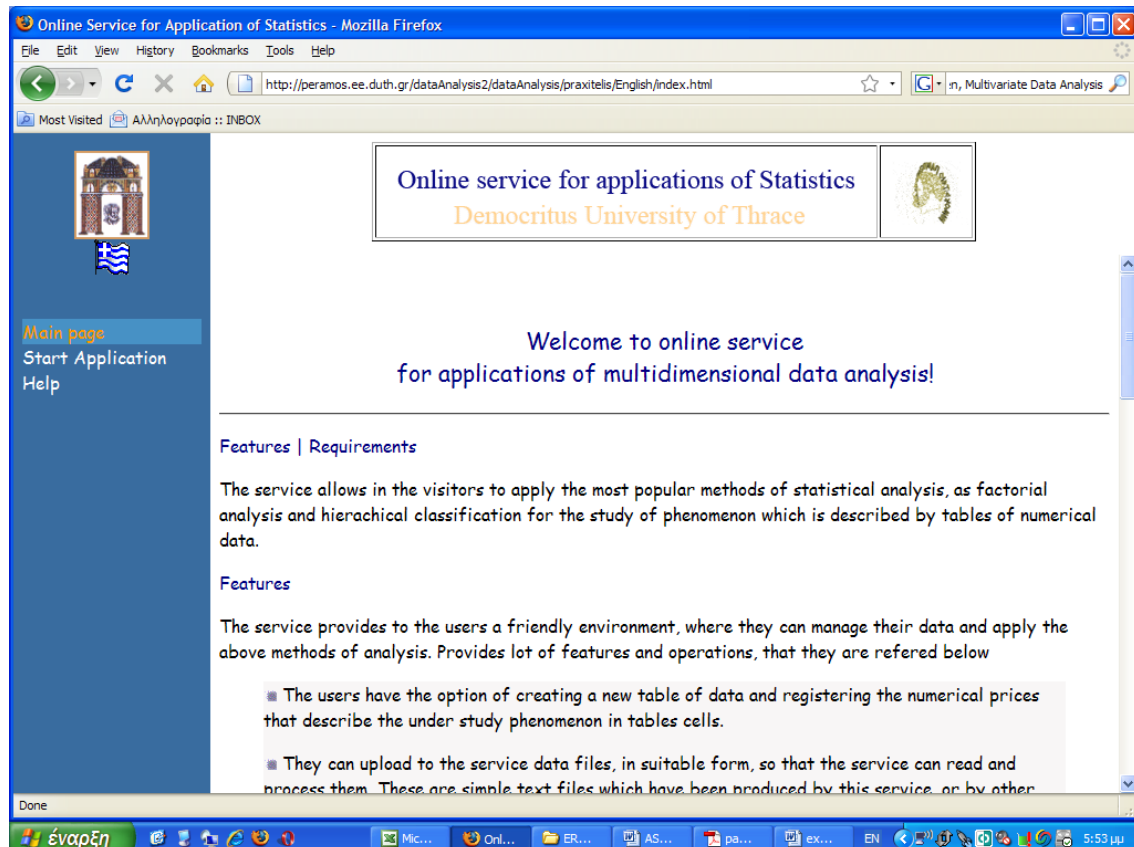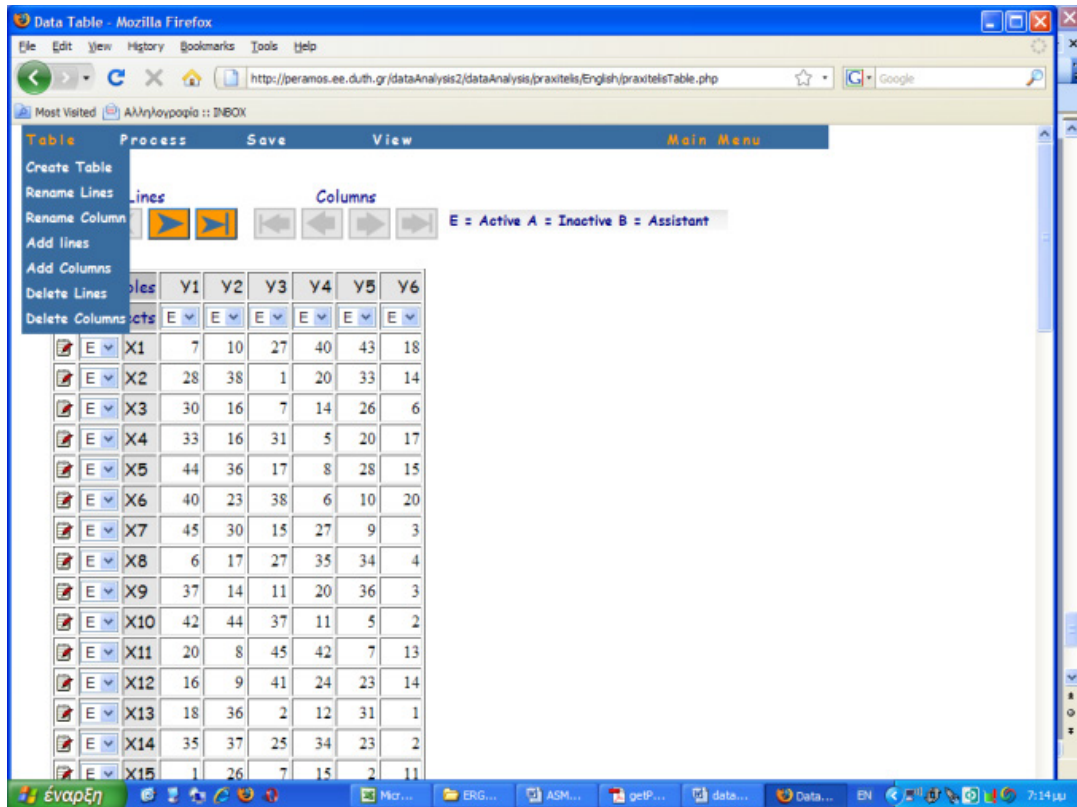

Figure 4. The main panel

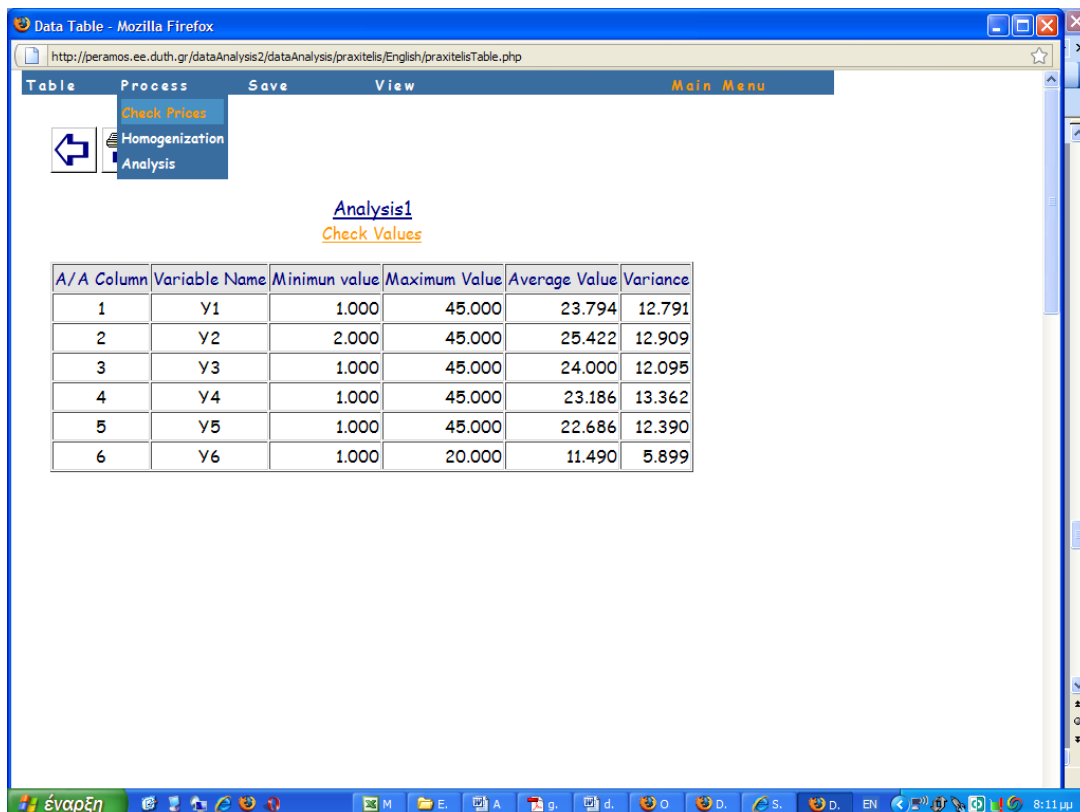**Figure 5.** A typical page of the program
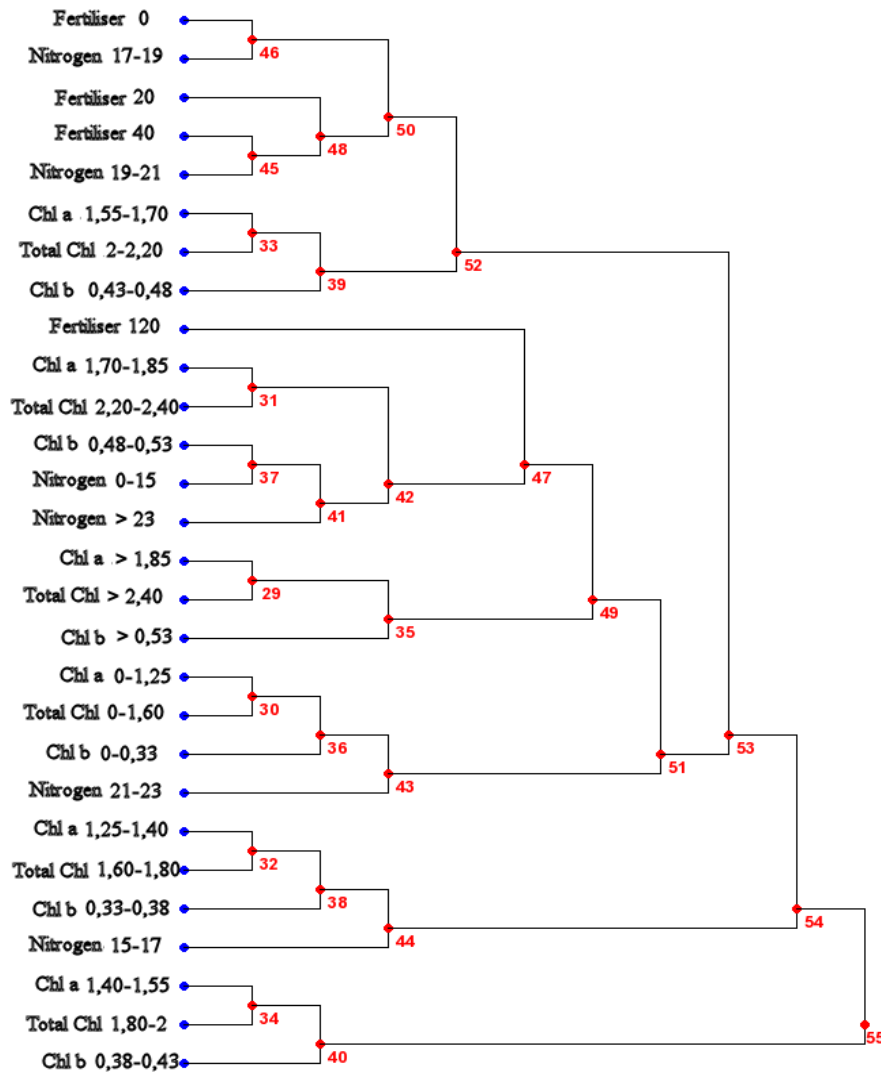

Figure 6. Check values results

Figure 7. The dendrogram of classification

5. Conclusions

Data analysis tools and data analysis software are typically used to sort data in order to identify patterns and establish relationships.

In times where web technologies advance rapidly and following the trend of reusing technologies which have been successfully employed in the past, AJAX is an effective alternative to Rich Internet Applications. Despite its undoubted advantages, AJAX has significant weaknesses which should be eventually removed if this novel technology wishes to be established in the future.

Furthermore the amount and the variety of AJAX applications appearing in the World Wide Web is an inspiring sign of AJAX's potential as well as an encouraging fact since web developers are already attracted to AJAX. However it is most uncertain whether AJAX will manage to develop into a dominant web design trend or will be casted out into oblivion in favor of other fresher technologies.

To demonstrate our approach to the new generation of data analysis tools, we present PRAXITELE, an online service for applications of data analysis. PRAXITELE is an easy to use Internet-based data analysis software package. The software runs inside Internet Explorer, Netscape and all others browsers. This means that anyone with a web browser, who is given authorization to use your information, can access it and perform analysis. The original tool was developed around 1988 and we have been testing the current scripting language package since 2006. Although our initial results are encouraging, we must study additional methods to fully evaluate the system and determine its effectiveness.

On Figure 8, we demonstrate the statistical resources actually available in PRAXITELE. This service provides to the users a friendly environment, where they can manage their data and apply on line different methods of Multivariate Data Analysis for free.

You can find PRAXITELE at:

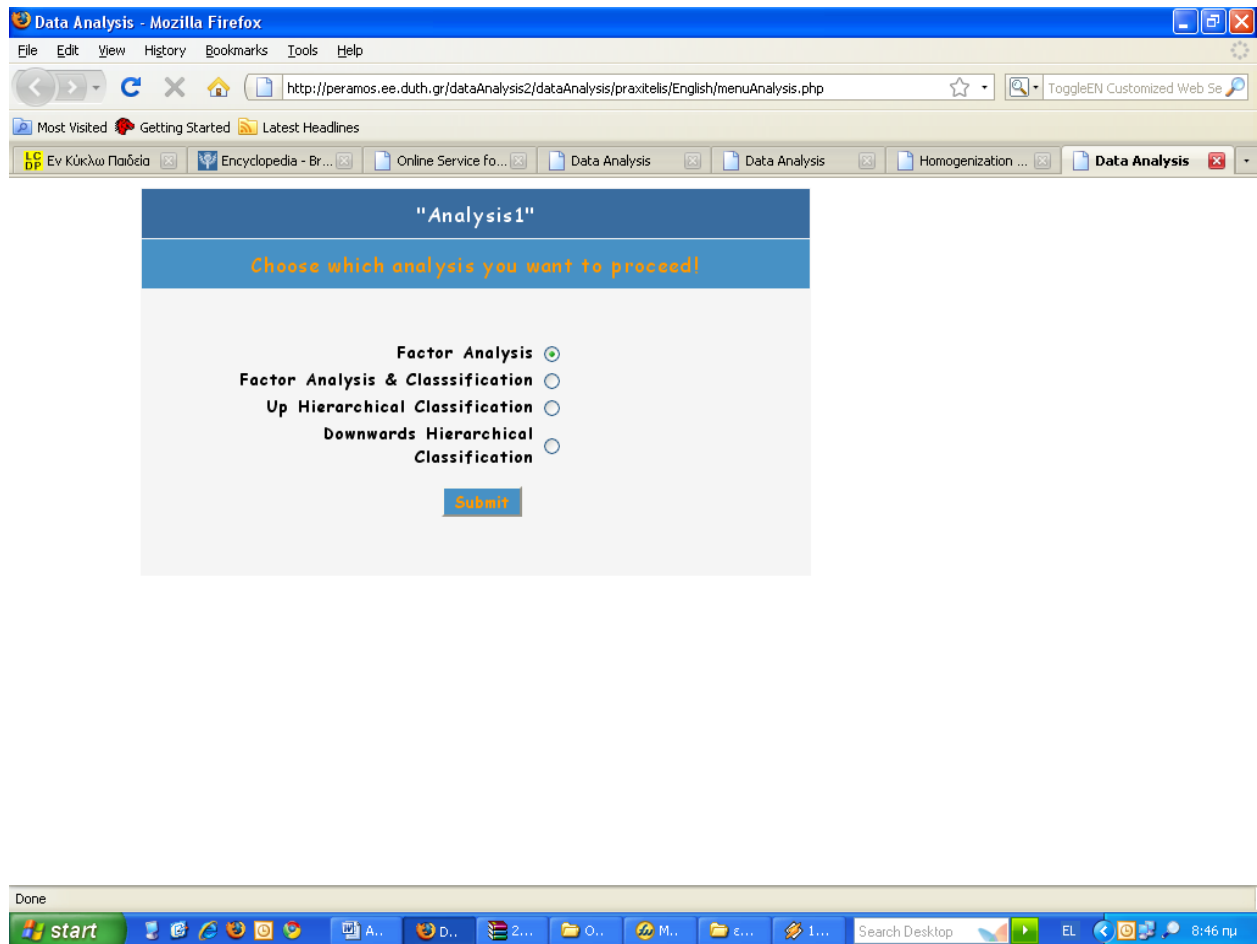http://peramos.ee.duth.gr/dataAnalysis2/dataAnalysis/praxitelis/English/index.html.

Figure 8. The statistical resources available in PRAXITELE

REFERENCES

[1] Esbensen Kim H., (2002). *Multivariate Data Analysis in practice (5th Edition)* Aalborg University, Esbjerg, Denmark

[2] http://en.wikipedia.org/wiki/Computer_assisted_qualitative_ data_analysis_software Accessed 20 July 2012

[3] http://www.surrey.ac.uk/sociology/research/researchcentres/ caqdas/about/index.htm   Accessed 20 July 2012

[4] Ann Lewins & Christina Silver (2004) *Choosing a CAQDAS Package* http://cue.berkeley.edu/qdaarticle.pdf   Accessed 20 July 2012

[5] http://www.ehow.com/internet-history/ Accessed 20 July 2012

[6] Austin, D., Brbir, A. and Garg, S. (2002), "*Web Services* architecture requirements", W3C, available at: www.w3.org/ TR/2002/WD-wsa-reqs-20020429, Accessed 18 July 2012

[7] Kirtland Mary. (2001). A Platform for Web Services. Technical Articles available at: http://cloverlink.com/whiteP apers/A_platform_for_web_services.pdf,   Accessed 20 July 2012

[8] Cabrera Luis Felipe, Kurt Christopher, Box Don. (2004). *An Introduction to the Web Services Architecture and Its*

Specifications. Web Services Technical Articles available at: http://msdn.microsoft.com/en-us/library/ms996441.aspx, Accessed 20 July 2012

[9] http://msdn.microsoft.com/en-us/library/ms950421.aspx Accessed 20 July 2012

[10] Draganova C. (2007). Asynchronous JavaScript Technology and XML (AJAX), *ITiCSE 2007, 12th ACM Annual Conference on Innovation & Technology in Computer Science Education,* Dundee International Journal of Science and Technology.

[11] Cheung David W.L, Thomas Y.T. Lee and Patric K.C. Yee, (2007). *Webformer: A Rapid Application Development Toolkit for Writing Ajax Web Form Applications* Springer Berlin/Heidenlberg.

[12] http://ajaxpatterns.org/ Accessed 20 July 2012

[13] Fangxiao Wang, Xiaoyi Zhang, and Hao Yan. Research and Application of a Web Data Validation Model Based on Asynchronous Communication. *Proceedings of the 2009 International Symposium on Web Information Systems and Applications (WISA'09) Nanchang, P. R. China, May 22-24, 2009, pp. 194-197*

[14] Paulson, L.D. (2005). Building rich applications with AJAX, *Computer, 38, 10*(Oct. 2005), pp 14-17.

[15] G. Lawton, "New Ways to Build Rich Internet Applications", *Computer, Vol. 41, No.8, pp. 10-12*, 2008.

[16] http://www.javalobby.org/articles/ajax/   Accessed 20 July 2012

[17] Eben A. Weitzman and Matthew B. Miles. (1995) *Computer Programs for Qualitative Data Analysis*, *SAGE Publications, Inc*

[18] Hair J., Black W.,Babin B., Anderson R. (2009). *Multivariate*

Data Analysis (7th Edition) Prentice Hall.

[19] Wikan Danar Sunindyo, Thomas Moser, Dietmar Winkler, Stefan Biffl (2010) Foundations for Event-Based Process Analysis in Heterogeneous Software Engineering Environments. Proceeding SEAA '10 Proceedings of the 2010 36th EUROMICRO Conference on Software Engineering and Advanced Applications Pages 313-322