

# Abnormal Data Detection for an E-Business using Object-Oriented Approach

Zongxiao Yang<sup>1</sup>, Yanyi Zheng<sup>1</sup>, Yanping Gao<sup>1</sup>, Chuanye Cheng<sup>1</sup>,  
Sheng Xu<sup>2</sup> and Hiroyuki Yamaguchi<sup>2</sup>

1 Institute of Systems Science and Engineering,  
HeNan University of Science and Technology, Luoyang 471003, China  
zxyang@mail.haust.edu.cn,

WWW home page: <http://www.haust.edu.cn/>

2 Research Center,  
Venture Link Communications Co. Ltd, Tokyo 107-6017, Japan

**Abstract.** An application service provider solution, named LINKCAFE for the food industry, offers customers a new means to leverage the internet to support line-of-business applications while reducing total cost of ownership. It is necessary for managers to maintain data accuracy and ensure data coherence in the online system. In this paper, abnormal data detection applications (ADDA) based on object oriented approach was developed by using unified modeling language models such as use-case diagrams and sequence diagrams. The programming has been accomplished for system managers for online service. Through verification and validation of the application in functionality to practical service for some years, the results have already proved that actual application of ADDA has not only enhanced the accuracy of data detection but also reduced data checking time to original 1/20 for system managers.

## 1 Introduction

Partners in technology have been continuously searching for new technologies that will enhance their client business. To most small- and medium-sized enterprises, they cannot usually afford to buy expensive applications essential for their development. In the late 1990s, a new kind of e-business called applications service provider (ASP) appeared and developed rapidly and vigorously these years. It offers the clients a new means of leveraging the Internet to support line-of-business applications while reducing total cost of ownership.<sup>[1-5]</sup>

An ASP solution named LINKCAFE has been put into practice successfully in the food industry for many years. Customers and client managers can more conveniently manipulate their service activities such as daily business management, marketing prospect prediction and technical supports by lease of the LINKCAFE system. They can experience immediate benefits through this online e-business without the need to purchase, install, upgrade and maintain costly servers, applications, databases and thus avoid hiring additional technical support for upgrade and maintenance. But it is of vital importance for system

managers to ensure data coherency because large amounts of online business data are processed everyday.

OOA-based abnormal data detection applications (ADDA) for LINKCAFE system will be developed in this research. UML (Unified Modeling Language) models are applied to identify the functional structures for requirement analysis and construct the entity relation models for the architectural design of ADDA. The applications are implemented by 4GL (fourth-generation language) such as PL/SQL, HTML and JavaScript. The effectiveness and timesaving performance of ADDA will be verified through online verification and validation of the system managers.

## 2 LINKCAFE-An ASP Solution

LINKCAFE system, based on Website and RDBMS, consists of three functional parts shown in Fig.1, which include: (1) data maintenance and management of online business by inserting and updating real-time data, maintaining client commercial information, and saving & modifying business data in the system etc. (2) data analysis of sales, orders and costs to clients such as league users, owner users, leader users and supervisor users for their prediction estimation, business administration etc. (3) IT-based automatic franchise enterprise support by POS system, electronic order system and work procedure management system.

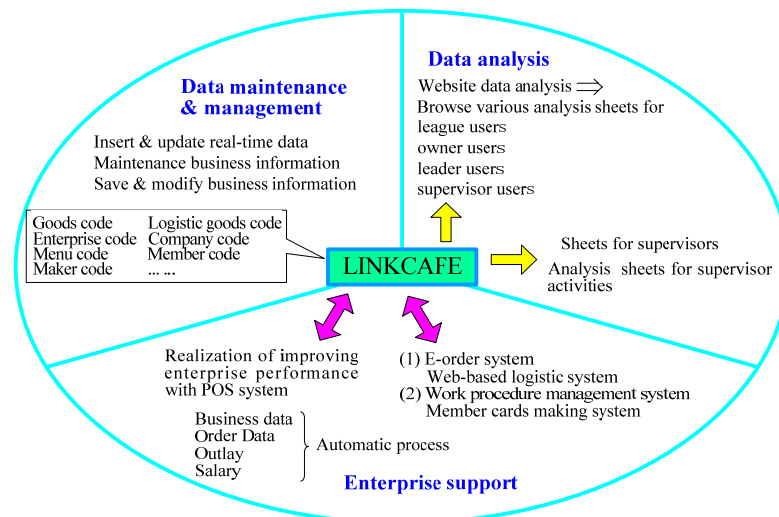


Fig.1. Outline of LINKCAFE system

The architecture of LINKCAFE shown in Fig.2 is mainly composed of data registries, basic sheets and TM (Technical Manual) sheets. In the data registry sub-system, twelve business data registry programs, such as enterprise calendar registry program, cashbook management program, can be provided for franchisor customers. For the basic sheets sub-system, three format categories including more than forty business report packages are developed for online business supports and enterprise activities to various franchise customers. Concerning TM sheets sub-system, nine profit-and-loss analysis packages are provided with many online supports such as prediction estimations, target

analysis and administration. All of the functions make franchise clients manage their business more conveniently and economically.

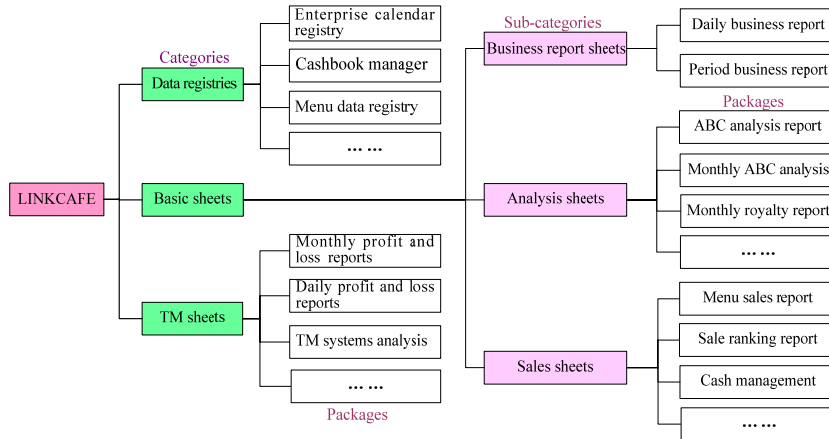


Fig.2. Architecture of LINKCAFE system (Partial)

To ensure accuracy and coherence of all the business data is necessary for system managers of this large-scale online system LINKCAFE everyday. The existence of abnormal data may affect the normal operation of the e-business and lower reputation of the franchise clients or even cause the ASP solution to face survival crisis. Therefore, it is necessary for the data managers to develop abnormal data detection package of the online system in order to improve labor efficiency and save working time.

### 3 OOA-based Construction of ADDA

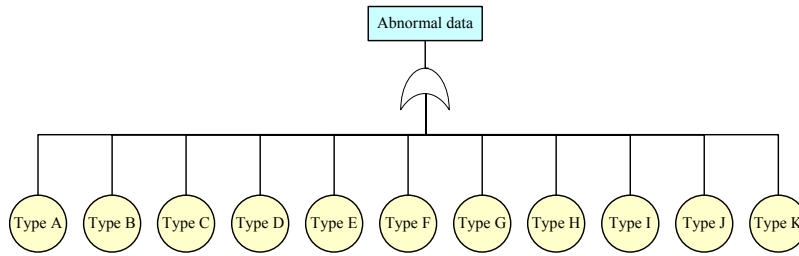
#### 3.1 Fault Tree Analysis for Abnormal Data

Abnormal data in LINKCAFE must be clarified in order to make development of ADDA more convenient. They are defined as 12 types which are stated as: (1) adjustment of department sales, (2) adjustment of sale data, (3) adjustment of goods sales, (4) adjustment of sum, (5) adjustment of owners and members, (6) adjustment of customer number, (7) adjustment of sale sum 35, (8) adjustment of sale sum 36, (9) overlap of sale data, (10) double business day, (11) presence of supplementary sales, and (12) adjustment of tax included in a discount.

Fault tree analysis (FTA)<sup>[6]</sup> is adopted to describe contribution of each abnormal data, and evaluates the reliability of a large-scale system qualitatively or quantitatively. The abnormal data are embodied through a fault tree in Fig.3. Every abnormal data is replaced with a capital Latin letter in order of alphabet. For instance, Type A represents the abnormal data of adjustment of department sales. Each abnormal data itself will alone lead to the occurrence of abnormal data, thus the twelve types are connected with an “OR” gate. By means of set theory, the cause-effect relationship can be expressed by the following equation:

$$D(A) = \bigcup_{i=1}^{12} \text{Type}(i) \tag{1}$$

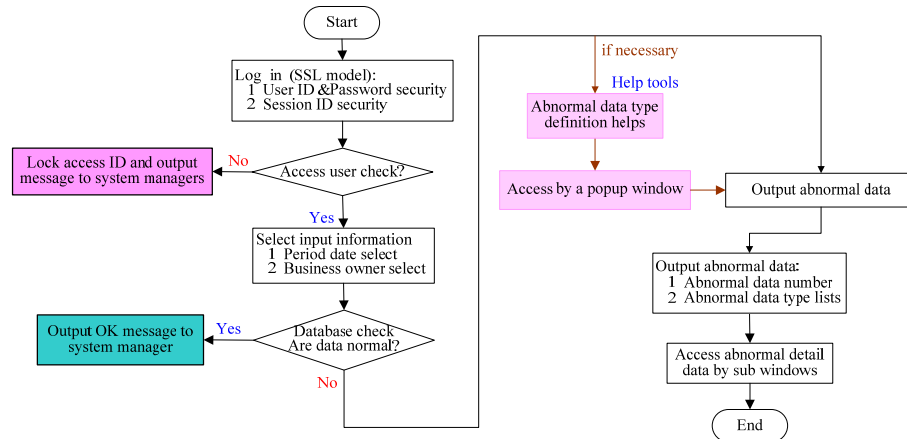
Where,  $D(A)$  represents the top event “abnormal data” while  $Type(i)$  denotes the abnormal data types. The employment of FTA can help system analysts find out abnormal data and perfect LINKCAFE system in a more high-efficient way.



**Fig.3.** A fault tree for abnormal data in LINKCAFE

**3.2 Architecture of ADDA**

A functional flow chart of ADDA is constructed on the basis of requirement analysis for LINKCAFE system shown in Fig.4.

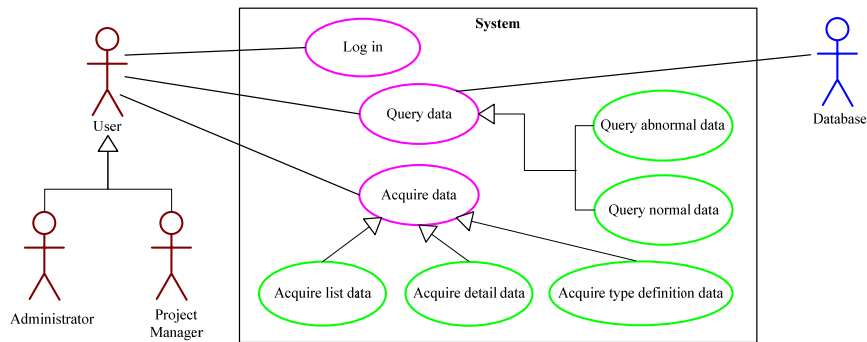


**Fig.4.** Functional flow chart of ADDA

In ADDA, we design the **Login** platform based on SSL (Security Socket Layer) model background with user ID, password security and session ID checking for system managers and database users. Certain access user would be locked if the access is not permitted and the application can output the access information to system managers for system security and safety; otherwise, the login user can select the input information including the period date and business owner code from the menu options, and then search the real time data from the online system. If the online data is identified on normal status, **OK** message will be output to the system managers; otherwise the abnormal data information including abnormal data number and abnormal data type list will be output. A popup window can be opened for abnormal data type definition help whenever necessary. The user can open sub-windows while he wants to check more abnormal detail data.

### 3.3 OOA-base Modeling

UML, which has become a standard modeling language for object oriented modeling [7-8], was applied to construct models of ADDA. The system requirements are discovered by means of discussion with various users or questionnaires etc., and are finally illustrated through use case diagrams. A high-level use case diagram shown in Fig.5 gives the general functions of the applications, which corresponds with the functional flow chart in Fig.4. As can be seen, the actor Users can inspect abnormal data by logging into the system, querying data and finally acquiring data.



**Fig.5.** A use case diagram for the system general function

On the base of use case diagrams, classes can be identified. A class is a set of objects that share a common structure and common behavior called attributes and operations respectively. In ADDA, there are total ten kinds of classes named as *ProcShowList*, *ProcTop*, *ProcTopLeft* etc. For class *ProcTop*, its operations consist of *DataCheck*, *ListData*. Likewise, other classed contains a set of their own attributes and operations.

A sequence diagram shown in Fig.6 describes information exchange among all classes and models the behavior of objects during the process of abnormal data detection. Every object is an instance of a class in the class diagram, interacting with each other through messages in order of time. For example, to realize the function of data query, the object "User" sends message *ListData* to the object *ProcTop* first, and then the object *ProcTop* calls its method *FuncGetData* itself so as to search the abnormal data in the database of LINKCAFE system, finally the abnormal data can be detected. Entity-relationship diagrams about ADDA were easily constructed based on UML models but omitted here. They enabled us to develop ADDA by computer languages conveniently.

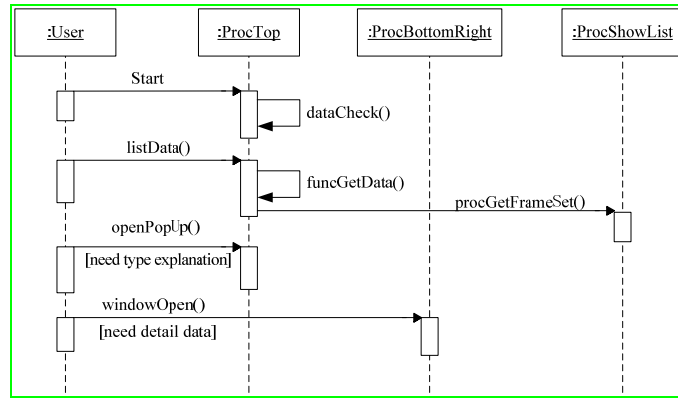


Fig.6. A sequence diagram for the abnormal data detection

### 3.4 Software Development

ADDA are easily implemented by means of 4th generation computer languages such as PL/SQL, HTML and JavaScript by OOA-based modeling [9]. The package body is composed of a main control procedure, seven platform control sub-procedures, five data control sub-procedures, one data catch sub-function and some common super-procedures for the diagnosis of twelve abnormal data types (symbolized by: A, B, C, D, ...) as well as some abnormal detail data. As an example, partial source codes of the method *FuncGetData* written with PL/SQL are given in Fig.7. And development of the ADDA was completed successfully in the same way for system managers in consideration of software performance tuning in programming [10].

```

FUNCTION FuncGetData (GCode IN VARCHAR2,
                    SDate IN VARCHAR2)
RETURN BOOLEAN IS
BEGIN
/*
  Input Information
  GCode : Business owner code
  SDate : Selected date information
  Output Information
  TRUE or FALSE
*/
DECLARE
CURSOR CS14(SCode IN VARCHAR2, SDate IN DATE) IS
SELECT NVL(SUM(HECODE(A.CD1406,
DECODE(SUBSTR(SCode,1,2),
'CD1', A.CD1406,
'CD6', DECODE(SUBSTR(A.CD1408,8,1),
'1',DECODE(SIGN(TO_NUMBER
(SUBSTR(A.CD1408,12,2))-21),
-1, A.CD1406, A.CD1405 * A.CD1410),
A.CD1405 * A.CD1410),
0, 1)),0) Cnt
FROM TB1400 A
WHERE A.CD1401 = SCode
AND A.CD1402 = SDate ;
RecCS14 CS14%ROWTYPE;
mSCode VARCHAR2(7);
BEGIN
RecCS14.Cnt := 0 ;
OPEN CS14 (mSCode, TO_DATE(SDate,'YYYYMMDD')) ;
FETCH CS14 INTO RecCS14;
CLOSE CS14 ;
RETURN TRUE;
EXCEPTION
WHEN NO DATA FOUND THEN
RETURN FALSE;
WHEN OTHERS THEN
RETURN FALSE;
END;
END;
    
```

Fig.7. Sources of the Method *FuncGetData* (Partial)

## 4 Verification and Validation

The developed ADDA for LINKCAFE system have been put in practice for online verification and validation of the effectiveness, speediness and timesaving performance for some years.

When LINKCAFE system operates normally, a large quantity of online business data are inserted and updated automatically into the system everyday. The occurrence of abnormal data is unavoidable at any time. For instance, what

if the salesperson in one of the shops received money but input wrong data into the POS system? This will cause inconsistency between the bill and actual total sum of money. Checking accuracy, coherence and trustiness of business data was a time-consuming task for data managers before ADDA was installed. Actually, to find out all the abnormal data will cost data managers too much of time that may even be counted by hours! But by employment of the diagnostic package, the abnormal data could be searched rapidly and the detection result was obtained within several minutes at most! It is given in the form of an abnormal data diagnostic list, which is illustrated by a screenshot shown in Fig.8.

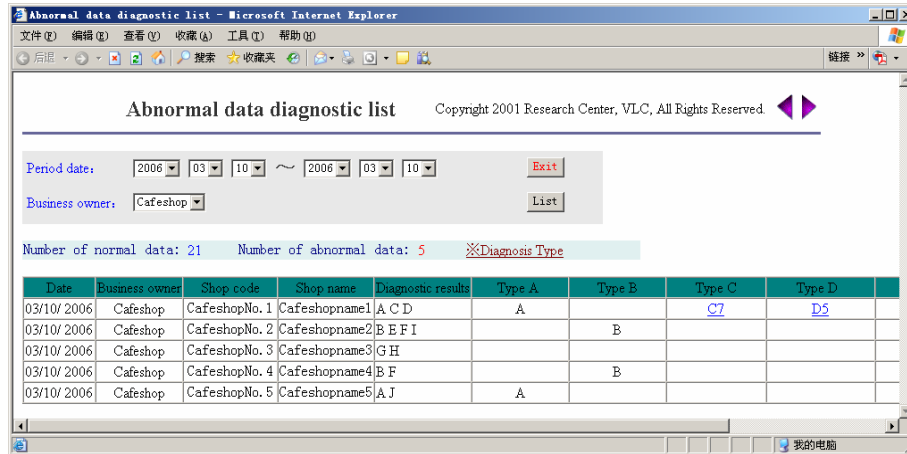
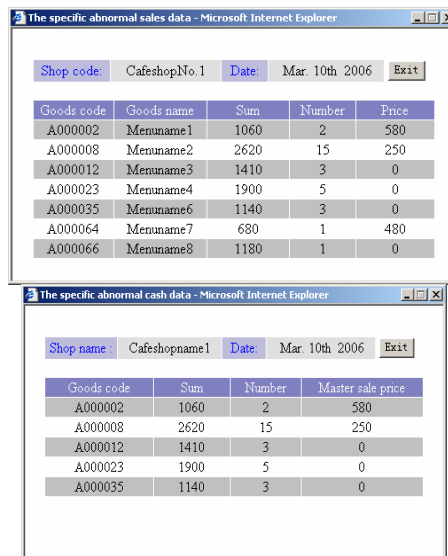


Fig.8. An abnormal data diagnostic list of ADDA



(a) An abnormal detail data for Type C (b) An abnormal detail data for Type D

Fig.9. Diagnostic results for the detailed abnormal data

As is shown in Fig.8, the detection result is obtained when clicking the button *List* after *period date* and the *business owner code* are selected. Such

symbols as A, B, C represent the types of abnormal data. Some sub-windows could be linked for more abnormal detail data information, for example C7 means that abnormal data type C has seven abnormal detail data records shown in Fig.9 (a). Another detailed detection result for D5 is illustrated in Fig.9 (b). The online verification and validation results have proved the applications improved the accuracy and coherence of data detection, increased the nonstop run trustiness of LINKCAFE system, saved data maintenance time of administrative staffs to original 1/20. Good reputation of LIKKCAFE system was established among the customers in food industry.

## 5 Concluding Remarks

In this research, OOA-based ADDA were developed for LINKCAFE system, an online e-business of ASP solution, for purpose of improving abnormal data detection accuracy and saving labor. The requirements were obtained and expressed by UML modeling such as use-case diagrams and sequence diagrams. The realization of package relied on 4GL such as PL/SQL, HTML and JavaScript.

After the diagnostic package are put into online service for some years, the effectiveness, speediness and timesaving performance of abnormal data detection have been verified and validated to LINKCAFE system successfully. The abnormal data detection results could be obtained within only a few minutes. This brings highly trustiness of LIKKCAFE system among clients in the food industry. Also the development of applications provides a methodology of online system management for abnormal data detection to system managers.

## References

1. Universal Solution Systems, <http://www.u-s-systems.com/>, 2005.
2. Hitachi Information Systems, <http://www.bistromate.com/>, 2004.
3. Alphax Food System, <http://www.afs.co.jp/product/>, 2005.
4. S.Y.Jing, Z.Wang, S.H.Yu and C.D.Lu, "Computer-aided industrial design system based on application service provide", *Computer Integrated Manufacturing Systems*, 10(10), 1184-1190 (2004).
5. M.A.Smith and R.L.Kumar, "A theory of application service provider (ASP) use from a client perspective", *Information and Management*, 41(8), 977-1002 (2004).
6. Z.X.Yang, X.B.Yuan, Z.Q.Feng, etal: "A Fault Prediction Approach for Process Plants using Fault Tree Analysis in Sensor Malfunction", *Proceedings of IEEE ICMA2006*, 3, 2415-2420(2006).
7. J.Schmuller, *Sams Teach Yourself UML in 24 Hours*. In: Li H., Zhao L.G.: tran. 3rd Edition. Beijing: Posts & Telecom Press (2004).
8. I.Sommerville, *Software Engineering*, 7th Edition. New York: Pearson Education Limited (2004).
9. S.Urman, Oracle9i PL/SQL Programming. Tokyo: McGraw-Hill (2001).
10. M.Gurry and P.Corrigan, Oracle Performance Tuning. 2nd Edition. Paris: O'Reilly (1998).