國立交通大學

多媒體工程研究所

碩 士 論 文

設計與實作智慧型問卷分析輔助系統

Design and Implementation of an Intelligent Questionnaire Analysis Assistant System

研 究 生：蔡雨杰
指導教授：曾憲雄 博士

中 華 民 國 九 十 六 年 六 月

設 計 與 實 作 智 慧 型 問 卷 分 析 輔 助 系 統

Design and Implementation of an Intelligent Questionnaire Analysis Assistant System

研 究 生：蔡雨杰　Student：Yu-Jie Tsai

指導教授：曾憲雄 博士　Advisor：Dr. Shian-Shyong Tseng

國 立 交 通 大 學

多 媒 體 工 程 研 究 所

碩 士 論 文

A Thesis

Submitted to Institute of Multimedia Engineering

College of Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Computer Science

June 2007

Hsinchu, Taiwan, Republic of China

中華民國九十六年六月

設計與實作智慧型問卷分析輔助系統

研究生：蔡雨杰　　　　　　　　　　指導教授：曾憲雄博士

國立交通大學多媒體工程研究所

摘要

當社會科學研究人員分析問卷時，常需要使用一些統計輔助工具，例如SPSS，來協助分析，然而現今這些工具只能給予使用者統計方法運算上的輔助，並沒有提供使用者分析問卷時的決策輔助，例如驗證假設的統計方法的選擇。

在這個研究中，藉由之前我們提出的輔助問卷分析相關的研究。我們設計並且實作一個智慧型問卷分析輔助系統，我們制定了系統設計相關的圖表，讓系統將來能夠方便於擴充以及維護。這個系統使用三種指標器來計算統計上的顯著性差異的程度以及規則來決定哪個指標器有較高的程度有統計上的顯著性差異。我們基於指標器和規則來建立了顯著性差異觀察器。系統也提供了建議關於合適的統計方法來驗證假設以及提供解釋給使用者。系統亦提供了學習教材給使用者學習這些建議的方法。

我們實作了這個系統的原型，並且做了一個個案研究以及使用者滿意度調查的實驗。實驗的結果顯示系統分析出來的結果幾乎正確，而使用者對於系統的概念很感興趣。我們在未來會增加更多有用的功能到系統中。

關鍵字：問卷分析，決策支援系統，指示器，線上學習。

Design and Implementation of an Intelligent Questionnaire Analysis Assistant System

Student: Yu-Jie Tsai Advisor: Dr. Shian-Shyong Tseng

Institute of Multimedia Engineering
National Chiao Tung University

Abstract

When social science researchers analyze questionnaires, they often need to use some statistic-related tools, for example SPSS, to assist their analysis. However, nowadays these tools only support users the computation of statistic, they do not support the decision when users analyze their questionnaire. For example, the decision of statistical methods used to verify their hypotheses.

In this study, those our previous researches on assistant questionnaire analysis are used to design and implement an intelligent questionnaire analysis assistant system. We formulate the figures for system design to make our system easy to extent and maintain in the future. This system uses three kinds of indicators to compute the degree of statistically significant difference and metarules to determine which indicator having more degree of statistically significant difference. A Significant Difference Viewer is constructed based on the indicators and metarules. This system also provides suggestion for appropriate statistics methods to test hypotheses and gives explanations for users. This system gives the learning materials for users to learn these suggested methods.

The prototype of this system is implemented, and the experiments of a case study and the satisfaction of users are also done. The experiment's result showed the analysis results of the system were almost correct, and users were interested in the idea of this system and we will add more useful functions into the system in the future.

Keyword ： Questionnaire Analysis, Decision Support System, Indicator, E-Learning

致謝

首先，我要感謝我的指導教授，曾憲雄博士，老師非常有耐心，不厭其煩，一次又一次的討論，指導我撰寫這篇論文，讓我不只是完成這篇論文，而且學習到很多研究方法、思考方法、以至於表達方法，這是碩士班期間最寶貴的知識。感謝洪宗貝老師、楊錦潭老師與王景宏博士，在口試時給了許多寶貴的意見，讓這篇論文更完善，並且讓我了解到這個研究的價值，拓展了我的視野。

再來，非常感謝曲衍旭學長，和學長討論使我的思考與研究得到相當大的啟發，沒有學長，就不會有這篇論文。此外，瑞鋒學長和喚宇學長，在我寫作期間給了我許多寶貴的建議與討論。昂叡、信男、東權、芙民、曉涵、嘉妮、學長姐以及學弟妹，感謝大家在我的碩士生涯一起努力、互相扶持。

最後，我要感謝我的家人，讓我能夠專注於研究上，並且在遭遇困難時鼓勵我，讓我能夠有自信地完成這篇論文。

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

In the traditional quantitative research, the researchers will firstly propose several hypotheses of the subject according to their experiences, and then determine whether there exists a statistical significance in some hypothesis in a try-and-error manner. The quality of the result using the manner, of course, is in accordance with the quality of the hypotheses made by the researchers, and the process of finding the statistically significant differences is highly dependent on researchers' intuition and experience. For example, in a questionnaire survey of elementary school students' Internet usage behavior, a researcher might make a hypothesis, "There is a difference between different genders about the hours they access the Internet every week," and then use appropriate inferential statistics method according to his/her knowledge of statistics to test this hypothesis. Without making the hypothesis, the statistically significant difference can not be found even if it really exists. Therefore, how to acquire the knowledge and experience of senior researchers might be helpful for the junior researchers.

Besides, granularity of original questionnaire data may not be good enough to find the statistically significant differences. For example, in a questionnaire survey of elementary school students' Internet usage behavior, if there is no significant difference between different resident regions about the hours students access the Internet every week, the researcher might conclude there is no significant difference between different resident regions. However, each region may contain several counties in geography. By drilling down the student's resident dimension to lower levels of granularity, it is still possible to find a significant difference between different counties.

When researchers analyze a questionnaire data, they have to make some hypotheses of the possible statistically significant differences from the data and make some selections of appropriate inferential statistic methods to test their hypothesis according to their intuition, experience, and knowledge.

However, the statistic-related assistant tools nowadays are not intelligent. These tools do not support users assistants when they analyze questionnaire, for example give some advices for the selections of statistical methods. They only provide assistants on computations of statistical methods.

In our previous research, we have proposed some techniques on assistant questionnaire analysis. But these research are theoretical, we have not use these research on a real system for users to use. We want to use these research to design and implement an intelligent questionnaire analysis assistant system.

In order to assist junior researchers in analyzing questionnaire data, we acquire the knowledge and experience of senior researchers to construct a forward-chaining rule-base expert system. This expert system assists researches in making hypotheses of the possible statistically significant differences from the data and making selections of inferential statistic methods to test researchers' hypotheses. According to the experts' experiment and knowledge, three indicators, Increase, StepDown, and Dice, are designed to help researchers finding the possible statistically significant differences. The experts also define metarules to determine degree of indicators. Hence, a significant difference viewer is constructed based on the indicators and metarules in the expert system to assist junior researchers in exploring data. For the need of selecting appropriate inferential statistic methods to test researchers' hypothesis, the expert system is also designed to give suggestions for appropriate statistic methods to test hypotheses. The expert system gives explanations about why these methods are appropriate to analyze the data. Since the methods are suggested from the expert system, researchers may not necessarily understand what these methods are and how to use them. But researchers may want to know the detail information of the methods, for example, meaning of the method, how to use the method. The expert system is designed to provide a learning platform for junior researchers to learn these methods. Therefore, junior researchers can learn which methods are appropriate to use.

One of the issues in designing a real online system is how to design a system having good maintainability and extensibility. This is the most difficult part for system design and implementation. We need to design functions provided for users to assist them, design architecture of the system supports those functions, and formulate clearly and completely those diagrams about system design, for example, use case diagram. Besides, when implementing this system, we need to integrate and adjust those techniques used in the system to adapt for requirements of the system. This is also a difficult part in system implementation.

The rest of this thesis is organized as follows. In Chapter 2, we introduce some preliminaries about the techniques used in the system, data warehouse and OLAP, significant difference, indicator, DRAMA/NORM, Ontology-based Learning Sequences Construction Algorithm [8], Ontology-based Adaptive Learning Sequences Construction Algorithm [5], and the e-learning architecture proposed by Chang [5]. Chapter 3 presents the methods we used to design and implement the system and functions, some schemas we designed for the system. Chapter 4 shows the overall system architecture and describes the system in detail. Chapter 5 gives the results of the experiments we designed to evaluate the system. Finally, concluding remarks are given in Chapter 6.

Chapter 2. Preliminaries

For those books, magazines, and research we have studied, there is less research about questionnaire analysis assistant, except the research we proposed. Hence, we list and illustrate some techniques and research we used in the system

2.1. Data Warehouse and OLAP

The data warehouse could consist with several data cubes or single data cube. For each data cube, it has several records and a star schema to describe the schema of the structure of data cube. In other word, the star schema can describe the dimensions with concept hierarchy and some measures of the data cube. And, the data warehouse supports an analysis tool: On-Line Analytic Processing (OLAP) [6][25]. It is a useful tool assistant to user exploring the data cube. OLAP can organize and present data in various formats in order to accommodate the diverse needs of the different analysis approaches. OLAP server provides server operations for analyzing multidimensional data cube:

- *Roll-up*: the roll-up operation collapses the dimension hierarchy along a particular dimension(s) so as to present the remaining dimensions at a coarser level of granularity.
- *Drill-down*: in contrast, the drill-down function allows users to obtain a more detailed view of a given dimension.
- *Slice*: Here, the objective is to extract a slice of the original cube corresponding to a single value of a given dimension. No aggregation is required with option. Instead, server allows the user to focus on desired values.
- *Dice*: A related operation is the dice. In this case, users can define a sub cube of the original space. In other words, by specifying value ranger on one or more dimensions, the user can highlight meaningful blocks of

aggregated data.

- *Pivot*: the pivot is a simple but effective operation that allows OLAP users to visualize cube values in more natural and intuitive ways

However, the OLAP is a discovery-based analysis tool, and it can not detect the significant difference pattern automatically or semi-automatically.

2.2. Significant Difference

In the questionnaire analysis, finding whether there is significant difference between two or more groups in one measure is one of the major problems in research. For example, in a survey of junior high school students' current status, "Is there significant difference between different genders' IQ?" and "Is there significant difference between the mathematics grades of different areas in Taiwan?" are two interesting phenomenon that researchers want to know. [12] categorized the research questions into degree of relationship among variables, significance of group differences, prediction of group membership, and structure, which significance of group differences is used to find the significant difference. Therefore, finding possible significant difference between different groups is a very important research issue.

However, finding possible significant difference namely is difficult for social science researchers. In our observation, there are two main causes may lead to this issue.

The first cause is that researchers find the significant difference by their intuition and experience. For example, a junior researcher might consider that there is significant difference between different genders' IQ. She/He could make a hypothesis, "There will be difference between different genders' IQ," and then use inferential statistics method to test this hypothesis. This is basically a hypothesis-based search method. Once the hypotheses are not made the significant difference can not be found even if it really exists. However, senior researchers might find it easier because of their rich experiences.

The second cause is that the original questionnaire data may be not good enough to find the significant differences. For example, in the survey of junior high school students' current status, the student's resident dimension doesn't have granularity, and just contains the region attribute. If there is no significant difference between the mathematics grades of different regions, the researcher can just say there is no significant difference between the mathematics grades of different regions. However, if the researchers combine their collected data with secondary data which are collected by other researches [30] like the government official statistical data, geographic information, or other researches data, and assume the student's resident dimension has granularity, they would drill down the dimension to find whether there is significant difference between the mathematics grades of different cities.

Significant difference, a specific term in statistics, represents two or more groups exist obviously different on a continuous variable. In Luo's research [16], he defined a significant difference as a Significant Difference Pattern. This pattern is used to describe the significant difference of data. He also proposed an algorithm, WISDOM, to find out Significant Difference Pattern from analyzing data in the data warehouse.

2.3. Indicator

The data warehouse supports the analysis tool OLAP, and users can use some OLAP operations, like roll-up, drill-down, dice, etc., to explore the data cubes. However, the exploring process is not automated, and users still need to explore the data cube by her/his intuition and experience. Sunita Sarawagi [23] proposed a Discovery-driven Exploration of OLAP Data Cubes approach, which provides the following three kinds of precomputed indicators to assist users to explore the data cubes.

- *SelfExp*: This indicates the degree of surprise of the cell value, relative to other cells at the same level of aggregation.
- *InExp*: This indicates the degree of surprise somewhere beneath the cell, if we were to drill down from it.
- *PathExp*: This indicates the degree of surprise for each drill-down path from the cell.

However these indicators are lack of finding statistically significant difference and researchers can not easily explore original questionnaire data. Chu [9] interviewed the experts. They use the similar idea to precompute some indicators. These indicators generate metarules to assist users in making hypotheses of the possible statistically significant difference easily.

He first built a data warehouse, which has subject-oriented, integrated, time-variant, and nonvolatile features, by combining the questionnaire data and secondary data. Nowadays, researchers generally firstly collect the required data, and then find and learn the appropriate functions to explore and analyze questionnaire by utilizing database, excel, or SPSS softwares, but the process is very hard. Therefore, he applied data warehousing technology and used OLAP to explore the data online from various views. Although OLAP is easy to explore data, it's not good enough for finding statistically significant differences. Hence, a decision support system with three kinds of indicator is proposed by him to assist researchers explore the data cubes in data warehouse and find possible statistically significant differences.

- *Increase*: This shows the degree of statistically significant difference of changing the view by increasing an additional dimension.
- *StepDown*: This shows the degree of statistically significant difference of changing the view by stepping down this dimension.
- *Dice*: This shows the degree of statistically significant difference of changing the view by dicing for some value.

Furthermore, the experts define thresholds for these indicators. These thresholds are used the determined if the data has the possible statistically significant differences or not. If the degree of statistically significant difference is over the thresholds for the indicator, the system shows there may be possible statistically significant differences. In addition, the experts also define some rules to determine which indicator should be more degree having the possible statistically significant differences if there are two or more indicators over the thresholds. Therefore, the experts conclude these threshold and rules to metarules to assist users. There are some examples of metarules below.

IF degree of statistically significant difference in Increase >= 0.8 THEN

Indicator appears the signal which means there has statistically significant difference.

IF Increase and Dice appear the signal THEN Dice disappears the signal.

2.4. DRAMA/NORM

In traditional forward-chaining rule-base expert system, the rule base consists of all rules and facts. The system needs to go through every matching rule when conducting inference for the proper result. This might become inefficient when the number of rules and facts become large. Therefore, many researches aim to improve the maintenance of rule-based expert system by incorporating the objected-oriented approach.

We apply the DRAMA/NORM package for building up the expert system. DRAMA is a rule-based, client-server tool/environment for KBS development. It can assist knowledge engineers in building up an expert system. Briefly, DRAMA contains lots of innovative techniques including Object-Oriented technology, knowledge inheritance, etc. It also contains useful tools, like rule verification tool, knowledge acquisition assistant tool and the inference server. Using the client-server architecture of DRAMA, the knowledge base is maintained on a server and clients could access this server for inference services.

The kernel knowledge model of DRAMA, named NORM (New Object-Oriented Rule-base Model), was developed by the KDE Lab at Dept. of Computer & Information Science of National Chiao-Tung University. The working model of NORM, containing knowledge classes (KCs) and the relations between KCs, as shown in Figure 2-1, is based on the principles about how people ponder and learn to acquire knowledge. According to domain expertise, when a person is trying to learn something, there are often some topics for him/her to study. A lot of new knowledge is built upon the original knowledge according to the discipline of Educational Psychology. Thus, new knowledge about the topics could easily be built one by one after the person successfully studies them. And, these topics could be transformed to KCs easily. In other words, learning is an activity to construct the relations between different KCs. Since this knowledge model fits in quite well with the thought of human and KCs are modularized, we can build and maintain the knowledge base more conveniently. It is very important to use such knowledge model for the knowledge engineers. Whenever there is a need to update some knowledge, it is unnecessary to change all the knowledge base. All we have to do is just to add or modify the modules involved. In addition, the client-server architecture of DRAMA makes the web services plausible and more easily. Thus, the benefits of the expert system approach can be utilized throughout the Internet.

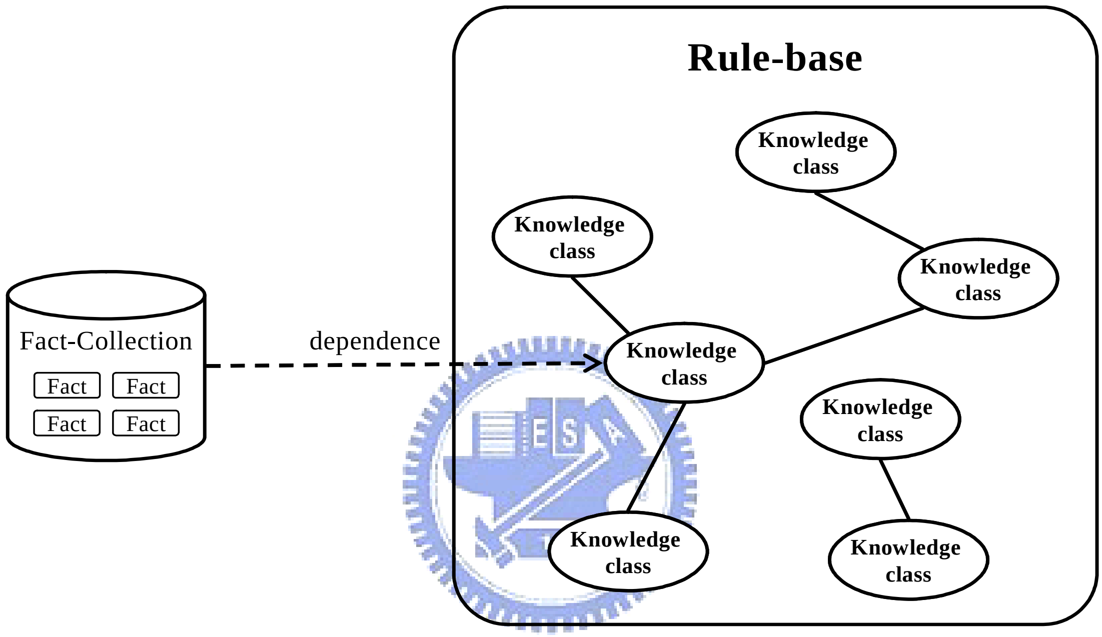


Figure 2-1: NORM-based knowledge base

2.5. Ontology-based Learning Sequences Construction

As shown in Figure 2-2, in [8], the Ontology-based Learning Sequences Construction Scheme is proposed in order to transform a domain ontology to a basic course scheme. There are three primary parts in this construction module:

- Transformations between domain ontology relations and learning sequences.

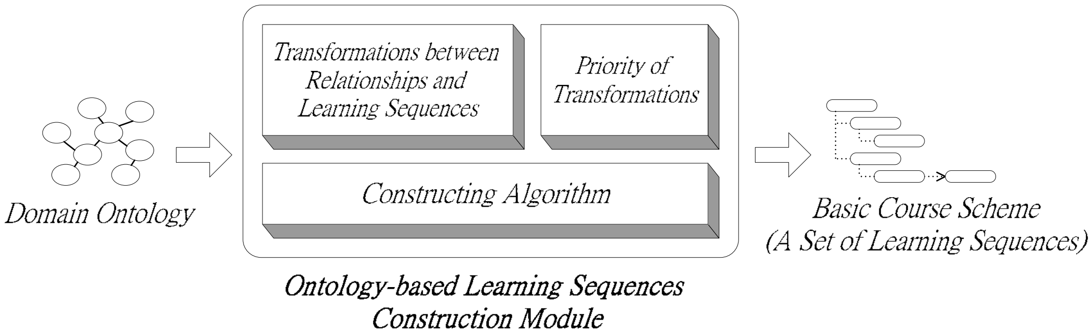


Figure 2-2: The Ontology-based Learning Sequences Construction Scheme

With the help of the transformation process, we could transform the ontology relationship into some kinds of learning sequence according to the properties of the relationship.

- Priority ordering definition among those transformations.

Usually, there are more than one kind of relations in an ontology and each relationship is supposed to correspond to a kind of transformation. Therefore, if we would like to transform a domain ontology into a basic course scheme, we have to decide the priority ordering definition among these transformations since there are so many co-existed relations. On the other hand, just like the definition of transformations, the priority ordering definition is also domain-dependent. Thus, it is supposed that the priority ordering is defined with the help of domain experts

- The ontology-based learning sequences constructing algorithm they posed.

Due to the complexity of this algorithm, please refer to [8] for more detail.

2.6. Ontology-based Adaptive Learning Sequences Construction

The Ontology-based adaptive learning sequences construction (OALSC) algorithm is used to generate adaptive learning sequences in the Remedial Tutoring Module. As shown in Figure 2-3, in [5], the inputs of the algorithm are domain ontology, students' learning portfolios and inference chains. The outputs of the algorithm are adaptive learning sequences. The ontology is defined by domain experts for representing the common error problems and related learning concepts. The students' learning portfolios keep the students' learning statuses. Inference chains contain the information of diagnosis process.

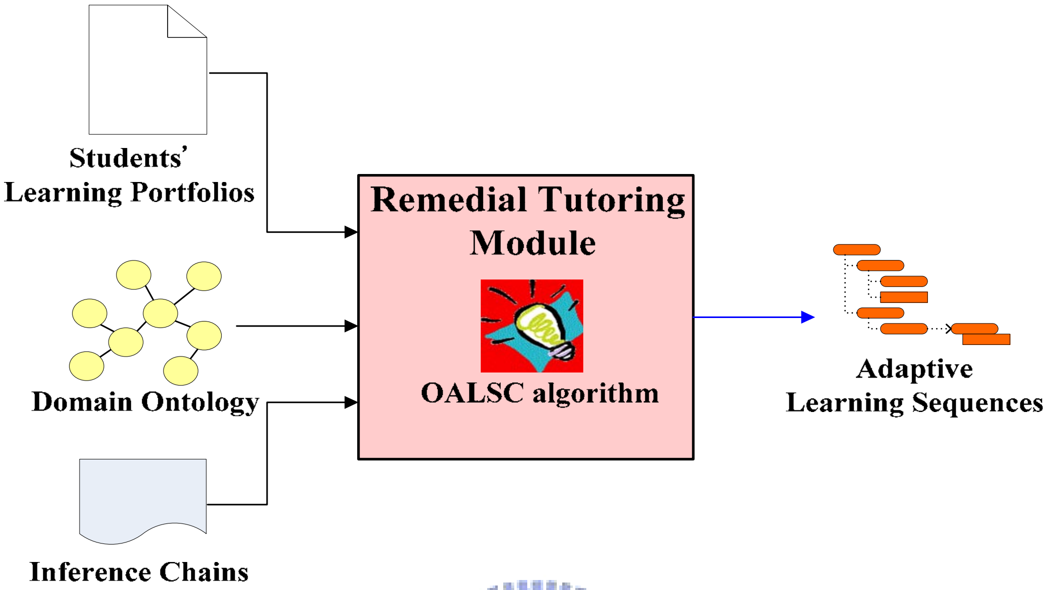


Figure 2-3: The inputs and outputs of the OALSC algorithm

The main idea of OALSC algorithm is to integrate inference chains containing the diagnostic information of tasks with specific ontology consisting of error problem nodes and the related learning concept nodes. Through the DRAMA, the inference chains can be mapped to rule class chains easily since the knowledge base of DRAMA is NORM-based. In addition, since the Two Phase Knowledge Acquisition Process is used to transform ontology to the rules and rule classes of the knowledge base, the rule class chains can be mapped to ontology easily.

On the other hand, since the specific ontology consists of leaning concepts nodes and error problem nodes, the related leaning concepts about the students' encountered problems can be found out easily if the inference chains are mapped to the paths on ontology correctly. Thus, the Remedial Tutoring Module takes the related learning concepts found on ontology and generate the adaptive learning sequences as remedial tutoring to help students solve their encountered problems. Moreover, the students' learning portfolios are used to make the learning sequences more adaptive through the OALSC algorithm.

2.7. E-learning Architecture

In E-learning, most of Learning-by-Doing systems lack the capability of providing adaptive remedial tutoring information. Therefore, as shown in Figure 2-4, Chang [5] proposed a systematic methodology to build the Learning-by-Doing Remedial Tutoring System for helping students solve their encountered problems. According to the general process of Learning-by-Doing, he designed three modules in this system, namely, Learning Module, Diagnosis Module, and Remedial Tutoring Module.

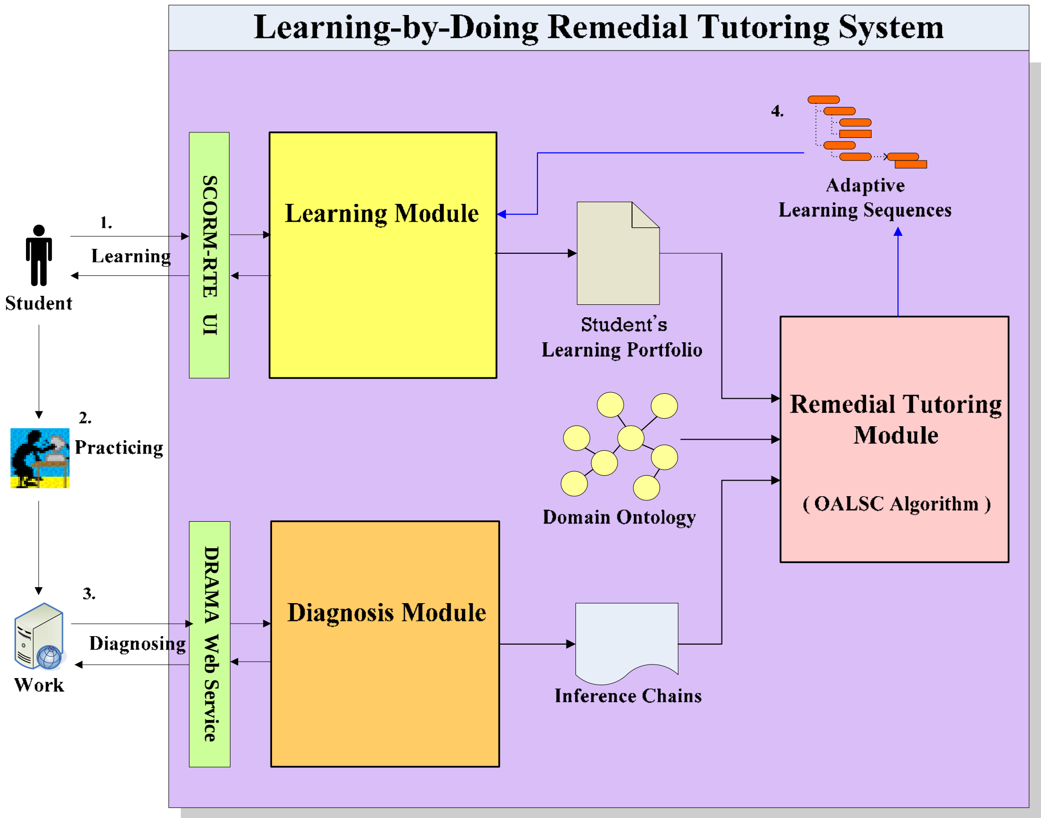


Figure 2-4: Overview of the Learning-by-Doing Remedial Tutoring System architecture

The Learning Module of the system is a learning platform which is designed to provide students learning sequences with theoretic courses for learning the required domain knowledge. Besides, this module is constructed with the SCORM 2004 Sample Run-Time Environment (Version1.3.2) and used to show the learning sequences. In SCORM RTE, the students' learning statuses and learning activities are recorded into the learning portfolios. The information is an important input source for Remedial Tutoring Module

Next, the Diagnosis Module is used to help students identify their encountered problems. The diagnostic information is recorded on the inference chains which are also important input sources for Remedial Tutoring Module to generate adaptive remedial tutoring information. The Diagnosis Module is constructed by using the DRAMA/NORM. Since the DRAMA is object-oriented, it is utilized to represent the concept classes of the domain ontology for identifying the students' problems and the related learning concepts in Learning-by-Doing.

Finally, the Remedial Tutoring Module is used to generate the adaptive learning sequences as remedial tutoring information for students by his proposed OALSC algorithm he proposed. If the diagnosis results show that the tasks have problems, the OALSC algorithm takes the students' learning portfolios, domain ontology and inference chains of diagnosis process as input sources and generate adaptive learning sequences. These adaptive learning sequences are stored in the repository of Learning Module. Later, when students need remedial tutoring information, these adaptive learning sequences can be retrieved for helping them improve their learning. In brief, this design of Learning-Diagnosis-Remedial Tutoring System is suitable to generate adaptive remedial tutoring in the Leaning-by-Doing.

Chapter 3. Intelligent Questionnaire Analysis Assistant System

At the beginning of design for the assistant system, the requirements which users would require when they analyze questionnaire data need to be analyzed. Therefore, the functions designed for the system would be appropriate to users. The requirements are analyzed first, and according to these requirements the functions of the system are designed. For each function, some techniques need to be used to achieve the requirements of the function. Finally, some schemas are formulated for each function, like ontology, rules, and data schema.

3.1. System Design Methods

To build an online system, one of the most difficult parts is how to build an online system with well maintainability and extensibility. Some techniques need to be used to help design the system.

- *Software Engineering*

The system is designed based on the flow of design and develop software described in software engineering. First of the flow, analyze which kind of system is. Second, collect and analyze users' requirements they need to assist to analyze questionnaire data. Third, design the diagrams of each function and the architecture of the system based on object-oriented technique. Therefore, the functions, architecture, and modules in the system with well maintainability and extensibility could be formulated well.

For the general of the diagrams designed for the system, to make communication easier between others, these diagrams are designed based on a standard.

- *Unified Modeling Language (UML)*

UML standard are used to design these diagrams of modules in the system, for examples, use case diagram, class diagram, and sequence diagram.

3.2. System Requirements

There are two main requirements in assisting users in analyzing questionnaire data. When users start to analyze questionnaire data, they explore the data first. However the granularity of original questionnaire data may not be good enough to explore the data. For instance, there is no significant difference between different values in higher levels of granularity. But it is still possible to find a significant difference in lower levels of granularity. Therefore, the first requirement is to ease users exploring the full data to make hypotheses of the possible statistically significant differences. In addition, after users make hypotheses of the possible statistically significant differences, they make a hypothesis and use the inferential statistics method to test this hypothesis. Since there are several methods can be used to test hypotheses, the methods user used to test hypotheses may not be appropriate. Therefore, the second requirement is to suggest users to select appropriate inferential statistics methods to test hypotheses according to the data. The system providing assistants for these two requirements can ease users finding the statistically significant differences from their questionnaire data.

3.3. System Functions

According to users' requirements, some functions for the system are designed to assist users. All functions of the system are divided into four parts, and there are several functions in each part. The functions are shown in Figure 3-1, and are described the detail below.

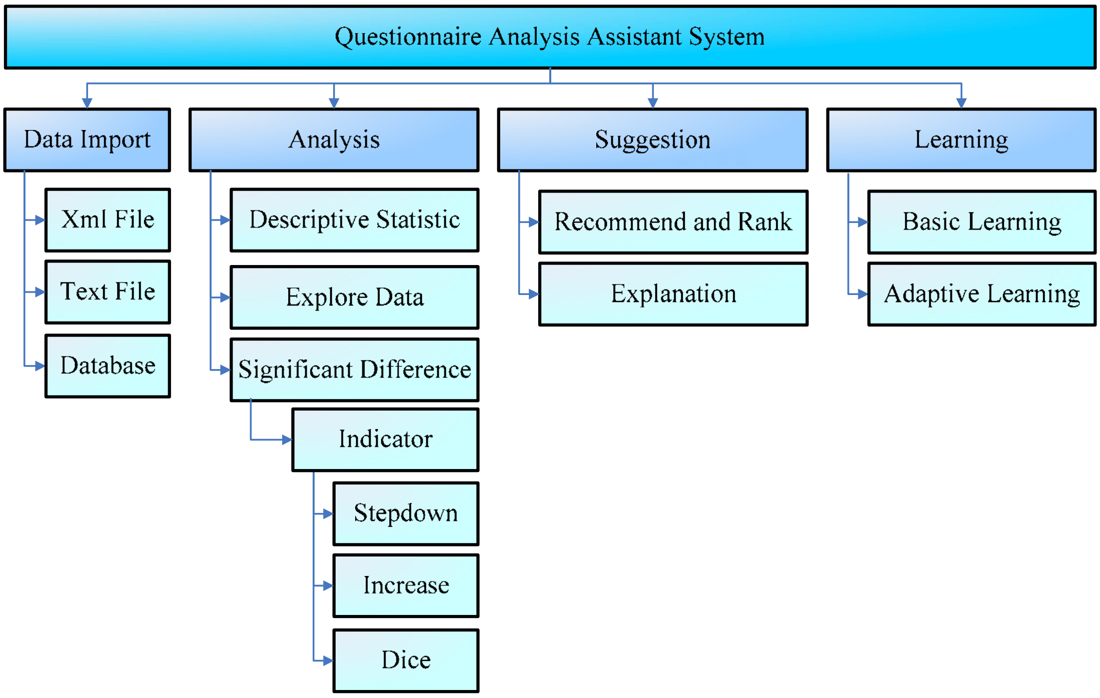


Figure 3-1 System Functions

3.3.1 Data Import

For the first part of functions, the system provides functions for users to import their questionnaire data. Users may store their data in different kinds of file types. The system needs to support users several kinds of file types. Three kinds of file types are supported for users to import their data into the system, xml file, text file, and database file. In addition to users import their questionnaire data into the system, users are also asked to provide metadata of their questionnaire data for the system. This makes the system easier to build data warehouse and to analyze data. The detail formulation of the metadata will be described in section 3.7.

3.3.2 Analysis

In the second part of functions, the system provides three functions for users to assist them analyzing questionnaire data. The first function, Descriptive Statistic, shows some descriptive statistics of the data, percentage, sum, mean, and standard deviation of each basic attribute data. The second function, Explore Data, uses OLAP to help users observe the distributions of data. The third function, Significant Difference, analyzes the data for possible significant differences in the data, and uses three kinds of indicators, Stepdown, Increase, and Dice, which are proposed by Chu [9] to construct a viewer to help users find attributes with rich information.

3.3.3 Suggestion

In the third part of functions, users need to use some statistic methods to test the data which they thought to be valuable. The system suggests some statistic methods which are appropriate to analyze the data. The first function, Recommend and Rank, suggests some statistic methods for users and ranks them from most appropriate to least. The second function, Explanation, explains the reason why the system suggests them. Therefore, users are easier to select some statistic methods to analyze. These suggestions and explanations are provided according to the domain experts' knowledge.

3.3.4 Learning

The final part of functions, a platform is provided for users to learn statistic methods. Two kinds of learning materials are provided for users. Basic Learning, this function aims at foundations of statistics to provide learning materials of basic statistics. The system construct the static learning sequences using the Ontology-based Learning Sequences Construction Scheme proposed by Chen [8] to transform a domain ontology to a basic course scheme. Adaptive Learning, provides learning materials of particular statistic methods to users who want to learn them. The adaptive learning sequences are constructed using the Ontology-based adaptive learning sequences construction algorithm proposed by Chang [5] to generate adaptive learning sequences from domain ontology, students' learning portfolios and inference chains.

3.4. System Implementation Methods

According to the functions designed for the system and because this system is developed by integrating the research we proposed before. We refer these research [5][9][16] to use the following technique to implement the system.

- *Data Warehouse*

The system needs storage to store users' questionnaire data. The data also needs to preprocess and be analyzed. Data warehouse is used to help the system preprocess, analyze, and store data.

- *On-Line Analytical Processing (OLAP)*

To give supports for users to observe their data, OLAP is used to help the system analyze the distributions of data and observe the data. Because OLAP provides real-time, quick analysis, and uses data warehouse as source.

- *Ontology*

The sources of knowledge in the system which is used to suggest statistic methods and provide learning materials are from experts interviewed about data analysis. Therefore, to present and store this knowledge for the system, ontology is used.

- *Knowledge Base*

Besides ontology is used to present and to store experts' knowledge for the system, the system also needs to transform and store the ontology in it to use. The system uses knowledge base to store the transformed knowledge and use it when the system needs to suggest statistic methods or to provide learning materials.

- *Inference Engine*

The system has to use the knowledge in the knowledge base to provide users some suggestion or materials like the experts choose some statistic methods and use these methods based on their knowledge. An inference engine is built to simulate experts' behaviors.

3.5. Domain Ontology

In the system, in order to use domain experts' knowledge to suggest appropriate statistic methods and provide adaptive learning materials, we interview experts of statistics and data analysis and discuss the knowledge from books about data analysis [27] with them. We use repertory grid method to acquire this knowledge. Therefore, a domain ontology is utilized to represent domain experts' knowledge in the system.

Figure 3-2 shows a simple example of domain ontology. This ontology describes the rules of selection about appropriate statistic methods to analyze significance of group differences. The detailed meanings of the relations are explained as follows:

- *Type of*: If concept classes A and B are all the types of concept class C, it means that C has two kinds of types, namely A and B. For example, data with one continuous dependent variable and data with multiple continuous dependent variables are two kinds of types of significance of group differences.
- *Strategy of*: If concept class A is a strategy of concept class B, it means that A is one strategy of B. For example, One-way ANOVA and T-Test are strategies of data with one discrete independent variable.

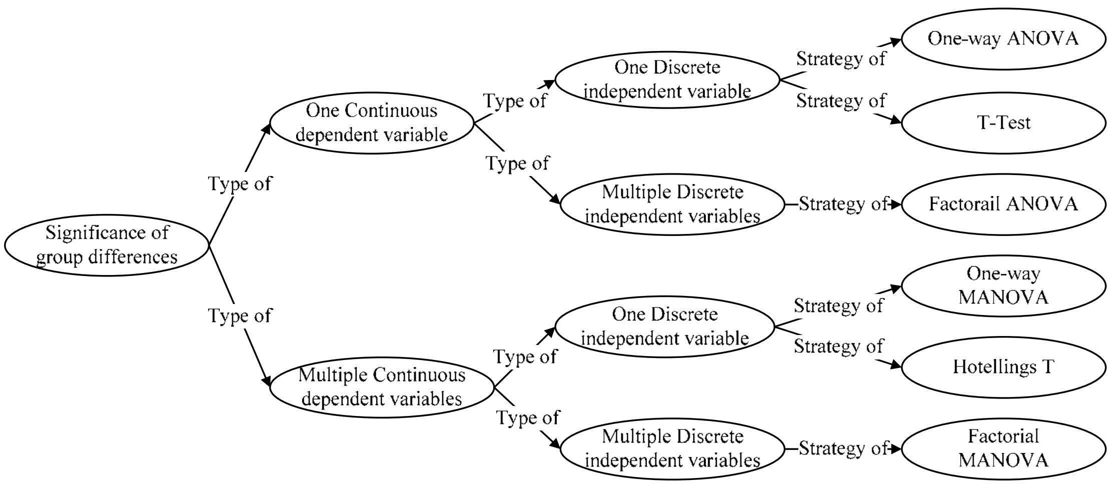


Figure 3-2 A simple example of domain ontology

For example, one part of the ontology describes one strategy experts used to select appropriate statistic methods. When analyzing the data with one continuous dependent variable and with multiple discrete independent variables, experts used factorial ANOVA to analyze this kind of data.

3.6. Rules Transformation

The knowledge used by the system about which statistic methods are appropriate to test hypotheses is referred from the example of ontology in Section 3.5. The system transfers it into rule classes which can be stored in the knowledge base of the system. Therefore, the inference engine in the system can use these rules to infer the appropriate statistic methods. There are 7 rule classes in the system used to infer appropriate methods. In each rule class, it verifies some conditions and triggers to another rule class or to results. The conditions used for inferring appropriate methods are the number of dimensions in data, the number of values in each dimension, and the data type of each dimension. Some examples of rules are listed below.

IF dimension number = 4 THEN data dimension = Three or above.

IF data dimension = Three or above THEN class trigger = Class Verify data scale.

IF data scale = Nominal or data scale = Ordinal THEN strategy = T-Test.

3.7. Data Import Schema

As we mentioned above, when users import their questionnaire data into the system, they are also asked to import a metadata of their questionnaire data. This makes the system easier to build a data warehouse of their data. The metadata describes the attributes in the questionnaire data. It includes two parts of data. The first part is about users' basic data including data type, scale, and hierarchy of this basic data. The second part is about question in the questionnaire. This part includes identify number, scale of the question. Figure 3-3 shows the DTD format which formulates the xml file of metadata. Figure 3-4 gives an example of metadata based on this DTD format.

```
<?xml version="1.0" ?>
<!DOCTYPE Questionnaire [
        <!ELEMENT Questionnaire (Description?,Basicdatapart,Questionpart)>
        <!ELEMENT Basicdatapart (Basicdata+)>
        <!ELEMENT Basicdata (Scale,Datatype,Hierarchy)>
        <!ELEMENT Hierarchy (key,Level+)>
        <!ELEMENT Questionpart (Question+)>
        <!ELEMENT Question (Description?,Measure)>
        <!ATTLIST Questionnaire name CDATA #REQUIRED>
        <!ATTLIST Questionnaire type CDATA #REQUIRED>
        <!ATTLIST Basicdata id CDATA #REQUIRED>
        <!ATTLIST Level id CDATA #REQUIRED>
        <!ATTLIST Question id CDATA #REQUIRED>
]>
```

Figure 3-3 The DTD format for xml file of metadata

```
<Questionnaire name="Health" type="csv">
        <Description>NCTU Health Promote Life Style and Health Education Course Requirement Questionnaire
                Survey</Description>
        <Basicdatapart>
                <Basicdata id="Gender">
                        <Scale>Nominal</Scale>
                        <Datatype>int</Datatype>
                        <Hierarchy>
                                <Key>Gender</Key>
                                <Level id="1">Name</Level>
                        </Hierarchy>
                </Basicdata>
                <Basicdata id="Institute">
                        <Scale>Nominal</Scale>
                        <Datatype>int</Datatype>
                        <Hierarchy>
                                <Key>Institute</Key>
                                <Level id="1">Name</Level>
                        </Hierarchy>
                </Basicdata>
                <Basicdata id="Group">
                        <Scale>Nominal</Scale>
                        <Datatype>int</Datatype>
                        <Hierarchy>
                                <Key>Group</Key>
                                <Level id="1">Name</Level>
                        </Hierarchy>
                </Basicdata>
        </Basicdatapart>
        <Questionpart>
                <Question id="a1">
                        <Description>I do extension excercise three times every week</Description>
                        <Measure>4</Measure>
                </Question>
                <Question id="a2">
                        <Measure>4</Measure>
                </Question>
                <Question id="a3">
                        <Measure>4</Measure>
                </Question>
        </Questionpart>
</Questionnaire>
```

Figure 3-4 An example of metadata

Chapter 4. System Implementation

To implement the system based on the functions as mentioned in the previous chapter, the system architecture is designed.

4.1. System Architecture

To construct the system, first of all, we need to build some database for the system, for example, data warehouse and knowledge base. Therefore, the system can be used to assist users in analyzing questionnaire data.

4.1.1 Data Preprocessing

When users import their questionnaire data into the system, they are also asked to import the metadata about questionnaire data. There are also some existent legacy database about geographical data, population data, and related research data, etc. Therefore, the system must integrate these data first to store them into data warehouse.

There are two processes to integrate the imported data. First, the system import user' data based on the metadata user imported with the questionnaire raw data. The metadata describes the formats of the questionnaire data, for example, data type, scale, and hierarchy of attributes. Second, the system integrates the imported data and the other legacy data. This process is accomplished through the tool, Microsoft SQL Server 2005. Finally, the integrated data stores in the data warehouse. The total process of data preprocessing is shown in Figure 4-1.

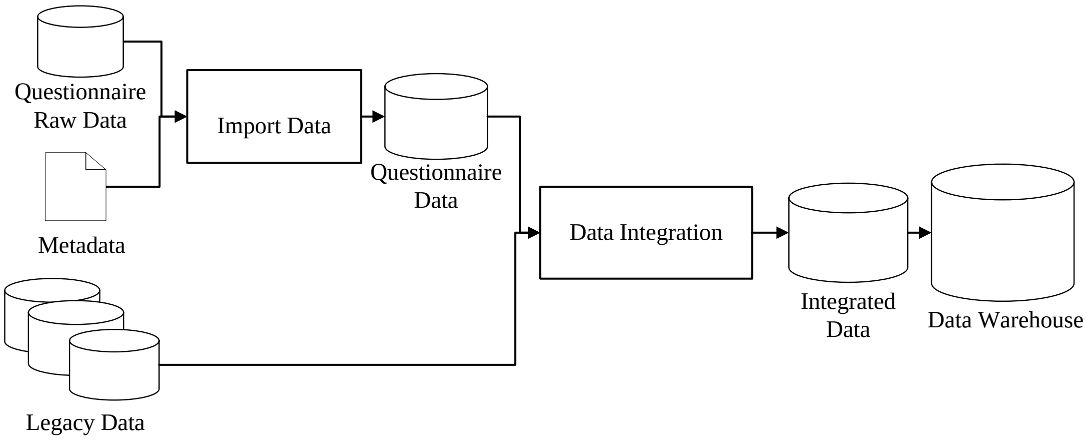


Figure 4-1 Data Preprocessing Module of System

4.1.2 Knowledge Transformation

To use the experts’ knowledge in the system, the system needs to transform the ontology to the knowledge base and content package repository. Therefore, the system can use this knowledge to assist users.

The system transforms the knowledge presented by the ontology through three processes to knowledge base and content package repository. The Knowledge Acquisition process transforms the ontology to the rule classes used in rule-based knowledge base in the system for the inference engine to infer appropriate statistic methods. OLSC process generates static learning sequences from the knowledge in the ontology. Parser process parses the ontology to get the materials of learning courses. The static learning sequences and courses store in content package repository for learning platform to get the learning materials. The total process of knowledge transformation is shown in Figure 4-2.

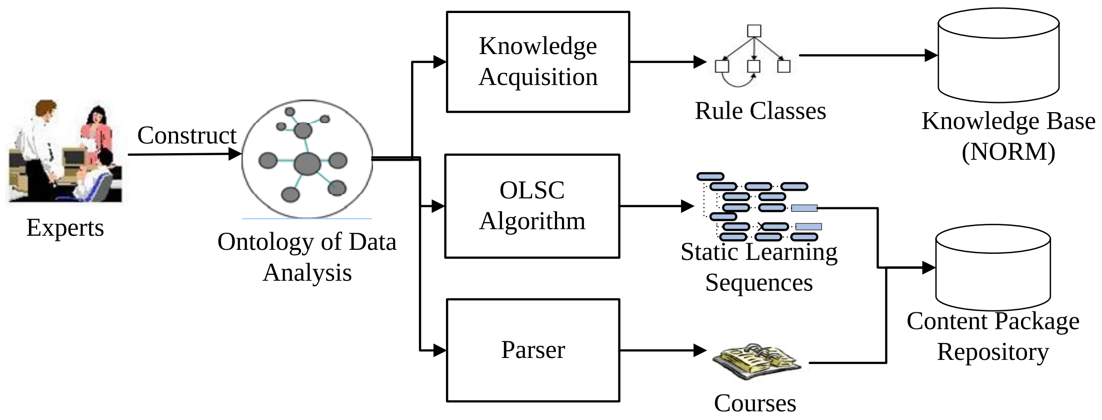


Figure 4-2 Knowledge Transformation Module of System

According to the users' requirements when they analyze questionnaire data, the architecture of the system is detailedly designed and is shown in Figure 4-3 System Architecture.

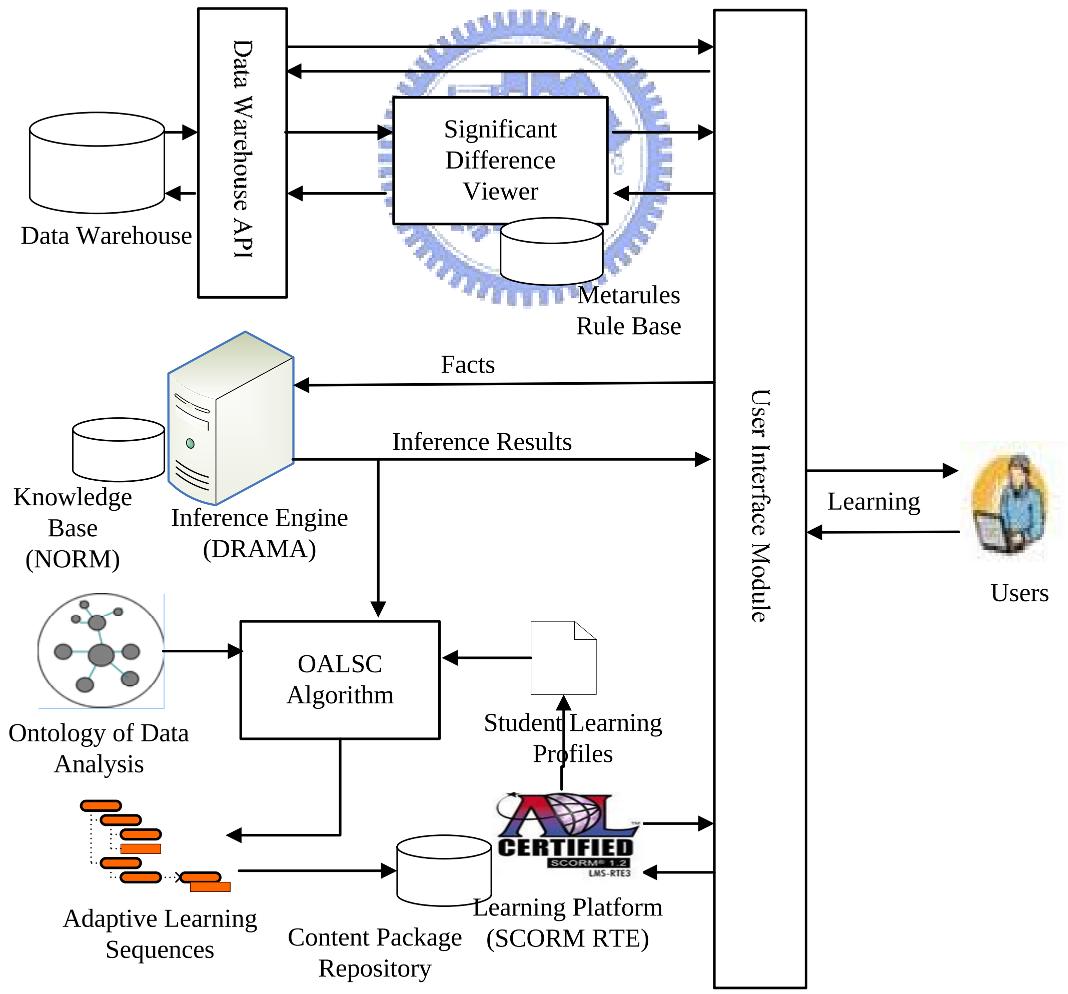


Figure 4-3 System Architecture

4.1.3 Data Exploration

In order to assist users in exploring their questionnaire data, first, the system provides some results of descriptive statistics about the data, for example, mean, standard deviation. The system access the data warehouse through the data warehouse API to get the data and compute the descriptive statistics of the data.

Second, the system provides a tool to assist users in viewing the distributions of the data. This is done by constructing OLAP to help the system. OLAP is constructed in the data warehouse API. The system accesses the analysis reports from OLAP to assist users.

Third, to assist users in finding the possible statistically significant differences, the system uses the three kinds of indicators designed according to experts' experiments. Significant Difference Viewer is constructed to generate the indicators. This process generates the indicators based on the degree of possible statistically significant differences in the data and the metarules defined to control the indicators. Therefore, users can make hypotheses of the possible statistically significant differences according to these indicators.

4.1.4 Inference Process

After users make hypotheses, they need suggestions from experts which statistics methods are appropriate to test the hypotheses. Therefore, the system use the experts' knowledge transformed in the knowledge base. An inference engine is constructed to use the knowledge of knowledge base. The inference engine can infer that the appropriate statistics methods like experts. The system first analyzes the characteristics of the data in the hypotheses, for example, data type, data scale. Then the system transmits these facts to inference engine. The inference engine infers the appropriate statistics methods based on these facts and the rules in the knowledge base. The system also provides explanations about these results according to the rules used in the inference process.

4.1.5 Learning Platform

For those statistics methods suggested by the system, the system also provides learning courses about these methods for users. A learning platform is constructed for this purpose. In order to provide courses adapt to each user, we use OALSC algorithm proposed by Chang [5] to generate the adaptive learning sequences for each user. OALSC algorithm generates the adaptive learning sequence based on the students' profile, ontology, and the method of inference results. The learning platform provides the learning courses according to the adaptive learning sequences.

4.2. System Implementation

When implement the system, some tools are used to assist implementing.

- *Data Warehouse and OLAP*

Microsoft SQL Server 2005 is used to help the system construct data warehouse and OLAP in it. Microsoft SQL Server 2005 provides well integration between data warehouse and OLAP and easy interface to construct them. It also assists the system preprocess data. It integrates well with Microsoft Visual Studio 2005. Therefore, it also gives advantages when constructing user interface.

- *Ontology*

Protégé is used to build the ontology for the system. Protégé provides complete functions to build ontology. Furthermore, it can generate an xml file about ontology, it eases the system to use the ontology.

- *Knowledge Base and Inference Engine*

DRAMA is used to construct knowledge base and inference engine in the system. DRAMA is a rule-based, client-server tool/environment for KBS development. It can assist knowledge engineers in building up a system. Briefly, DRAMA contains lots of innovative techniques including Object-Oriented technology, knowledge inheritance, etc. It also contains useful tools, like rule verification tool, knowledge acquisition assistant tool and the inference server.

- *Learning Platform*

Learning platform is constructed by SCORM RTE 2004. SCORM standard is most popular standard of learning materials. The learning materials in the system are also referred to this standard. So, SCORM RTE 2004 is used to construct learning platform. Users can view learning materials on it.

- *User Interface*

Microsoft Visual Studio 2005 ASP.NET with C# is used to construct user interface. It provides very convenient tools to construct user interface. Therefore, the user interface is constructed easily. Also as mentioned above, it integrates well with data warehouse and OLAP.

At last, the following figures are some example pages of the system. After login in the system, it shows the main page to users in Figure 4-4. Users can use functions they want to use by choosing the menu at the left part of the page.

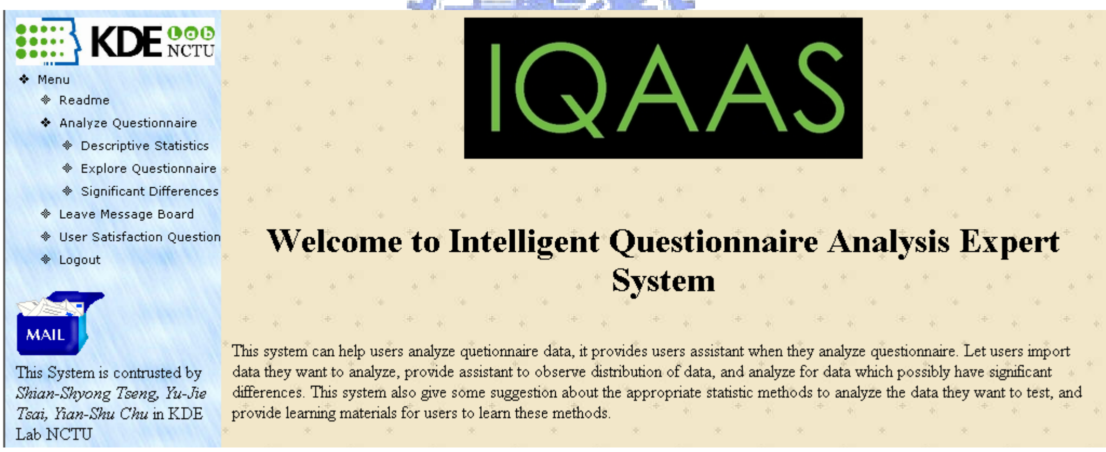


Figure 4-4 Main Page of System

First of all, user can choose Descriptive Statistic node. He/She can operate to get descriptive statistic of his/her data. User can select the questionnaire data he/she wants to analyze using the drop down list on the top, and then selects some basic data and some questions he/she wants to analyze. Finally, he/she presses the Observe button to get the results like Figure 4-5.

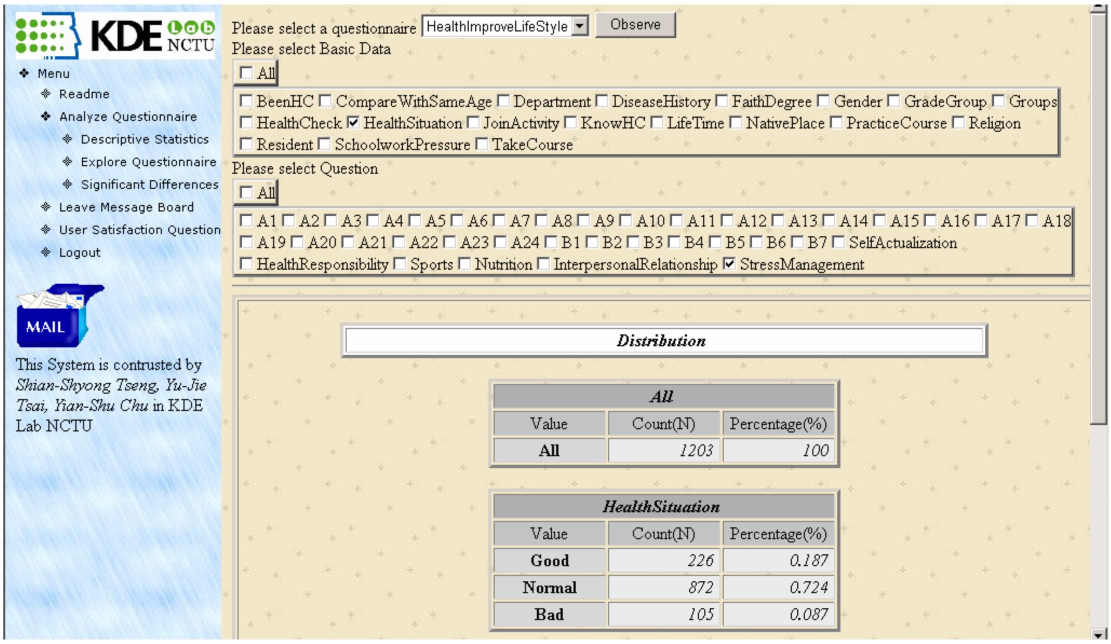


Figure 4-5 Descriptive Statistic Function

Second, user can choose Explore Questionnaire Data node, he/she can operate to get some tables or diagrams about his/her data. The same with Descriptive Statistic function, user selects which questionnaire data first. He/She can add the basic data or questions to the table or diagram to get the analysis results. Figure 4-6 is an example of a table and a diagram of data.

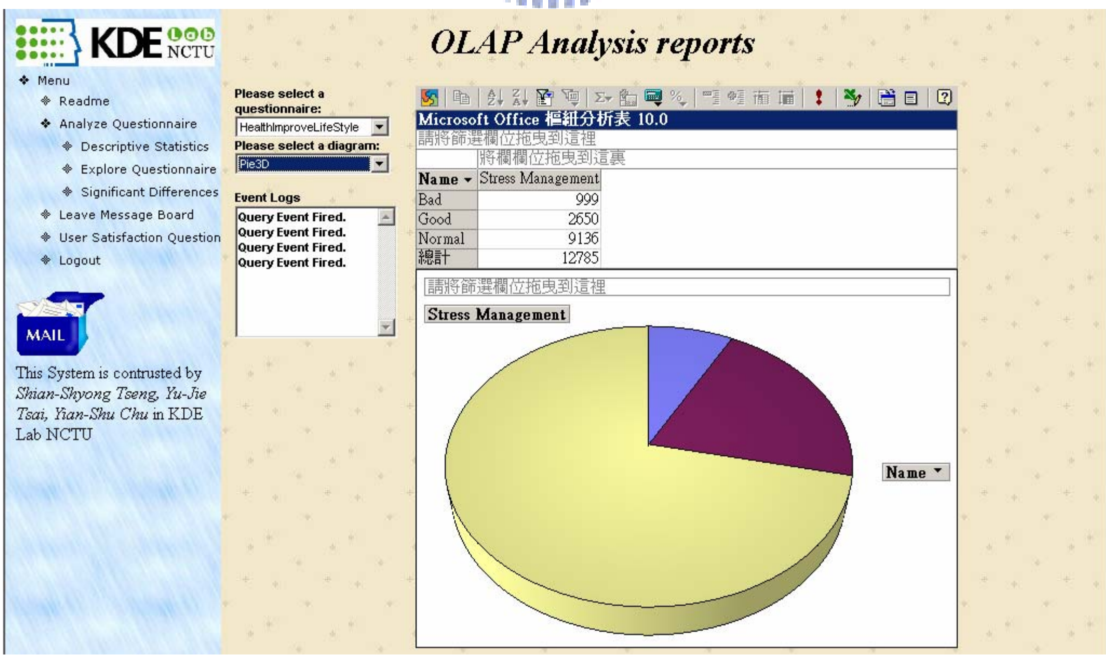


Figure 4-6 Explore Data Function

Third, user can choose Significant Difference node, he/she operates this function to get some assistants for analyzing his/her data. It is similar to the two functions above; user chooses his/her questionnaire, and the question he/she want to analyze first, and then he/she press Observe button to begin analyzing. The system displays the analysis results. User can compare the color of each cell in the table to see where the system analyzed might have significant differences in it. He/She can click one cell to get further analysis, or press Suggestion for Appropriate Analysis Methods button to get suggestion from the system to analyze these data. Figure 4-7 shows a sample for this.

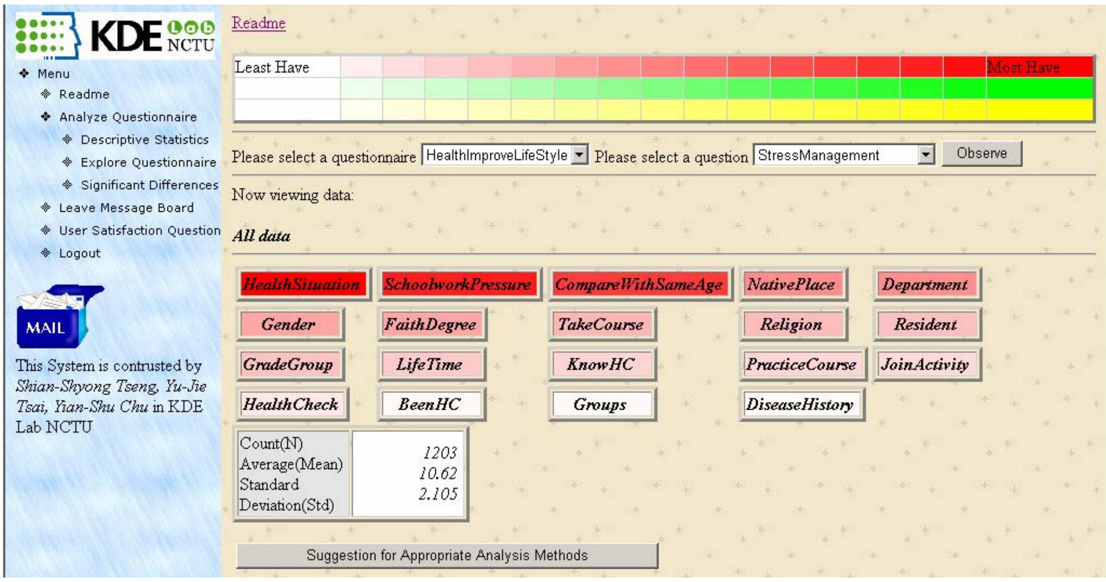


Figure 4-7 Significant Differences Function

The suggestion results display like Figure 4-8. User can operate to get more information about the inference results or the learning materials about the method by pressing Detail or Learn button. Figure 4-9 shows the example of explanation. Figure 4-10 gives learning materials of one method.

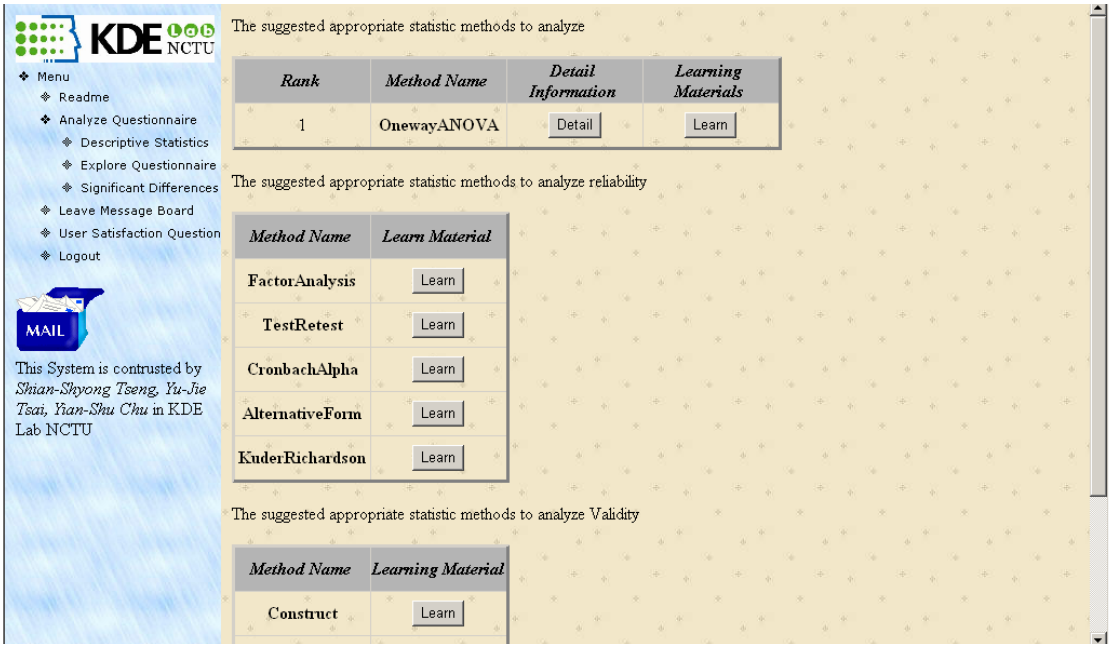


Figure 4-8 Recommend and Rank Function

Figure 4-9 Explanation Function

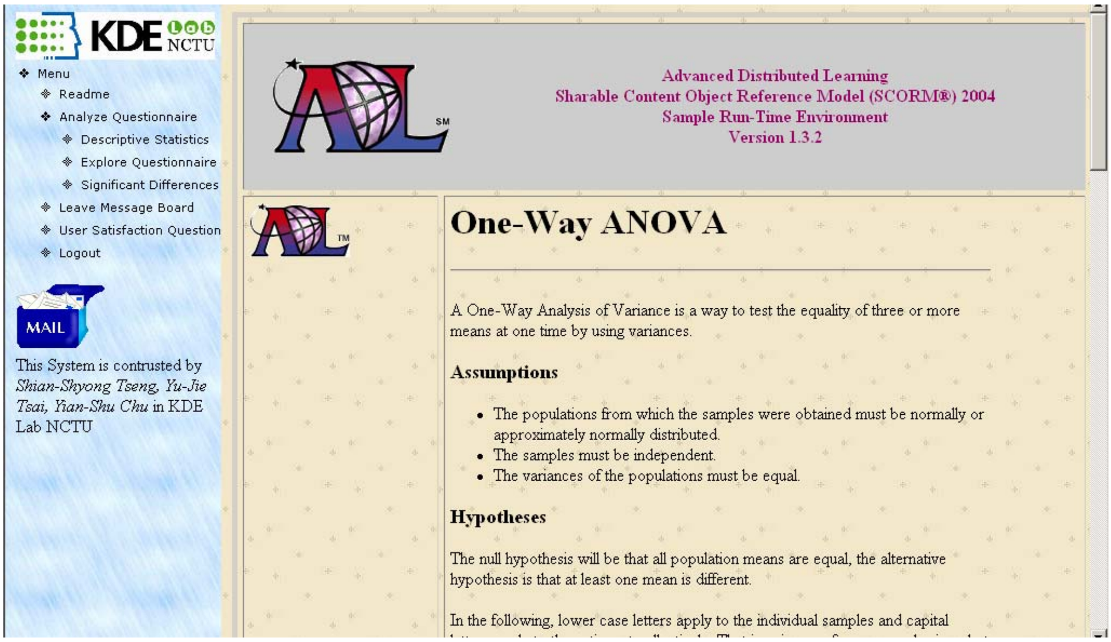


Figure 4-10 Adaptive Learning Function

Chapter 5. Experiments

At the end of this thesis, two experiments are designed to evaluate the system. In order to evaluate if the system really provides users assistants for questionnaire analysis or not, in the first experiment, a case study is given by the system. The analysis results and other assistant are described by the system. A satisfaction questionnaire survey is designed for the second experiment. To get satisfaction of users who used the system to assist analyzing questionnaire data.

5.1. Case Study

The questionnaire of the case study is NCTU health improve life style and health education course requirements questionnaire survey. This questionnaire is proposed by NCTU Sanitary and Health Care Center.

5.1.1 Data information

This questionnaire is designed to be a reference for planning and pushing sanitary and health activities in the future in NCTU. It has three parts in the questionnaire, basic data part, life style part, and course requirements part. The system analyzed the first two parts. There are 19 attributes in basic data part, for example gender, department, and score. There are 24 questions in life style part and each question belongs to an item. There are 6 items designed to analyze, self-actualization, health responsibility, sports, nutrition, interpersonal relationship, and stress management. There are 1203 records in the database.

5.1.2 Experiment Results

The results of analysis from the system are shown in Table 5-1. For self-actualization item, these attributes may have significant differences in them, schoolwork pressure, health situation, and comparison with same ages. To analyze the attribute, schoolwork pressure, the system suggested one-way ANOVA and factor analysis for reliability.

Item	Possible Attributes have Significant Difference			
Self-Actualization	School pressure	Health situation	Compare with same ages	
Health Responsibility	Resident*	Native place*		
Sports	Resident	Compare with same ages	Health situation	
Nutrition	Resident*	Health situation	School pressure	Native place*
Interpersonal Relationship	Health situation	School pressure	Compare with same ages	

Table 5-1 Case study analysis results

We compared the results with the results from using T-test and One-way AONVA. Most results are the same. Some attributes do not have significant differences using T-test or One-way ANOVA to verify. These attributes are marked a star sign in the results table. We analyzed the data by ourselves and found the reason why the results are different. It is because the number of records is significant difference between different values of the attribute. For the other attributes, the results analyzed by the system really have significant differences.

5.2. Satisfaction Survey

For evaluating the system if it could provide enough assistants for users. An experiment is designed for this purpose. We gather the satisfaction of users for their comments about using the system to assist them.

5.2.1 Satisfaction Questionnaire Experiment Design

Survey the satisfactions of users who used the system to assist analyzing questionnaire data online.

- Find 30 testers from the internet.
- Let these testers use the system to assist them analyzing questionnaire data.
- After they finished their analysis, they filled a satisfaction questionnaire.

5.2.2 Evaluation

The satisfaction questionnaire includes the items as follow.

- System user interface
- System response time
- System functions corrections
- Add other functions to the system
- Satisfaction of using the system

15 questions are designed for these five items and Likert 5 scale is used to evaluate the degree of users' satisfaction, from very disagree (1) to very agree (5).

5.2.3 Experiment Results

The results of satisfaction questionnaire survey are shown in Table 5-2

The mean points of each question and the mean points of each item are listed in the right two columns. The mean points of system user interface are 3.32. This is the lowest points between the five items which are evaluated. This means the user interface of our system should be improved to be friendlier. The mean points of add other functions to our system are 4.08. It shows users hope the system can be developed further to provide more functions. The mean points of Satisfaction of using the system are 4.02.

NO.	Question	Mean of one question	Mean of one item
Q01	The user interface of this system is convenient to use.	3.57	3.32
Q02	The user interface of this system is friendly to use.	3.27	
Q03	The user interface of this system is clear to use.	3.13	
Q04	The response time of this system is fast.	3.93	3.93
Q05	This system can assist analyzing questionnaire correctly.	4.17	3.95
Q06	This system can assist analyzing questionnaire conveniently.	3.93	
Q07	This system provides complete functions to analyze questionnaire.	3.67	
Q08	This system provides rich information to assist analyzing questionnaire.	4.03	
Q09	I hope this system can analyze other functions.	4.03	4.08
Q10	I hope this system can provide other data of descriptive statistics.	4.00	
Q11	I hope this system can add more statistic methods to suggest.	4.20	
Q12	I will use this system again in the future.	4.23	4.02
Q13	Generally speaking, I am satisfied with the spending time when I used this system to assist analyzing questionnaire data.	3.80	
Q14	Generally speaking, I am satisfied with the correction of results when I used this system to assist analyzing questionnaire data.	4.07	
Q15	Generally speaking, I am satisfied with this system.	3.97	

Table 5-2 Satisfaction questionnaire survey results

Chapter 6. Conclusions

6.1. Conclusions

In the questionnaire analysis, how to find a significant difference between two or more groups in one measure is one of the major problems which social science researchers are concerned about. However, finding possible significant differences is difficult for social science researchers. In order to assist junior researchers in finding possible significant differences, in this paper, we build a system to help users find the possible significant differences from the data cube. The system provides the significant difference viewer to assist users in exploring data. The viewer uses three kinds of indicators designed according to the experts' experiments and the metarules defined by experts. The system also provides suggestion to select appropriate statistics methods to test users' hypotheses. The system gives explanations about the suggestions and provides learning courses for these methods.

The prototype of the proposed system was also implemented, and we designed two experiments at last to evaluate the functions of the system if these functions could assist users or not. In the first experiment, we got a questionnaire data to do case study. To evaluate if the functions of our system could run normally and return some results this system analyzed. In the second experiment, we found some testers and let them use the system to assist analyzing questionnaire data. After they finished their analysis, we gathered the satisfaction questionnaire data their filled.

The results of first experiment show the analysis results by the system using one case of questionnaire data, and the results are almost correct. In the results of second experiment, users were interested in the idea of this system and they hope more useful functions can be added into the system. Generally speaking, they were satisfied with the system.

6.2.Future Works

There are some works to be done to improve the system to be friendlier and to analyze data more correct.

First, according to the results of the first experiments, there are still some causes the experts did not consider. Therefore, we will interview the experts to tell them these causes and make them define the Indicators more deeply to increase the precision.

Second, we will also consult some experts of data analysis and get some advices to improve the user interface of the system to be friendlier and easier for users to assist them in analyzing questionnaire data.

Reference

[1] S. Agarwal, R. Agrawal, P. M. Deshpande, etc., “On the Computation of Multidimensional Aggregates”, International Conference on VLDB, pp. 506-521, Mumbai, India, 1996.

[2] D. J. Berndt, A. R. Hevner, and J. Studnicki, “The Catch data warehouse: support for community health care decision-making”, Decision Support Systems, Vol. 35, Issue 3, pp. 367-384, June 2003.

[3] B. P. Butz, M. Duarte, and S. M. Miller, “An Intelligent Tutoring System for Circuit Analysis”, IEEE Transactions on Education, 2006.

[4] J. M. Caroll and A. P. Aaronson, ”Learning By Doing With Simulated Intelligent Help”, Communications of the ACM, Volume31, Number9, September 1988.

[5] C. H. Chang, “Building a Learning-by-Doing Remedial Tutoring System for DNS Management”, National Chiao Tung University, Master Thesis, 2006.

[6] S. Chaudhuri and U. Dayal, “An Overview of Data Warehousing and OLAP Technology”, SIGMOD, Record 26(1): pp. 65-74, 1997.

[7] C. S. Chen. “A unifying framework for intelligent DNS management”, International Journal of Human-Computer Studies. Volume 58, Issue 4, Pages 415-445, April 2003.

[8] R. Y. Chen, S. S. Tseng, C. L. Liu, C. S. Chang and C. S. Chen, “Learning Sequences Construction Using Ontology and Rules”, International Conference on Computers in Education, 2005.

[9] Y. S. Chu, S. S. Tseng, R. J. Luo, Y. J. Tsai, “A Framework for Finding Statistically Significant Differences”, The 24th Workshop on Combinatorial Mathematics and Computation Theory, 2006.

[10] A. Datta and H. Thomas, “The cube data model: a conceptual model and algebra

for on-line analytical processing in data warehouses", Decision Support Systems, Vol. 27, Issue 3, pp. 289-301, December 1999.

[11] S. Fischer, "Course and exercise sequencing using metadata in adaptive hypermedia learning systems", ACM Journal of Educational Resources in Computing, Vol. 1, No. 1, Article #3, 21 pages, Spring 2001.

[12] G. D. Garson, Advances in Social Science and Computers, A research annual, H61.3 A 38 v.1, 1989.

[13] H. Gupta and I. S. Mumick, "Selection of Views to Materialize in a Data Warehouse:, IEEE, Alencia, Spain, 1998.

[14] C. L. Liu, S. S. Tseng, C. S. Chen, "Design and Implementation of an Intelligent DNS Management System", ICACT, Korea, pp.373-386, Feb. 5-7, 2002.

[15] F. M. Lpez, "Overview of Methodologies for Building Ontologies", IJCAI-99 workshop on Ontologies and Problem-Solving Methods, Stockholm, Sweden, 1999.

[16] R. J. Luo, "A study of Statistically Significant Difference Pattern Detection", National Chiao Tung University, Master Thesis, 2006.

[17] S. E. Middleton, N. R. Shadbolt, and D. C. De Roure, "Ontological User Profiling in Recommender Systems", ACM Transactions on Information Systems, Vol. 22, No. 1, Pages 54-88, January 2004.

[18] T. Nodenot, C. Marquesuzaa, P. Laforcade, and C. Sallaberry, "Model based Engineering of Learning Situations for Adaptive Web Based Educational Systems", international World Wide Web conference, 2004.

[19] R. Páircéir, S. McClean, B. Scotney, "Discovery of Multi-level Rules and Exceptions Discovery of Multi-level Rules and Exceptions", ACM SIGKDD, 2000.

[20] T. Palpanas, N. Koudas, and A. Mendelzon, "Using Datacube Aggregates for

Approximate Querying and Deviation Detection", IEEE Trans. on Knowledge and Data Engineering, Vol. 17, No. 11, pp. 1465-1477, 2005.

[21] L. Rokach and O. Mainmon, "Top-down Induction of Decision Trees Classifiers - a Survey", IEEE Transaction on Systems, Man, and Cybernetics, 2005.

[22] N. Sala, "The Internet And The Learning by Doing Strategy In The Educational Processes: A Case of Study", IEEE, 2001.

[23] S. Sarawagi, R. Agrawal, and N. Megiddo, "Discovery-driven Exploration of OLAP Data Cubes", Research Report, IBM Almaden Research Center, 1998.

[24] S. Sarawagi, "User-adaptive exploration of multidimensional data", International Conference on VLDB Conference, pp. 307-316, Cairo, Egypt, 2000.

[25] A. Shoshani, "OLAP and Statistical Databases Similarities and Differences, PODS, Tucson Arizona USA, 1997.

[26] Sommerville , Software Engineering, Pearson, England, 2004.

[27] Tabachnick, Using multivariate statistics, Collins College Publishers, 1996.

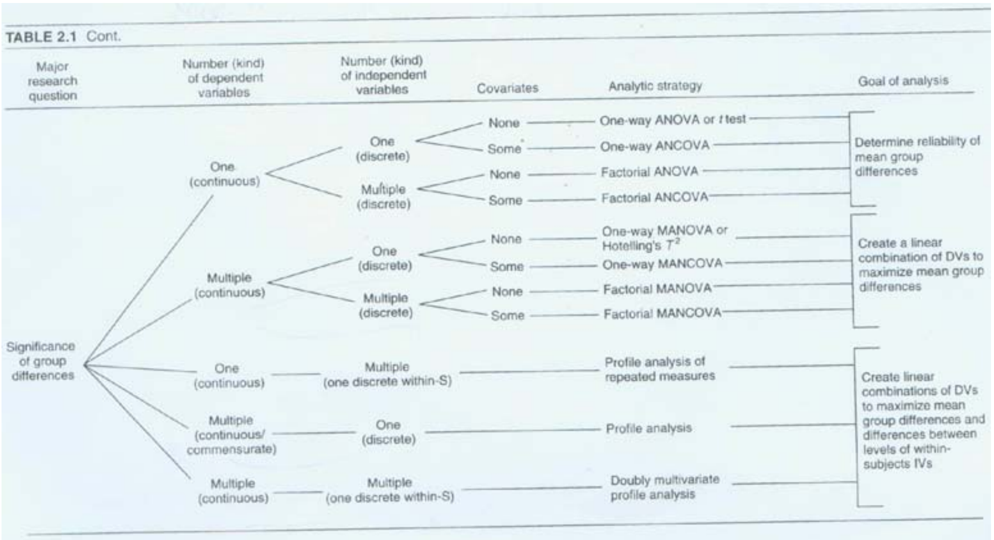
[28] W. Wang, J. F. Weng, J. M. Su and S. S. Tseng, "Learning Portfolio Analysis and Ming in SCORM Compliant Environment", 34th ASEE/IEEE Frontiers in Education Conference.

[29] H. Zhang and B. Padmanabhan, "On the Discovery of Significant Statistical Quantitative Rules", ACM KDD Trans., 2004.

[30] 邱皓政，量化研究法（二）：統計原理與分析技術，雙葉書廊，民國 94 年。

Appendix

Barbara Tabachnick [27] categorized the research questions into four types: degree of relationship among variables, significance of group differences, prediction of group membership, and structure. He also constructs a decision tree for each type to select an analytic strategy. In this thesis, we only considered the significance of group differences. These trees are shown in below.



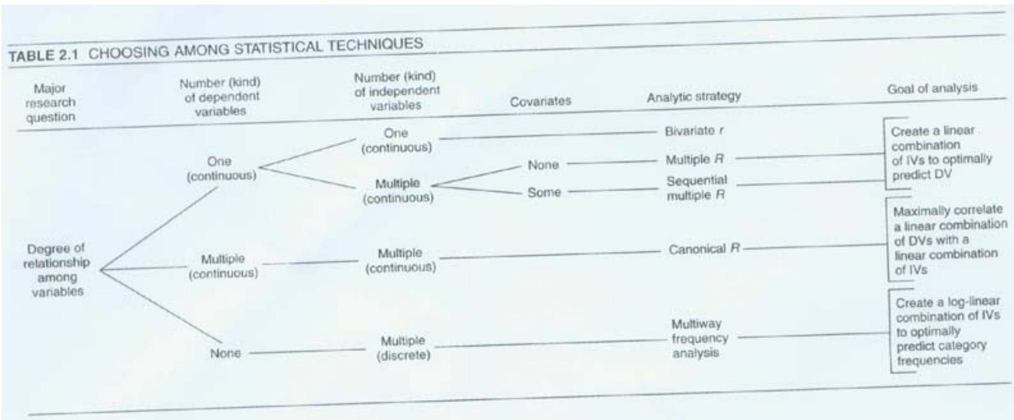
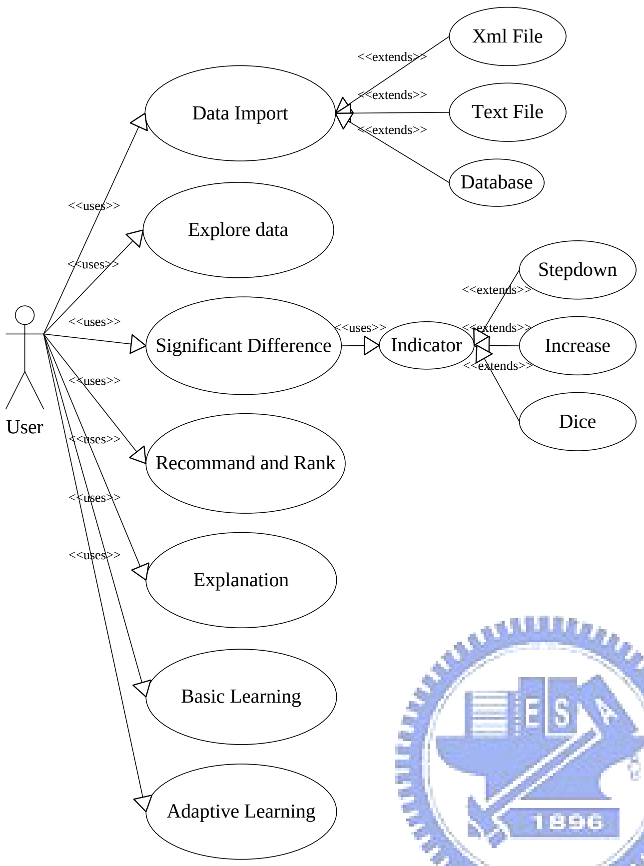
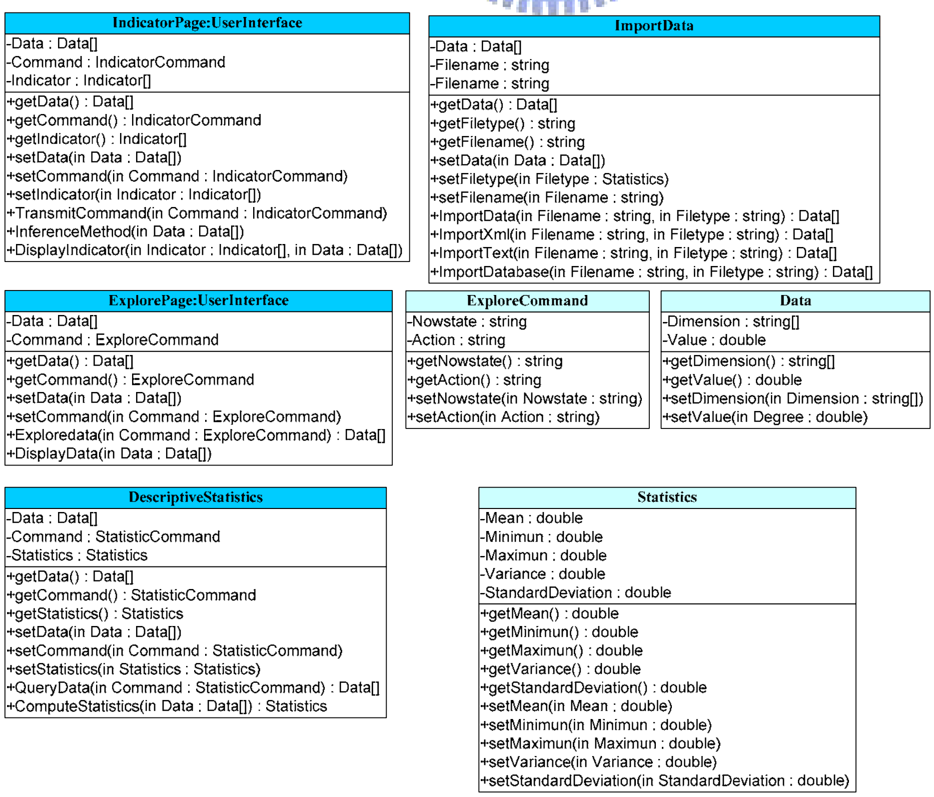


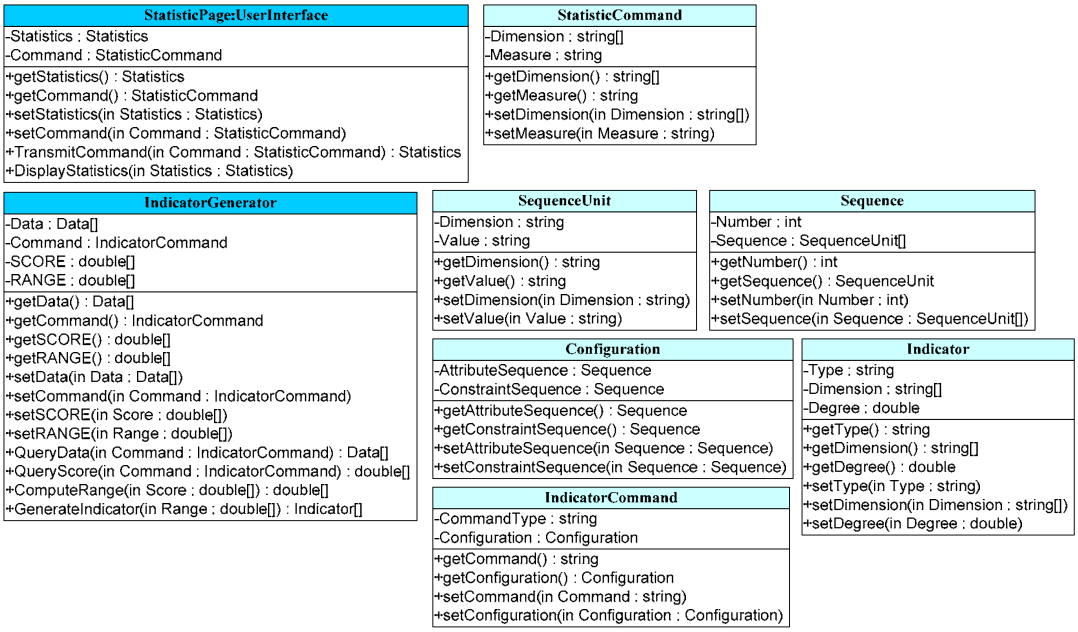
TABLE 2.1 Cont.

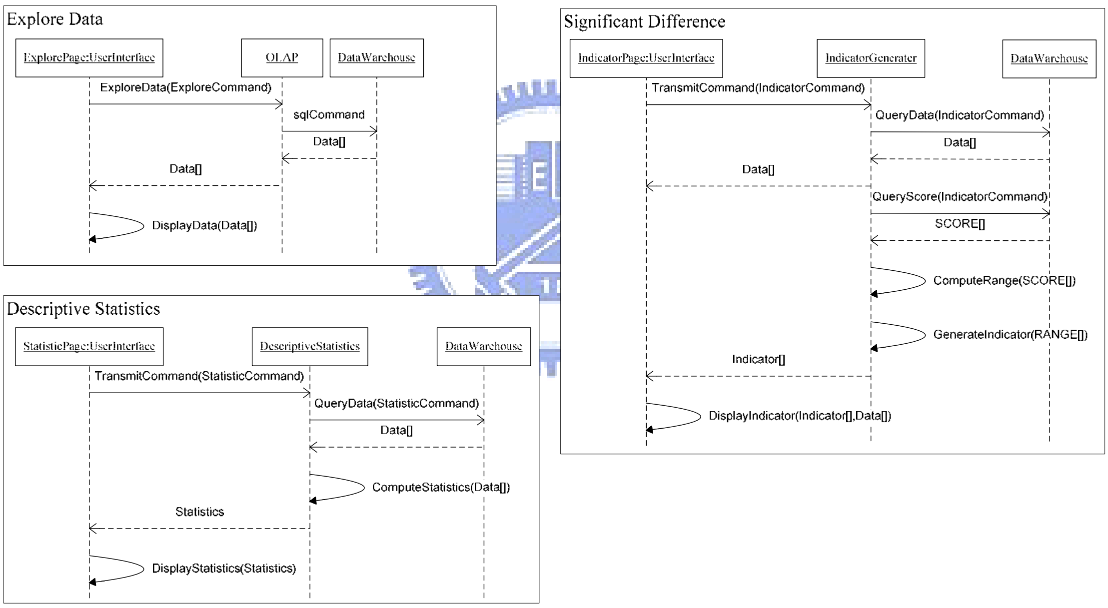
Prediction of group membership	One (discrete)	Multiple (continuous)	None	One-way discriminant function	Create a linear combination of IVs to maximize group differences
			Some	Sequential one-way discriminant function	
		Multiple (discrete)		Multiway frequency analysis (logit)	Create a loglinear combination of IVs to optimally predict DV
		Multiple (continuous and/or discrete)	None	Logistic regression	Create a linear combination of the log of the odds of being in one group
			Some	Sequential logistic regression	
	Multiple (discrete)	Multiple (continuous)	None	Factorial discriminant function	Create a linear combination of IVs to maximize group differences (DVs)
			Some	Sequential factorial discriminant function	
Structure	Multiple (continuous observed)	Multiple (latent)		Factor analysis (theoretical)	Create linear combinations of observed variables to represent latent variables
	Multiple (latent)	Multiple (continuous observed)		Principal components (empirical)	
	Multiple (continuous observed and/or latent)	Multiple (continuous observed and/or latent)		Structural equation modeling	Create linear combinations of observed and latent IVs to predict linear combinations of observed and latent DVs

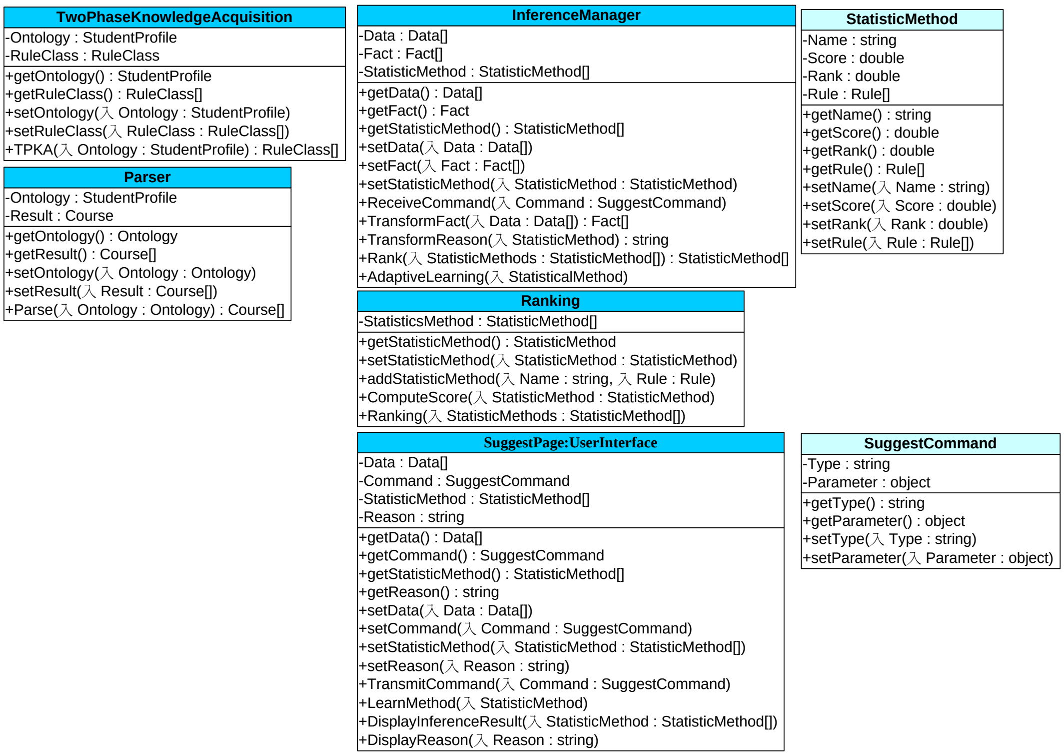
The use case diagram, class diagrams, and sequence diagrams designed for the system are listed below.

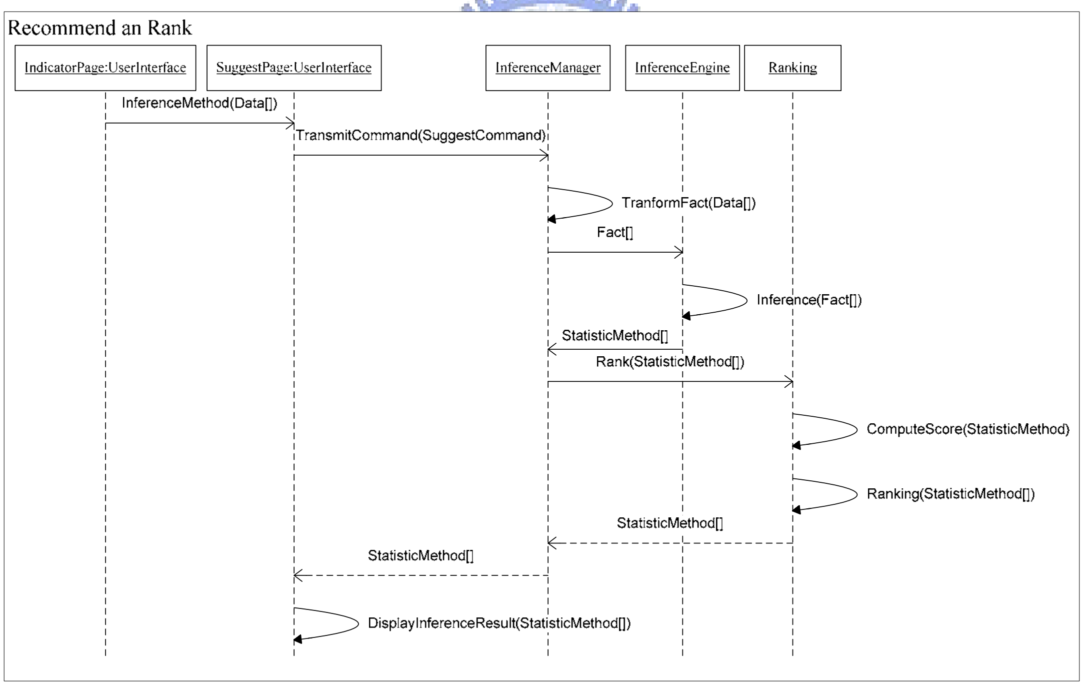


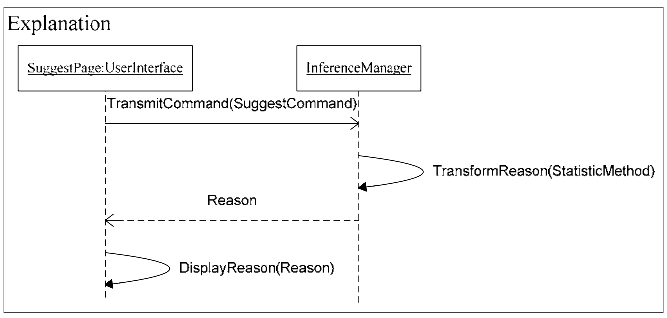



StatisticPage:UserInterface
-Statistics : Statistics
-Command : StatisticCommand
+getStatistics() : Statistics
+getCommand() : StatisticCommand
+setStatistics(in Statistics : Statistics)
+setCommand(in Command : StatisticCommand)
+TransmitCommand(in Command : StatisticCommand) : Statistics
+DisplayStatistics(in Statistics : Statistics)
StatisticCommand
-Dimension : string[]
-Measure : string
+getDimension() : string[]
+getMeasure() : string
+setDimension(in Dimension : string[])
+setMeasure(in Measure : string)
IndicatorGenerator
-Data : Data[]
-Command : IndicatorCommand
-SCORE : double[]
-RANGE : double[]
+getData() : Data[]
+getCommand() : IndicatorCommand
+getSCORE() : double[]
+getRANGE() : double[]
+setData(in Data : Data[])
+setCommand(in Command : IndicatorCommand)
+setSCORE(in Score : double[])
+setRANGE(in Range : double[])
+QueryData(in Command : IndicatorCommand) : Data[]
+QueryScore(in Command : IndicatorCommand) : double[]
+ComputeRange(in Score : double[]) : double[]
+GenerateIndicator(in Range : double[]) : Indicator[]
SequenceUnit
-Dimension : string
-Value : string
+getDimension() : string
+getValue() : string
+setDimension(in Dimension : string)
+setValue(in Value : string)
Sequence
-Number : int
-Sequence : SequenceUnit[]
+getNumber() : int
+getSequence() : SequenceUnit
+setNumber(in Number : int)
+setSequence(in Sequence : SequenceUnit[])
Configuration
-AttributeSequence : Sequence
-ConstraintSequence : Sequence
+getAttributeSequence() : Sequence
+getConstraintSequence() : Sequence
+setAttributeSequence(in Sequence : Sequence)
+setConstraintSequence(in Sequence : Sequence)
Indicator
-Type : string
-Dimension : string[]
-Degree : double
+getType() : string
+getDimension() : string[]
+getDegree() : double
+setType(in Type : string)
+setDimension(in Dimension : string[])
+setDegree(in Degree : double)
IndicatorCommand
-CommandType : string
-Configuration : Configuration
+getCommand() : string
+getConfiguration() : Configuration
+setCommand(in Command : string)
+setConfiguration(in Configuration : Configuration)


Explore Data
ExplorePage:UserInterface
OLAP
DataWarehouse
ExploreData(ExploreCommand)
sqlCommand
Data[]
Data[]
DisplayData(Data[])
Significant Difference
IndicatorPage:UserInterface
IndicatorGenerator
DataWarehouse
TransmitCommand(IndicatorCommand)
QueryData(IndicatorCommand)
Data[]
Data[]
QueryScore(IndicatorCommand)
SCORE[]
ComputeRange(SCORE[])
GenerateIndicator(RANGE[])
Indicator[]
DisplayIndicator(Indicator[],Data[])
Descriptive Statistics
StatisticPage:UserInterface
DescriptiveStatistics
DataWarehouse
TransmitCommand(StatisticCommand)
QueryData(StatisticCommand)
Data[]
ComputeStatistics(Data[])
Statistics
DisplayStatistics(Statistics)


TwoPhaseKnowledgeAcquisition
-Ontology : StudentProfile
-RuleClass : RuleClass
+getOntology() : StudentProfile
+getRuleClass() : RuleClass[]
+setOntology(入 Ontology : StudentProfile)
+setRuleClass(入 RuleClass : RuleClass[])
+TPKA(入 Ontology : StudentProfile) : RuleClass[]
Parser
-Ontology : StudentProfile
-Result : Course
+getOntology() : Ontology
+getResult() : Course[]
+setOntology(入 Ontology : Ontology)
+setResult(入 Result : Course[])
+Parse(入 Ontology : Ontology) : Course[]
InferenceManager
-Data : Data[]
-Fact : Fact[]
-StatisticMethod : StatisticMethod[]
+getData() : Data[]
+getFact() : Fact
+getStatisticMethod() : StatisticMethod[]
+setData(入 Data : Data[])
+setFact(入 Fact : Fact[])
+setStatisticMethod(入 StatisticMethod : StatisticMethod)
+ReceiveCommand(入 Command : SuggestCommand)
+TransformFact(入 Data : Data[]) : Fact[]
+TransformReason(入 StatisticMethod) : string
+Rank(入 StatisticMethods : StatisticMethod[]) : StatisticMethod[]
+AdaptiveLearning(入 StatisticalMethod)
StatisticMethod
-Name : string
-Score : double
-Rank : double
-Rule : Rule[]
+getName() : string
+getScore() : double
+getRank() : double
+getRule() : Rule[]
+setName(入 Name : string)
+setScore(入 Score : double)
+setRank(入 Rank : double)
+setRule(入 Rule : Rule[])
Ranking
-StatisticsMethod : StatisticMethod[]
+getStatisticMethod() : StatisticMethod
+setStatisticMethod(入 StatisticMethod : StatisticMethod)
+addStatisticMethod(入 Name : string, 入 Rule : Rule)
+ComputeScore(入 StatisticMethod : StatisticMethod)
+Ranking(入 StatisticMethods : StatisticMethod[])
SuggestPage:UserInterface
-Data : Data[]
-Command : SuggestCommand
-StatisticMethod : StatisticMethod[]
-Reason : string
+getData() : Data[]
+getCommand() : SuggestCommand
+getStatisticMethod() : StatisticMethod[]
+getReason() : string
+setData(入 Data : Data[])
+setCommand(入 Command : SuggestCommand)
+setStatisticMethod(入 StatisticMethod : StatisticMethod[])
+setReason(入 Reason : string)
+TransmitCommand(入 Command : SuggestCommand)
+LearnMethod(入 StatisticMethod)
+DisplayInferenceResult(入 StatisticMethod : StatisticMethod[])
+DisplayReason(入 Reason : string)
SuggestCommand
-Type : string
-Parameter : object
+getType() : string
+getParameter() : object
+setType(入 Type : string)
+setParameter(入 Parameter : object)


Recommend an Rank
IndicatorPage:UserInterface
SuggestPage:UserInterface
InferenceManager
InferenceEngine
Ranking
InferenceMethod(Data[])
TransmitCommand(SuggestCommand)
TranformFact(Data[])
Fact[]
Inference(Fact[])
StatisticMethod[]
Rank(StatisticMethod[])
ComputeScore(StatisticMethod)
Ranking(StatisticMethod[])
StatisticMethod[]
StatisticMethod[]
DisplayInferenceResult(StatisticMethod[])


Explanation
SuggestPage:UserInterface
InferenceManager
TransmitCommand(SuggestCommand)
TransformReason(StatisticMethod)
Reason
DisplayReason(Reason)

OntologyBasedLearningSeuenceConstruction
-Ontology : StudentProfile -LerningSequence : LearningSequence[]
+getOntology() : StudentProfile +getLearningSequence() : LearningSequence[] +setOntology(入 Ontology : StudentProfile) +setLearningSequence(入 LearningSequence : LearningSequence[]) +OLSC(入 Ontology : StudentProfile) : LearningSequence[]

OntologyBasedAdaptiveLearningSeuenceConstruction
-Ontology : Ontology -StudentProfile : StudentProfile -InferenceResult : StatisticMethod -LerningSequence : LearningSequence[]
+getOntology() : Ontology +getStudentProfile() : StudentProfile +getInferenceResult() : StatisticMethod[] +getLearningSequence() : LearningSequence[] +setOntology(入 Ontology : Ontology) +setStudentProfile(入 StudentProfile : StudentProfile) +setInferenceResult(入 InferenceResult : StatisticMethod[]) +setLearningSequence(入 LearningSequence : LearningSequence[]) +OALSC(入 Ontology : Ontology, 入 StudentProfile : StudentProfile, 入 InferenceResult : StatisticMethod[]) : LearningSequence[]

LearningPage:UserInterface
-Command : LearningCommand -Course : Course[]
+getCommand() : LearningCommand +getCourse() : Course +setCommand(入 Command : LearningCommand) +setCourse(入 Course : Course) +GenerateCourse(入 Command : LearningCommand) : Course +DisplayCourse(入 Course : Course[])

LearningCommand
-Type : string -Parameter : object
+getType() : string +getParameter() : object +setType(入 CourseName : string) +setParameter(入 Parameter : object)