ANADOLU ÜNİVERSİTESİ SINAV HİZMETLERİ SINAV DEĞERLENDİRME OTOMASYONU

Ali YÜREKLİ

Yüksek Lisans Tezi

Bilgisayar Mühendisliği Anabilim Dalı

Temmuz, 2013

JÜRİ VE ENSTİTÜ ONAYI

Ali Yürekli'nin "***Anadolu Üniversitesi Sınav Hizmetleri Sınav Değerlendirme Otomasyonu"*** başlıklı **Bilgisayar Mühendisliği** Anabilim Dalındaki, Yüksek Lisans Tezi 27.06.2013 tarihinde, aşağıdaki jüri tarafından Anadolu Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin ilgili maddeleri uyarınca değerlendirilerek kabul edilmiştir.

	Adı-Soyadı	İmza
Üye (Tez Danışmanı):	**Yrd. Doç. Dr. ÖZGÜR YILMAZEL**	**.....**
Üye :	**Yrd. Doç. Dr. CİHAN KALELİ**	**.....**
Üye :	**Yrd. Doç. Dr. KAMİL ÇEKEROL**	**.....**

Anadolu Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun …… tarih ve …… sayılı kararıyla onaylanmıştır.

Enstitü Müdürü

ABSTRACT

Master of Science Thesis

EXAM EVALUATION AUTOMATION FOR ANADOLU UNIVERSITY EXAMINATION SERVICES

Ali YÜREKLİ

Anadolu University
Graduate School of Sciences
Computer Engineering Program

Supervisor: Assist. Prof. Dr. Özgür YILMAZEL
2013, 48 pages

Anadolu University, performing examination services perfectly for years, is a reliable, worldwide institution in open and distance education. It has been servicing millions of students in domestic and foreign exam organizations.

The increasing number of exam organizations and student capacity, and the advanced print options has raised the need for improvements and updates in evaluation of examinations. An alternative, reliable, complete automation is needed as well as the existing systems.

The present thesis covers the development of an evaluation system, which planned to be used in exam organizations of Anadolu University. The primary concentration of the study is to cover the requirements of evaluation stages in a reliable, error-free, and efficient way. While evaluating the performance and success of the system, previous examinations data were collected and used as test collection. The official student grades and the results produced by the system were compared. The improvements introduced by the use of the system were analyzed, and the performance effects during the whole process were investigated. In the light of the findings, a roadmap to deployment to the live system was obtained.

Keywords: Open education, distance education, examination services, and evaluation

ÖZET

Yüksek Lisans Tezi

ANADOLU ÜNİVERSİTESİ SINAV HİZMETLERİ SINAV DEĞERLENDİRME OTOMASYONU

Ali YÜREKLİ

Anadolu Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Yrd. Doç. Dr. Özgür YILMAZEL
2013, 48 sayfa

Anadolu Üniversitesi, uzaktan eğitim ve açık öğretim dallarında yıllarca kusursuz hizmet vermiş, dünya çapında marka olmuş, güvenilir bir kurumdur. Yurtiçi ve yurtdışında yürüttüğü sınav organizasyonlarıyla milyonlarca kişiye eğitim hizmeti sağlamaktadır.

Genişleyen organizasyon sayısı, öğrenci kapasitesi ve baskı sistemleri ile sınav organizasyonlarının değerlendirme sürecinde iyileştirmelere ve güncellemelere ihtiyaç doğmuştur. Var olan mevcut sistemlere alternatif, güvenilir ve kapsamlı bir otomasyon geliştirilmesi gerekmektedir.

Bu tezde, Anadolu Üniversitesi'nin sınav organizasyonlarında kullanılması planlanan, değerlendirme aşamalarındaki gereksinimleri hatasız, güvenilir ve hızlı bir şekilde karşılayacak otomasyon sisteminin geliştirilmesi ve kullanıma hazır hale getirilmesi amaçlanmıştır. Sistemin performansı ve başarısı değerlendirilirken, eski sınavlara ait verilerden yararlanılmış, sistemde üretilen veriler ile bu resmi veriler karşılaştırılmıştır. Geliştirilen otomasyonun kullanımıyla sağlanan iyileştirmeler incelenmiş, değerlendirme sürecindeki performans etkileri analiz edilmiştir. Elde edilen bulgular sonucunda, canlı sisteme geçiş haritası çıkartılmıştır.

Anahtar Kelimeler: Açık öğretim, uzaktan eğitim, sınav hizmetleri, değerlendirme

ACKNOWLEDGEMENTS

I would like to thank my advisor Asst. Prof. Dr. Özgür Yılmazel for his guidance and support during my study. It was a great pleasure to work with him during this study.

Ali Yürekli

July 2013

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AUCRAC	: Anadolu University Computer Research and Application Center
AUES	: Anadolu University Examination Services
AUPP	: Anadolu University Printing Press
AUSA	: Anadolu University Student Affairs
AUTRU	: Anadolu University Test Research Unit
FTP	: File Transfer Protocol
OCR	: Optical Character Recognition
OEF	: Open Education Faculty
OMR	: Optical Mark Recognition
UML	: Unified Modeling Language

1 INTRODUCTION

Anadolu University has been offering open and distance education at the tertiary level since 1982. The open and distance education system provides higher education opportunity to a great number of students and it plays an in important role in Turkish education system. Today, the system has over 1,700,000 students and 1,500,000 graduates [1].

The success and experience in open and distance education through the years has led Anadolu University to take responsibility in organizing institutional examinations. Many Turkish institutions in banking, government, and education have been working cooperatively with Anadolu University in order to organize exams. These examinations are mostly applied in order to choose candidates for employment, rank the candidates for employment, or decide promotions for higher positions.

Organization of an exam consists of many different steps that require cooperation of different units such as Anadolu University Student Affairs (AUSA), Anadolu University Test Research Unit (AUTRU), Anadolu University Printing Press (AUPP), and Anadolu University Examination Services (AUES). The evaluation process is the responsibility of AUES that is a subunit of Anadolu University Computer Research and Application Centre (AUCRAC).

In this thesis, we construct evaluation system automation for the examination services of Anadolu University. The tool has been designed to provide a secure and useful environment to progress the evaluation steps of examinations held by the university. It also concentrates on distributing the workload between different units having participated in evaluation steps. The performance of the system was evaluated in terms of three fundamental bases. These bases are correctness of the results, total amount of time for an evaluation, and the ease of use provided for the users. Firstly, a collection of ground data was created from previous examinations. When determining these exams, a various criteria such as student volume, exam characteristic, and session count have been taken into account. Secondly, the Optical Mark Reader (OMR) data files of these exams were used as the test collection. The

official student results were compared with the results calculated by the evaluation management tool. After validating the correctness of results, the parallelism among the evaluation steps was investigated. The amount of time per each exam evaluation was measured. Finally, the distribution of the workload and the effects in employee performance were surveyed and analyzed.

The organization of this thesis is as follows. Section 2 gives a background of the general concepts about the open and distance education in Anadolu University. It also describes the role of examination services in handling and managing the exams done by the university. Section 3 presents the details of our evaluation management tool including the system architecture, functionalities of the software, ground data construction, and the results introduced as the outcomes of the system. In section 4, concluding remarks and the future work are given.

2 BACKGROUND

Open and distance education is an alternative way of delivering instruction and education to the students who are not physically present in a traditional setting [2]. In this methodology, the source of information and the learners are separated by time and distance [3].

Although open and distance education has a long history, the advances in technology and computer science have increased the popularity and use of exponentially. Today, synchronous and asynchronous learning methods are available via educational television, audio and video recordings, message board forums, web conferencing, webcasts and webinars [4].

Open and distance education is a great option for many people who wish to get higher education. It creates the potential to equalize access to education and allows the internalization of learning opportunities. Since the students can determine time and place of class time, the system comes up with great flexibility. Compared to classroom-based tuition, distance education is more affordable in terms of enrolment fees.

2.1 Exam Services and Organizations in Anadolu University

2.1.1 Open and Distance Education

Turkey is one of the countries that have the greatest proportion of university students enrolled in distance education. Authorized as the national distance education provider in 1981, Anadolu University has been offering open and distance education at the tertiary level since the early 1980s. Open and distance education system not only serves students in Turkey but Turkish communities in the European Union and Northern Cyprus. Currently, the system has over 1,700,000 students and 1,500,000 graduates.

Open and distance education in Anadolu University addresses massive educational, economic and logistical challenges of serving huge numbers of students. The system continuously adopts new technologies and employs a mix of narrative media, interactive media, adaptive media, communicative media and productive media [5].

2.1.2 Institutional Services

Many institutions in Turkey require examination organizations and they look for partners that can cooperate in these organizations. The institutions in banking, education, and government search for new employees or promote existing employees. They need to filter and rank the candidates, so they organize examinations, which satisfy their requirements. Since the eliminations are critical for many people, the organizations must be reliable, secure, successful and satisfactory. As the number of candidates increases, the organizations become harder and complex to fully qualify the expectations.

Anadolu University plays an important role in organizing institutional examinations and cooperates with many institutions in Turkey. It offers reliable examination services to the institutions and helps in organizing big scale examinations. Each year, Anadolu University organizes more than ten examinations having a total capacity more than 100,000 candidate.

2.1.3 Anadolu University Examination Services

AUCRAC is the center of information technologies in Anadolu University. Besides being a computer research center, it also supports open and distance education by developing software systems and supplying technical infrastructure.

AUES, one of the subunits in AUCRAC, is a specialized unit in organizing examinations. This unit cooperates with different units such as printing press, test research center, logistics and student affairs.

In an organization process, student, course and booklet information is collected by AUES. By scheduling this data, the exam schema and sessions are created. While generating and printing exam documents such as answer sheets, student lists, building lists; exam employee information is collected from exam center offices. This unit manages logistic and economic procedures and it makes possible the sessions to be applied at the announced dates. At later stages of organization process, the evaluation of grades and the announcement of results are handled. The evaluation is done by AUES, and then the results are submitted to AUSA.

2.2 Optical Mark Recognition

2.2.1 OMR Background

The process of capturing human-marked data from document forms is called optical mark recognition. In this process, a scanner device shines a beam of light onto the form paper. The marked areas are detected by the contrasting reflectivity at predetermined positions on the form as they reflect less light than the blank areas of the form [6].

In every session of an exam organized by Anadolu University, an answer sheet form is printed for each student. Backup answer sheet forms are also printed with respect to some criteria. After the sessions are completed, these forms are collected in AUCRAC. A special subunit in AUCRAC feeds the scanners with answer sheets, and then OMR software prepares the data files for evaluation of student grades.

2.2.2 OMR Software

The OMR software makes the recognition possible by using an image scanner to process printed forms. In the evaluation system of AUES, the software recognizes predefined fields, and it creates a data line for each form feeding the scanners. Data lines are stored in data files for each examination center for a session of the exam.

In Figure 2.1, a sample answer sheet form is given. This form is currently in use and it stores two types of information, one of which is understandable by students, and the other one is understandable by the OMR software. Human understandable fields are necessary to give information to the people participating in the examination process. These fields contain information about the name and date of the exam, the student who owns the answer sheet, the booklet code, course names, additional warnings, and etc. OMR understandable fields are necessary to feed evaluation applications with usable data. The fields that must be recognized by the software are summarized at Table 1.

Table 2.1. Pre-defined fields of the answer sheet form

NAME	FILLED BY	DESCRIPTION
Session No	Printing Software	Identifies the session that the student takes place.
Binary Student No	Printing Software	Identifies the student who owns the form.
Binary Room No	Printing Software	Identifies the room that is assigned for the owner student.
Binary Desk No	Printing Software	Identifies the desk that is assigned for the owner student.
Student No	Student	Identifies the student who uses the form.
Absence Flag	Instructor	Filled when the owner student is absent for the session.
Backup Flag	Printing Software	Indicates that the form is a backup form.
Test Group	Student	Identifies the test group of the student who uses the form.
Answer Field	Student	Shows the answers given by the student who uses the form.

2.2.3 OMR Drawbacks

Beside the capabilities and the functionalities of OMR systems, there are also some drawbacks and limitations.

The relationship between printed forms and OMR software is the first drawback. OMR software is dependent to the graphical design of paper forms. Different types of forms require different programs for recognition process. If someone needs to create a new survey form, he should design the form first, and then he should develop the application suitable for that form.

Second drawback occurs when dealing with big data. When large amount of forms are processed, OMR complicates the data collection. In the scanning process, it becomes possible to duplicate, miss, spoil, or skew data [7].

Spoiling the mandatory fields for evaluation introduces a dangerous, vital problem. This problem will be investigated in later sections of this thesis.

The third drawback, increase in total time of scanning process is the outcome of spoiling data. In order to gather error-free data files, the scanning process is done twice for each form and the results are stored in separate files. If there exists differences for the same answer sheet, the more accurate scanning is treated as the primary scanning.

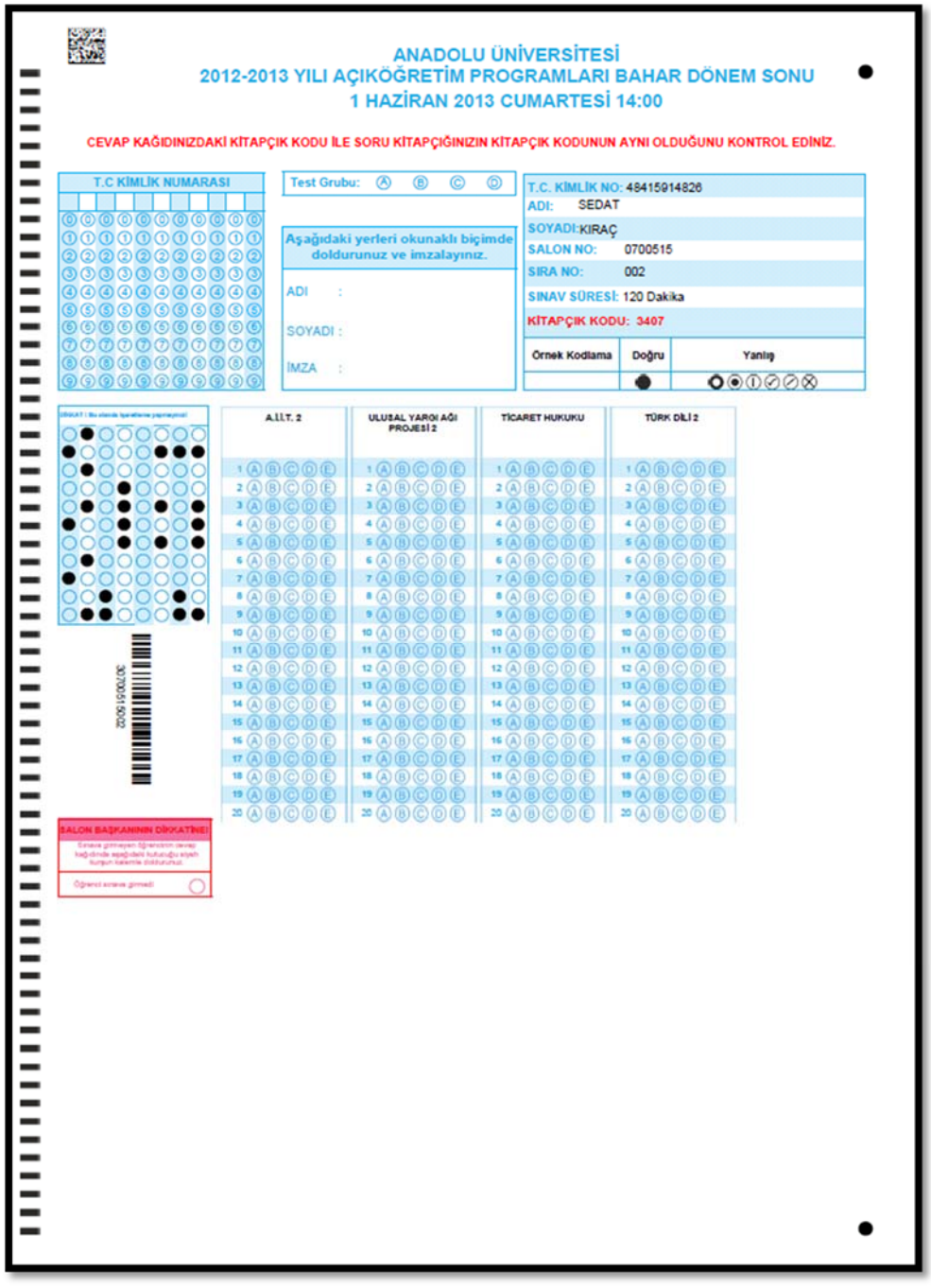

Figure 2.1. Personalized answer sheet form

2.3 Examination Architecture

An examination consists of a number of sessions. The number of sessions can vary from 1 to n. Generally; students can take place in each of the sessions with respect to booklet and course characteristics. The sessions are applied in examination application centers. The application centers have various types of exam buildings, and the buildings contain many exam rooms. In each session, different application centers can be used according to student data. In Figure 2.2, the structure of an exam is illustrated.

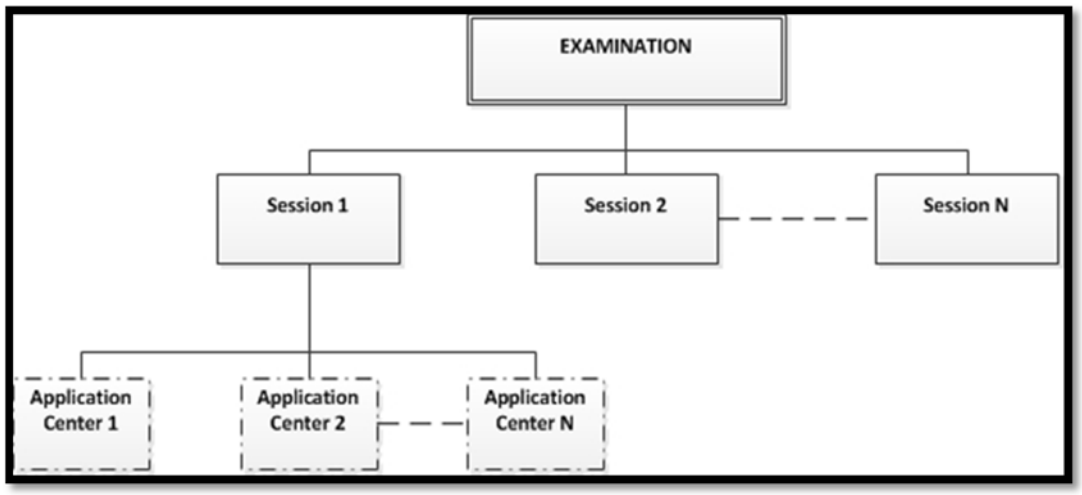

Figure 2.2. Illustration of an exam schema

One of the previous, big scale examinations was applied in January 2013. Having 4 sessions, 91 unique application centers, the statistics for this exam are shown in table below.

Table 2.2. Statistics of a big scale examination

Session Number	Application Centers	Examination Buildings	Examination Rooms	Registered Students
1	91	3186	53954	1112286
2	91	2984	50153	1042077
3	91	1802	28568	627592
4	91	2023	32836	711260

2.4 Related Work on OMR Based Evaluation

In the literature, there have been various studies about OMR systems including OMR software, OMR based exam evaluation, community surveys, automated attendance, and etc. Currently, there exist many different software applications that provide services for universities, high schools, and institutions that use OMR based systems.

The origin of OMR belongs to optical character recognition (OCR). This process is the conversion of scanned images of handwritten, typewritten or printed text into machine-encoded text [8]. The modern OCR technology has born in 1951 with the invention of robotic reader and writer [9].

OMR technology introduced the ability to read marks that have been made in pre-defined positions on paper forms. The pre-defined positions contain areas that are available for marking. These areas such as bubbles should be filled, and boxes to be checked off in order to recognize the optical marks [10].

The accuracy of recognition has also been a study field for many researchers. Although some OMR systems could not be accurate enough, most systems can reach up to 98% accuracy. High accuracy could be maximized with high quality input [11]. In other words, when the users fill in the bubbles carefully, the accuracy of recognition increases. The accuracy measurement is also going to be analyzed in this thesis. The relationship between data size and the accuracy will be discussed and illustrated.

3 EXAM EVALUATION AUTOMATION

Exam evaluation automation for Anadolu University is a system currently in development process, which aims to fully qualify the requirements of evaluation processes of examination services provided by Anadolu University. Providing functionalities for different user groups, the automation consists of several modules. The software is designed as a central database oriented application, developed with latest Java technologies.

3.1 System Requirements

a. The system should meet the demands of evaluation process of exam services. Supporting usability for different user groups, it must be helpful in distributing the workload of evaluation among the participants in the process.
b. The software must be robust, accurate and reliable since the process is very vital and critical for many people and institutions.
c. The system should be capable of handling different types of examinations. Since the requirements of evaluation and answer sheet design may vary at different exams, the system must provide flexible configurations for optic parser and evaluation criteria.
d. There should be a mechanism to provide communication between user groups and data file transfer operations should be moved to the digital platform. Examination data files must be stored in an accessible server.
e. OMR based errors must be detected and handled by the system. A tagging module should apply the detection and correction operations on the data files. The error types and the solutions should be configures by the system.
f. User actions and events must be tracked. The history of actions and events should be logged. Some reports and documentation can be generated by the system.

3.2 System Architecture

3.2.1 System Architecture Model

In this automation, a central database model is used. The architectural model is shown in Figure 4. A central database on a specific database server maintains data of examinations, students, courses and grades. The system services different types of user groups via modules for each group. A common GUI is providing the functionalities to the users by showing related content for the group of the users. While the administrator users can access all application content, examination services users can access the content except system configurations and developer-related functionalities. OMR users and TRU users can only use OMR and TRU modules.

In addition to a database server, an FTP server is used by the system to read and write OMR data files. These data files are filtered and transferred by OMR module, read back by evaluation module.

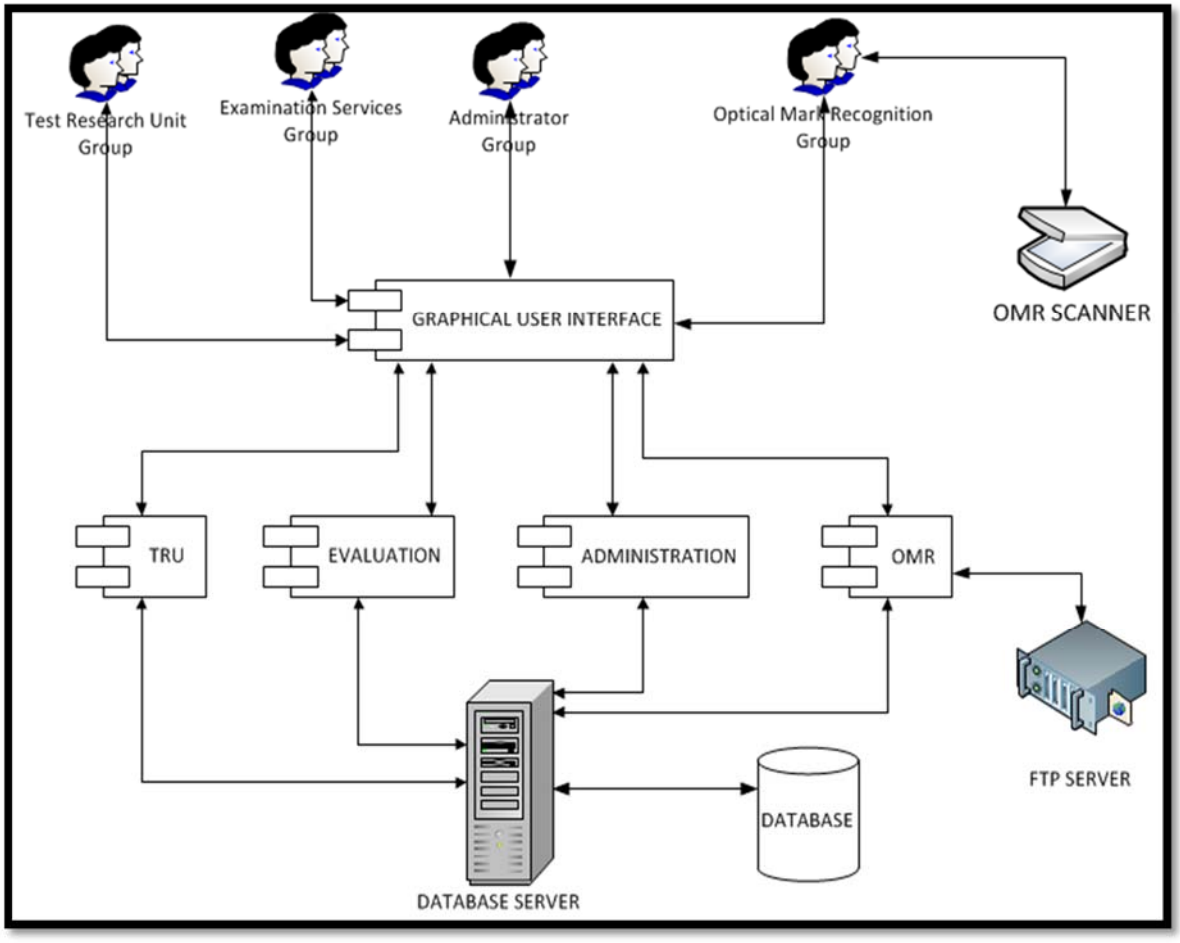

Figure 3.1. The architectural model of evaluation system

3.2.2 User Groups

The system is designed to support 4 different user groups:

a. Examination Services User (ES User): This user group is defined for AUES staff. The users of this group handle examination-based operations such as evaluation, data and process management.

b. Optical Mark Recognition User (OMR User): This user group is defined for AUTRAC staff responsible for OMR based operations.

c. Test Research Unit User (TRU User): This user group is defined for AUTRU staff. Item analysis study and statistical documentation will be the main functionalities provided to the users in this group.

d. Administrator: This user group is defined for administrator and developer staff in AUES staff. System and application based configurations are carried by the users in this group.

By extending an abstract class named AppUser, the user group hierarchy of the system is modeled. All user groups extend this base class, overriding the super constructor. The UML class diagram for user groups is shown in figure below:

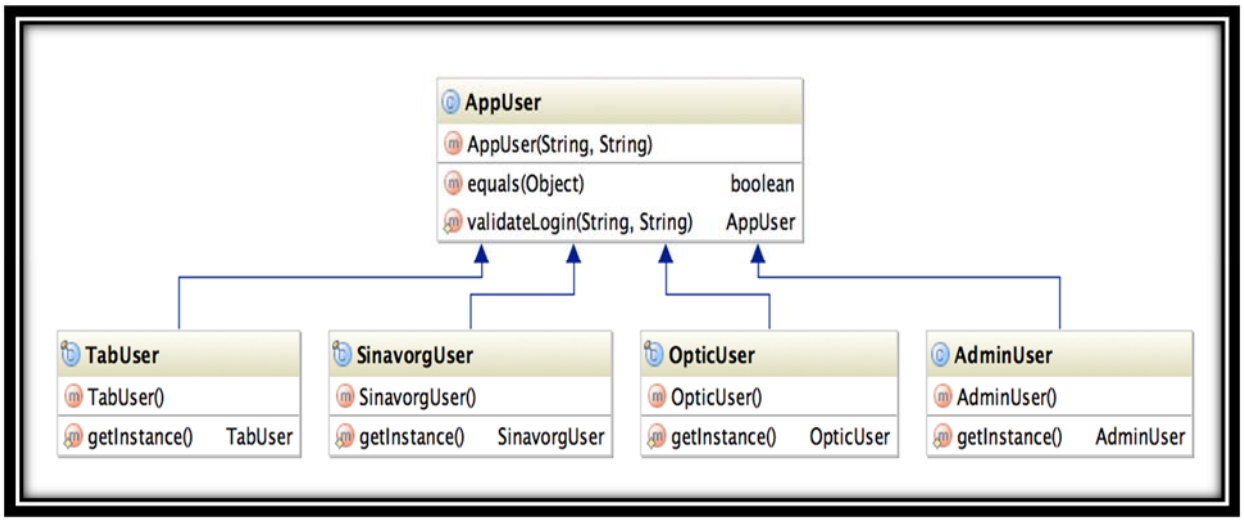

Figure 3.2. UML class diagram for user groups

3.3 System Modules

3.3.1 OMR Module

3.3.1.1 OMR Tagging

In order to prevent the effects of data corruption in evaluation data files, we designed a tagging tool to find solutions for the exceptions thrown by the corrupted data.

Once the user uploads data files, the system detects and stores the exceptions. After uploading, users can list and preview the exceptions. At that stage, users can find preview the suggested solutions for an exception when the exception is clicked to expand for details. User can tag a solution; update an existing solution for each of the exceptions. All of the exceptions must be tagged in order to continue to evaluate the grades of students.

An exception in the system can be solved in some ways. With respect to the reason of the exception, the system suggests a list of solutions. Being configured by the administrator users, the solutions for exceptions are listed below:

a. ***No Solution:*** This solution is the default state for an exception. When an exception is detected and thrown, it is stored with no solution. One of the other solution types should be matched with the exception.
b. ***Ignore:*** When an exception is tagged with this solution, the origin of the exception is ignored. In other words, the corruption in data is negligible.
c. ***Exclude:*** This solution excludes the answer sheet that causes the problem. Corresponding data from the answer sheet is not included in evaluation.
d. ***Accept Binary Student Number:*** When an exception is tagged with this solution, the value from the binary student number will be used in evaluation as the owner of the answer sheet.

e. ***Accept Student Number:*** When an exception is tagged with this solution, the value from the student number will be used in evaluation as the owner of the answer sheet.

f. ***New Student Number:*** This solution requires an entry of a student number.

g. ***Manual Swap Detection:*** This solution detects a swap situation when automatic swap detection algorithm could not be applied. Solution requires an entry for the owner of the answer sheet and an entry for the student that makes the swap.

h. ***Automatic Swap Detection:*** An exception with automatic swap type requires validation of the situation. This solution provides this validation, and it forces to apply swap evaluation algorithm during the evaluation of grades.

i. ***Accept Attended:*** The answer sheet will be accepted as an attended sheet. This solution requires test group information since OMR software do not recognize test group field in nonattendance data type.

j. ***Parse as Normal:*** This solution implies that the answer sheet data will be parsed with normal sheet parsing algorithm. It validates that the type of answer sheet is normal.

k. ***Parse as Backup:*** This solution implies that the answer sheet data will be parsed with backup sheet parsing algorithm. It validates that the type of answer sheet is backup.

3.3.1.2 Functionalities

- The users of the module can transfer OMR data files to a server via FTP. Controlling the pattern of the file names validates the files. For an examination center in a session, 2 data files must be submitted. One file contains the students that attend to the session, and the other file contains the students that do not attend to the session. Files can be uploaded multiple times.
- While uploading data files, data corruption is analyzed and exceptions are thrown. A menu lists these exceptions. Besides detecting the exceptions, missing answer sheets are also detected. A menu lists these missing sheets.
- Users can relate solutions with exceptions by the OMR tagging tool. Tool allows reverting or submitting the relations. Applying a solution may change the status of a missing answer sheet.
- Users can access examination information of a student by querying the student with his student number.
- Users can access details such as name, definition and description of exception types and solution types.
- User can track information about the uploaded examination files.

3.3.1.3 Problem Solution Map

When users try to find a solution for an exception, the system suggests possible solutions from the problem solution map. The administrator users create this map and they can modify it. However, modification of this map will require changes in implementation of the code. Thus, it is suggested to assign a modification issue to a system developer. In Table 3, the problem solution map is illustrated.

Table 3.1. Problem solution map

EXCEPTION TYPE	**SOLUTION TYPE**
Session Mismatch	Ignore
Session Mismatch	Exclude
Exam Center Mismatch	Ignore
Exam Center Mismatch	Exclude
Student Number Mismatch	Accept Binary Student Number
Student Number Mismatch	Accept Student Number
Student Number Mismatch	Manual Swap Detection
Attendance Uncertainty	Ignore
Attendance Uncertainty	Accept Attended
Student Not Found	New Student Number
Student Not Found	Manual Swap Detection
Auto Detected Swap	Automatic Swap Detection
Backup Student Number Mismatch	Exclude
Backup Student Number Mismatch	New Student Number
Backup Control Needed	Parse as Backup
Backup Control Needed	Parse as Normal

3.3.2 Administration Module

3.3.2.1 Optic Parser Configuration

Previously we reported the relationship between the design of answer sheet forms and the OMR software. OMR software produces data files that store the data extracted from the answer sheets. A generic parser for data files can handle different answer sheet designs; which bring flexibility to the procedure in designing the forms and modifying the OMR software.

In order to achieve the flexibility mentioned above, we designed a generic optic parser, which should be supported with a configuration. The main idea of the design is to force the configuration to provide necessary fields for evaluation. The pre-defined fields shown in Table 1 are assumed mandatory for evaluation. While configuring the parser, one should define the starting point and the length for each of the mandatory fields.

Table 3.2. Some configurations for optical parser

Field / Configuration	Default Configuration	Sample Configuration
Session No	<61, 1>	<2, 2>
Binary Student No	<40, 11>	<10, 4>
Binary Room No	<51, 7>	<4, 3>
Binary Desk No	<58, 3>	<8, 3>
Student No	<64, 11>	<14, 4>
Absence Flag	<75, 1>	<18, 1>
Backup Flag	<76, 1>	<20, 1>
Test Group	<63, 1>	<22, 1>
Answer Field	<77, 200>	<25, 20>

When the optic parser runs with sample configuration, it parses sample data *WY458GFNW72HFH377VHAH2348THG48GYUJW8AS4Q842775* as Session No = U4, Binary Student No = 710G, Binary Room No = 8DH, Binary Desk No = Y27, Student No = HBVN, Absence Flag = 3, Backup Flag = F, Test Group = W, Answer Field = 58GFNW72HFH377VHAH23.

3.3.2.2 Evaluation Configuration

The variety of examinations leads to different requirements for evaluation of exams. In order to handle these requirements in evaluation, we designed a generic configuration mechanism. This mechanism enables to apply some rules on procedures used while filtering OMR data corruptions and evaluating course grades. The items in configurations are listed below:

a. ***Cancelled Question Behavior:*** A question in a course can be cancelled due to ambiguity or contradiction. In this case, there are 2 main approaches. The question can be accepted true for all students, or it can be excluded from evaluation. This configuration item sets the approach for a cancelled question.

b. ***Exempted Question Behavior:*** A question in a course may be exempted for visually impaired students. In this case, there are 2 main approaches. The question can be accepted true for all visually impaired students, or it can be excluded from their evaluation. This configuration item sets the approach for an exempted question.

c. ***Fault Tolerate in Student No:*** While the optic parser checks the equality of the binary student number and student number, it uses this configuration item as a parameter to reduce and filter data corruptions. This item can take values between 0 and 5. The mismatching characters up to the selected value are tolerated, which reduces the number exceptions that need to be solved.

d. ***Space Tolerate in Student No:*** One of the common mistakes that students do is to forget to fill the student number. When a student number is not filled, the predefined area displays as a space. This configuration item is used when counting spaces in optic parser student control; it can take values between 8 and 11.

e. ***Precision Display Value:*** When calculating and announcing grades, a precision display value is used to provide flexibility for different examination requirements. This configuration item can take values between 0 and 3.

f. ***Precision Rounding Model:*** When calculating grades, a mathematical rounding model is used to provide flexibility for different examination requirements. The item can take values in Java RoundingMode mathematical library, namely CEILING, FLOOR, DOWN, HALF_DOWN, HALF_UP, HALF_EVEN, UNNECESSARY, and UP.

g. ***Backup Sheet Evaluation Model:*** In some examinations, due to booklet specifications, it is not possible to evaluate a backup answer sheet without physical control of course answers text parsing. This configuration item can be AUTOMATIC and MANUAL. In automatic mode, it is not required to check the sheets. In manual mode, it is obligatory to check the sheets, so these sheets are skipped by the application.

3.3.2.3 Functionalities

- New configurations can be defined for optic parser. This ability provides flexibility for new visual designs of answer sheet forms. The dependency between answer sheet forms and the OMR software is reduced to optic configuration level.
- New configurations can be defined for evaluation. This ability provides flexibility to meet demands of institutions about evaluation criteria.
- The types of OMR data corruption and solutions for these types can be configured. However, the changes in configuration may require updates at the implementation of the code. Thus, this functionality is available for only developers.
- Administration module is only available for administrator user group.

3.3.3 Evaluation Module

3.3.3.1 Solution Extraction

The smallest data chunk of evaluation is an examination center of an exam session. As the scanning processed is doubled for better OMR results, the smallest data chunk is also doubled. The data files are maintained with respect to the chunk format and the evaluation is applied for both the first scanning and the second scanning. The scanning processes are independent from each other, thus the exceptions are stored with the session, examination center and scanning number.

While uploading the data files to the system and reading the uploaded files for evaluation, the same algorithm of the optic parser is used. The answer sheet is parsed into smaller sections, the pre-defined fields are constructed from these sections, and finally these fields are tested in terms of correctness rules. The use of the same algorithm in both stages results in detecting the same problems for an answer sheet. As the system does not allow evaluating an answer sheet having an exception with no solution, it is guaranteed that every problem has one unique solution.

Having guaranteed that all exceptions have one unique solution, the solution extraction becomes possible. When reading data files for evaluation, the optic parser detects the same exceptions with the ones detected and stored at data upload process. The solutions belonging to the working data chunk are retrieved from the database and matched with the answer sheets by using the concatenation of binary student number and student number fields. If an answer sheet has one or more exceptions, it is forced to get the tagged solutions and apply procedural behaviors with respect to the solution types.

3.3.3.2 Functionalities

- Users can arrange the evaluation parameters for an examination by selecting a predefined configuration from the list of configurations created by administrators.
- Users can arrange the optic parser parameters for an examination by selecting a predefined configuration from the list of configurations created by administrators.
- Users can manage the evaluation stages of the answer sheets as they are uploaded to the system. When all the exceptions are matched with a solution, the system allows evaluating that upload.
- The system provides the users with the ability to apply the evaluation of cancelled questions. The procedure is applied incrementally. When a question is cancelled and defined to the system, it can be evaluated without any dependency.
- Users can try and examine different precision rounding models by selecting the model and the precision display value. Furthermore, mathematical descriptions of the models are provided in this functionality.
- Some students may be treated as temporarily virtually impaired for an examination. These students must be defined to the system before the evaluation. The system allows the users add/remove temporarily visually impaired students.
- Users are provided with the ability to clear all data related with an examination.
- The users can list the swap situations detected and granted in an examination.
- After the announcement of grades, some students may object to the grades via official letters of application. If the objection is due to a test group error, the grades of the student are re-calculated according to another test group. The users can manage this process and re-evaluate a student.

3.3.3.3 Examination Collections

An evaluation application needs certain information of answer keys, cancelled questions, exempted questions, and visually impaired students. In our system, we represent these data types as custom collections. By implementing factory methods, we provide a standardized approach to load, retrieve, and use these collections.

When an evaluation application is initialized, the identification number of exam is retrieved, and the necessary collections are created and loaded automatically. The inheritance between evaluator classes and the abstract evaluator provides this flexibility to system developers.

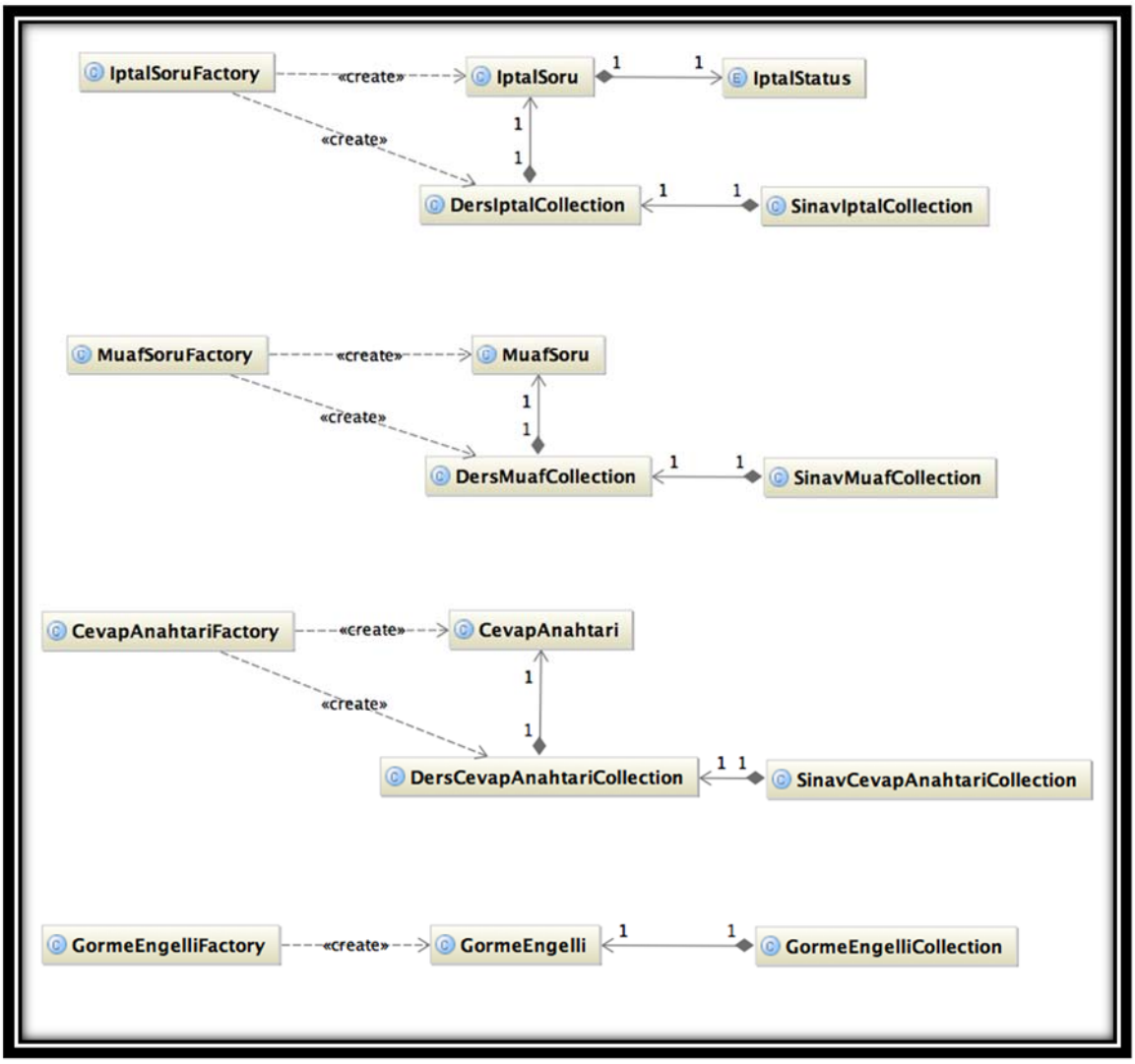

Figure 3.3. UML class diagram for collections in an exam

a. ***Cancelled Question Collections:*** *IptalSoru* represents a cancelled question in a course. An enumeration named *IptalStatus* maintains the current status of the question. The cancelled questions for a course are stored in collections named *DersIptalCollection*. *SinavIptalCollection* maintains all the collections for courses in the examination. *IptalSoruFactory* is the factory that creates and loads every data type about cancelled questions in an examination.

b. ***Exempted Question Collections:*** *MuafSoru* represents an exempted question for visually impaired students in a course. The exempted questions for a course are stored in collections named *DersMuafCollection*. *SinavMuafCollection* maintains all the collections for courses in the examination. *MuafSoruFactory* is the factory that creates and loads every data type about exempted questions in an examination.

c. ***Answer Key Collections:*** *CevapAnahtari* represents an answer key for a test group in a course. The instances of this data type are stored in collections named *DersCevapAnahtariCollection*, each of which is related with a single course in the examination. *SinavCevapAnahtariCollection* maintains all the collections of answer keys for courses in the examination. *CevapAnahtariFactory* is the factory that retrieves all answer keys, creates CevapAnahtari instances, and stores them in related collections.

d. ***Visually Impaired Student Collections:*** *GormeEngelli* represents a visually impaired student who is not responsible of exempted questions in an examination. *GormeEngelliFactory* is the factory that retrieves all visually impaired students, creates *GormeEngelli* instances, and stores these instances in a collection named *GormeEngelliCollection*.

3.3.3.4 Evaluator Model

In order to conduct answer sheet evaluation for an examination center of a session, there must be no unresolved problems or missing answer sheets for the current examination center. ES user accesses to the data files uploaded via FTP, and then he triggers the evaluation request for these files. Each line in data files are processed and converted to student data.

Cancelled question evaluation is applied after finalizing answer sheet evaluation. ES user retrieves the cancelled questions and triggers the evaluation request for those questions. The system gathers the students whose final grades will change after the evaluation of cancelled questions. Student grades are re-calculated and updated by applying the cancelled question behavior.

After the announcement of grades, official objections are delivered to AUCRAC. ES user can he can re-evaluate and update student grades if the objections are related with incorrect test groups.

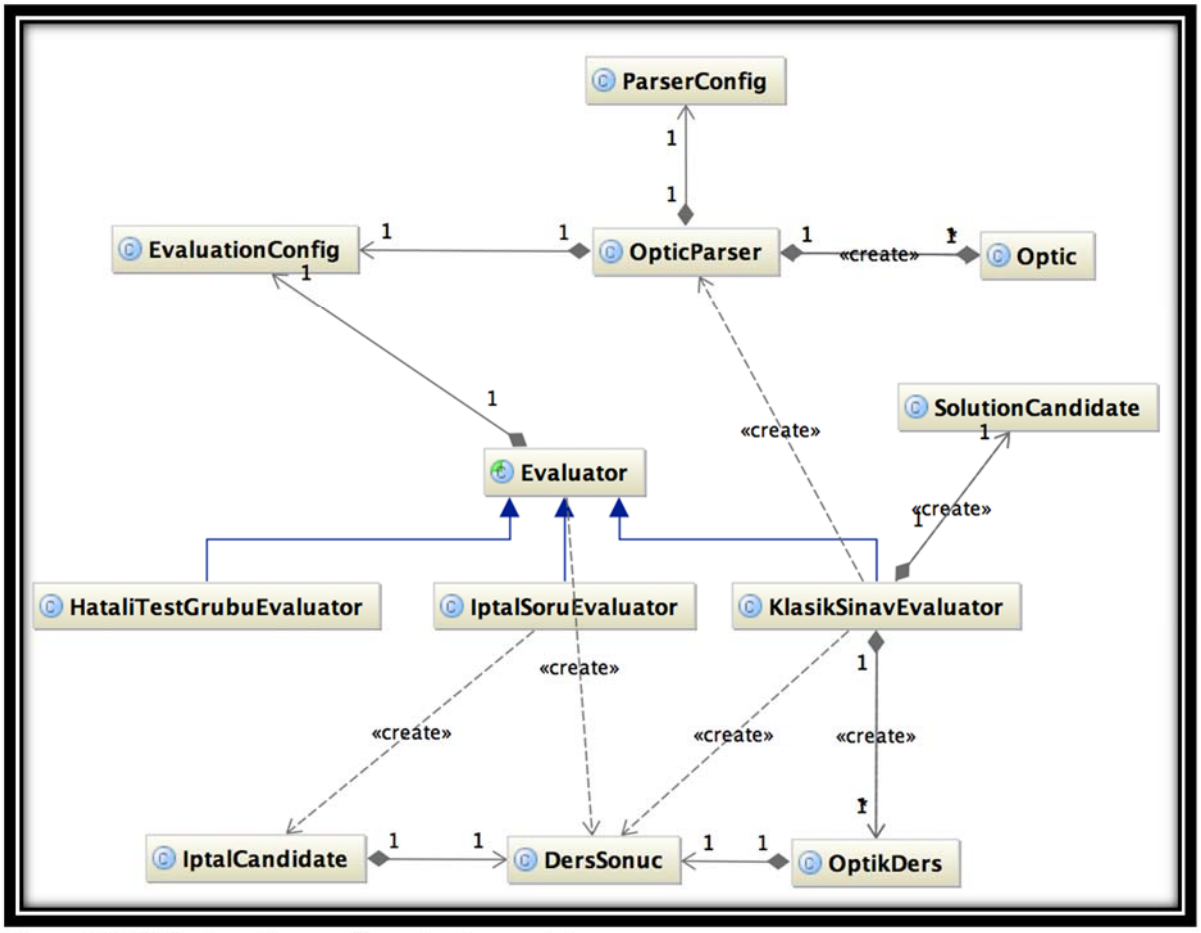

Figure 3.4. UML class diagram for evaluation model

In order to perform the operations described above, we designed a model that can be extended with specific functionalities. We created an abstract class *Evaluator* that immediately loads exam collections when extended and instantiated by a specific evaluator. Furthermore, it creates and loads the evaluation configuration represented by *EvaluationConfig*. *Evaluator* also provides methods of calculating statistics and points of students. Overriding these methods are not allowed, thus all extending classes use the same logic when performing calculation operations.

KlasikSinavEvaluator, which extends *Evaluator*, is responsible of evaluating the answer sheets. It is provided with the ability to connect the FTP server that maintains optical data. It has an embedded *OpticParser* instance that can parse data in optic files with parameters defined in parser configuration. This configuration is represented with *ParserConfiguration*. As the data files are read and processed, *Optic* instances are created. These instances maintain student examination information. *KlasikSinavEvaluator* collects optic instances and match them with *SolutionCandidate* that provide solutions for errors from optic files. Finally, courses are gathered from optics, and then they are associated with result objects named *DersSonuc*. The results are inserted to the database table via a batch insert procedure.

IptalSoruEvaluator, which extends Evaluator, is responsible of evaluating the cancelled questions. It gathers *IptalCandidate* objects that are the student grades affected from the questions. The grades of candidates are re-calculated with the calculation methods defined in super class *Evaluator*. The results are updated in the database table via a batch update procedure.

HataliTestGrubuEvaluator, which extends Evaluator, is responsible of evaluating test group objections. It requests the data of new test group. Bu using evaluation configuration and calculation methods, it re-calculates the grades of a student with new test group data. If ES user approves the evaluation, it updates results in the database table.

3.3.4 TRU Module

3.3.4.1 Item Analysis Study

Item analysis is a method of gauging the quality of an examination by looking at its constituent parts. It seeks to give some idea of how well the examination has performed relative to its purposes. The primary purpose of most examinations in higher education is that of a measurement tool, for assessing the achievements of the examination candidates and thus how future learning can be supported and directed. It is important for academic staff to have an understanding of item analysis - its methods, assumptions, uses and limitations in order that examinations can be assessed and improved [12].

Item analysis study helps TRU staff to investigate the integrity and selectivity of the questions in examinations. The study is done separately for each course in the examination. The participants of the analysis are the students that use test group "A". After retrieving student answers, the set is sorted according to the grades of students. By selecting 27% of the top group, and by selecting 27% of the bottom group, two different data sets are constructed. For each question in the course, the frequencies of multiple-selection based answers are calculated for the data sets. Finally, difficulty (p) and discrimination (r) values are computed. The observations are matched with a suitable comment. All of the findings are reported with addition information such as arithmetic mean, max grade, min grade, standard deviation, skewness, kurtosis, and XY line chart.

Computing difficulty and discrimination:

$$p = \frac{\text{Correct answers from top group} + \text{Correct answers from bottom group}}{\text{Size of top group} + \text{Size of bottom group}}$$

$$r = \frac{\text{Correct answers from top group} - \text{Correct answers from bottom group}}{\text{Size of top group}}$$

Interpreting difficulty and discrimination:

Case 1: The answer of the question is not the same as the dominant answer among the top group.

Case 2: The answer of the question is same as the dominant answer among the top group.

Table 3.3. Lookup table for interpreting item analysis result

CASE	**RANGE**	**COMMENT**	**DISPLAY**
1	Unnecessary	Inverted outcome	TERS ÇALIŞMA
2	$r < 0.2$	Bad	KÖTÜ
2	$0.2 \leq r < 0.3$	Medium	ORTA
2	$0.3 \leq r < 0.4$	Good	İYİ
2	$0.4 \leq r$	Very Good	ÇOK İYİ

Computing additional information:

Arithmetic mean is the sum of a collection of numbers divided by the number of numbers in the collection [13]. Standard deviation shows how much variation or dispersion from the average [14]. Skewness is a measure of the extent to which a probability distribution of a real-valued random variable “leans” to one side of the mean [15]. Kurtosis is any measure of the “peakedness" of the probability distribution of a real-valued random variable [16].

All computations in this category are done with the help of the Apache Commons Mathematics Library. An instance of DescriptiveStatistics class is created. This instance is loaded with student grades, and the statistical values are computed.

A sample item analysis study report is shown in Figure 3.5. Due to confidential issues, examination and course details are hidden at the figure.

DERS KODU :

DERS ADI :

Toplam Aday Sayısı	16775	En Yüksek Puan	100,00	Standart Sapma	17,07
Madde Analizi Aday Sayısı	7629	En Düşük Puan	0,00	Çarpıklık	-0,16
Sınava Girmeyen Aday Sayısı	1514	Ortalama Puan	61,50	Basıklık	-0,55

		A		B		C		D		E		BOŞ		GEÇERSİZ				
SORU	CEVAP	ÜST	ALT	ÜST	ALT	ÜST	ALT	ÜST	ALT	ÜST	ALT	ÜST	ALT	ÜST	ALT	p	r	YORUM
1	D	0,18	0,24	0,03	0,08	0,07	0,08	0,17	0,08	0,52	0,50	0,00	0,00	0,00	0,00	0,12	0,09	TERS ÇALIŞMA
2	E	0,11	0,20	0,06	0,23	0,03	0,14	0,07	0,22	0,71	0,17	0,00	0,01	0,00	0,00	0,44	0,53	ÇOK İYİ
3	C	0,11	0,43	0,00	0,01	0,81	0,31	0,02	0,07	0,03	0,15	0,00	0,00	0,00	0,00	0,56	0,50	ÇOK İYİ
4	C	0,00	0,02	0,00	0,02	0,98	0,84	0,00	0,05	0,00	0,04	0,00	0,00	0,00	0,00	0,91	0,14	KÖTÜ
5	A	0,96	0,74	0,01	0,06	0,00	0,03	0,00	0,05	0,00	0,08	0,00	0,00	0,00	0,00	0,85	0,22	ORTA
6	B	0,06	0,12	0,92	0,57	0,00	0,08	0,00	0,13	0,00	0,06	0,00	0,00	0,00	0,00	0,75	0,35	İYİ
7	D	0,06	0,24	0,03	0,14	0,12	0,21	0,67	0,19	0,09	0,19	0,00	0,00	0,00	0,00	0,43	0,47	ÇOK İYİ
8	B	0,00	0,07	0,82	0,37	0,06	0,24	0,05	0,17	0,04	0,11	0,00	0,02	0,00	0,00	0,59	0,44	ÇOK İYİ
9	C	0,08	0,20	0,01	0,11	0,82	0,37	0,02	0,12	0,03	0,16	0,00	0,01	0,00	0,00	0,59	0,45	ÇOK İYİ
10	B	0,01	0,21	0,95	0,31	0,00	0,20	0,01	0,17	0,00	0,08	0,00	0,01	0,00	0,00	0,63	0,63	ÇOK İYİ
11	B	0,00	0,08	0,97	0,40	0,01	0,28	0,00	0,13	0,00	0,08	0,00	0,01	0,00	0,00	0,68	0,56	ÇOK İYİ
12	D	0,00	0,11	0,00	0,10	0,00	0,20	0,96	0,38	0,02	0,17	0,00	0,01	0,00	0,00	0,67	0,58	ÇOK İYİ
13	D	0,00	0,09	0,00	0,11	0,02	0,34	0,95	0,33	0,00	0,08	0,00	0,01	0,00	0,00	0,64	0,62	ÇOK İYİ
14	A	0,84	0,12	0,01	0,08	0,05	0,30	0,01	0,14	0,07	0,33	0,00	0,01	0,00	0,00	0,48	0,71	ÇOK İYİ
15	C	0,05	0,20	0,00	0,17	0,86	0,27	0,06	0,21	0,00	0,09	0,00	0,02	0,00	0,00	0,56	0,58	ÇOK İYİ
16	A	0,79	0,41	0,07	0,11	0,06	0,15	0,05	0,20	0,00	0,09	0,00	0,01	0,00	0,00	0,60	0,38	İYİ
17	E	0,01	0,07	0,00	0,04	0,00	0,06	0,02	0,17	0,95	0,63	0,00	0,00	0,00	0,00	0,79	0,31	İYİ
18	E	0,01	0,04	0,01	0,08	0,00	0,02	0,00	0,05	0,96	0,78	0,00	0,00	0,00	0,00	0,87	0,18	KÖTÜ
19	A	0,79	0,46	0,01	0,04	0,03	0,08	0,03	0,11	0,12	0,28	0,00	0,00	0,00	0,00	0,63	0,32	İYİ
20	E	0,09	0,13	0,15	0,22	0,09	0,22	0,15	0,21	0,48	0,18	0,00	0,00	0,00	0,00	0,33	0,30	İYİ

Döküm Tarihi : 27/06/2013 1

Figure 3.5. Sample report showing item analysis results

3.3.4.2 Functionalities

- Users can conduct item analysis study and generate reports dynamically. These reports are exported in PDF format and they are stored in users' local machines.
- Users can generate statistical documents related with arithmetical means of course grades. These reports are exported in XLS format and they are stored in users' local machines.
- Report generation procedures can be applied at any time once pre-evaluation of examination is started.

3.3.4.3 Dynamic Document Generation

The primary feature of this module is dynamic document generation. This feature is designed upon a generic model that uses Java concurrency for multi-threaded operations. According to the type of document to be generated, a worker thread is initialized and triggered. For each of the document types, a generator class is implemented to produce documents in PDF or XLS file formats.

Our model consists of an interface, an abstract class and extending classes. *IGenerator* is the interface that defines the main methods in document production. *Generator* is the abstract class that is the super class of generator classes specialized for the document types. This class contains fields and methods common in specialized generators.

ArithmeticMeanGenerator is responsible for generating arithmetic mean documents in XLS format. It creates an instance of *ArithmeticMean* that maintains entities of arithmetic mean study.

ItemAnalysisGenerator is responsible for conducting item analysis study and generate documents in PDF format. For each course in examination, an instance of *ItemAnalysisStudy* is initialized. This class is capable of performing item analysis study for the given course. It retrieves student grades in form of *StudentResult*, and produces question outcomes in form of *QuestionOutcome*. Finally, the generator combines analysis results and converts them to entities of PDF format.

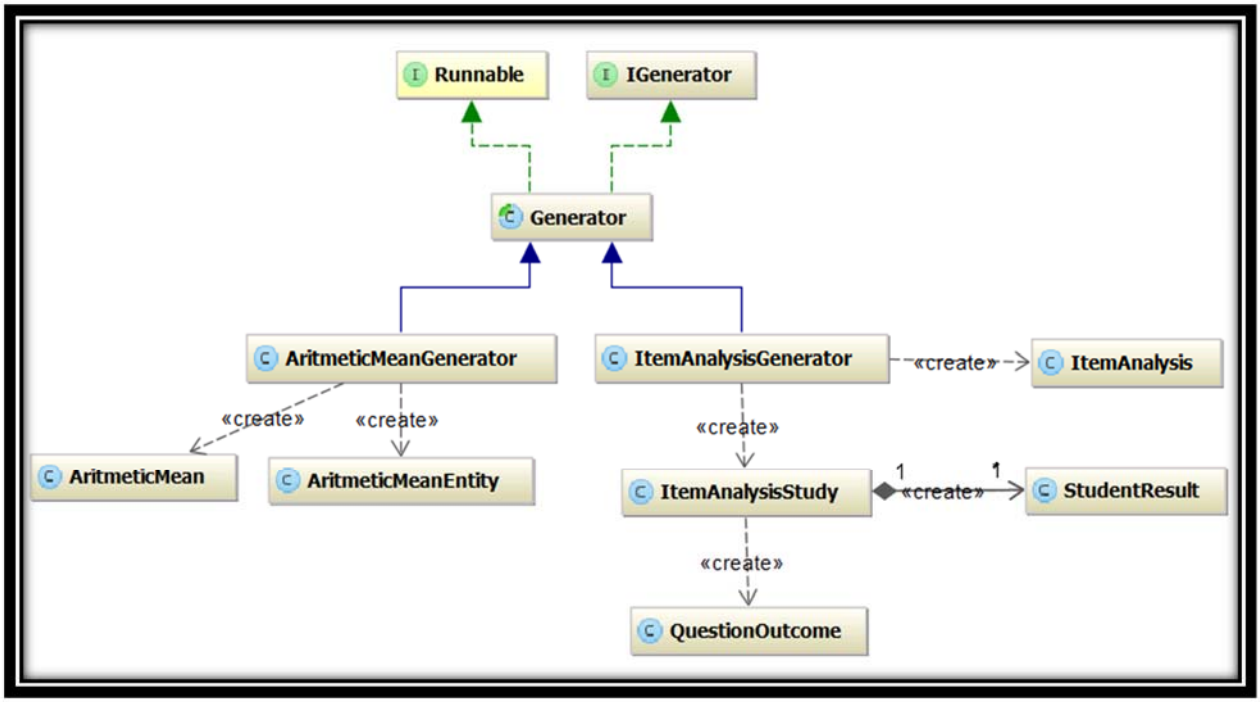

Figure 3.6. UML class diagram for document generation

3.4 User Groups and Module Access

The user groups are allowed to access different modules of the system. The table below shows the accessibility of modules among the different user groups.

Table 3.4. The accessibility of modules among user groups

User / Module	OMR Module	Administration Module	Evaluation Module	TRU Module
ES User	□	□	□	□
OMR User	□	□	□	□
TRU User	□	□	□	□
Administrator	□	□	□	□

3.5 Use Cases

3.5.1 OMR Module Use Cases

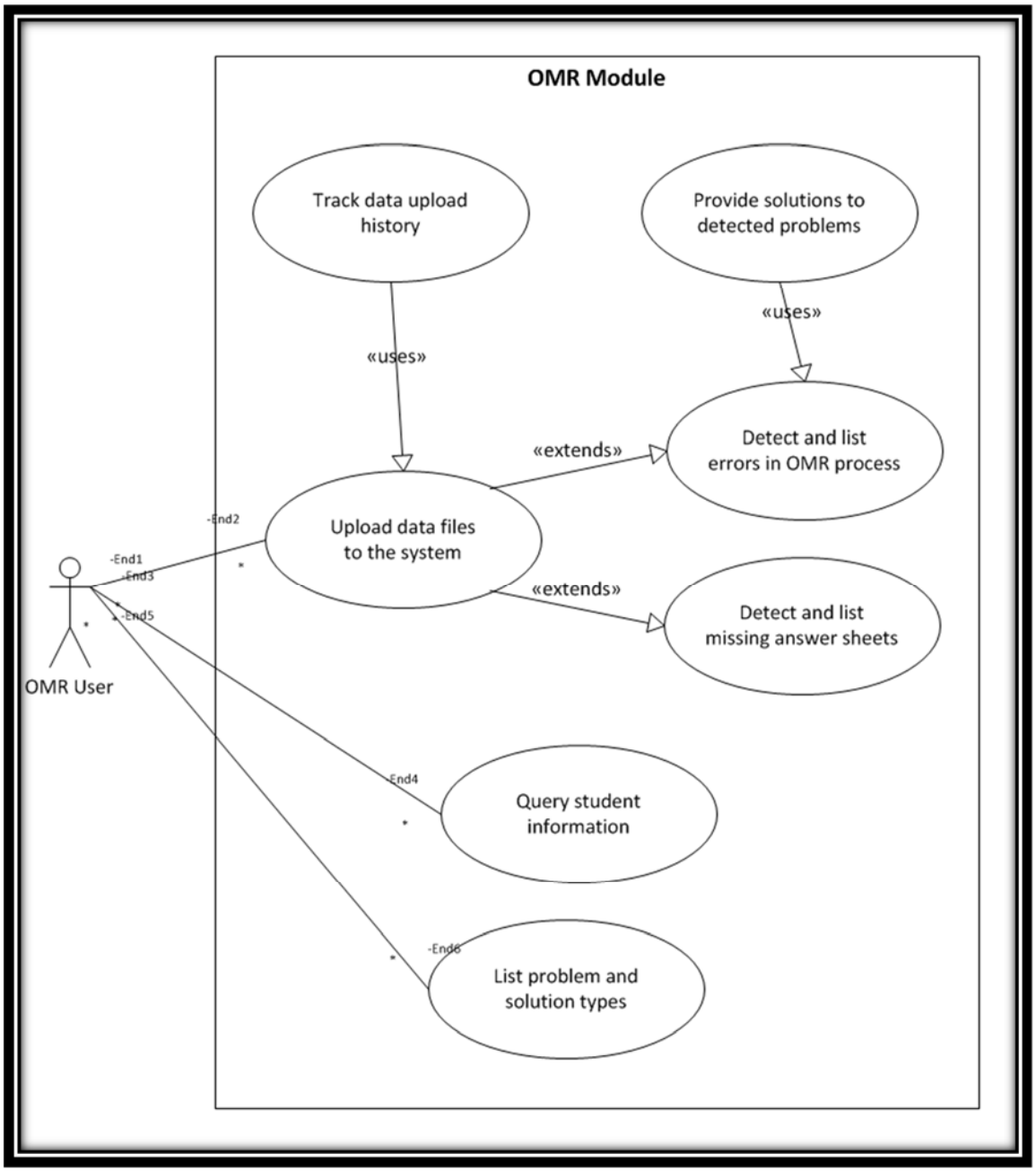

Figure 3.7. Uses case diagram of OMR module

3.5.2 Administration Module Use Cases

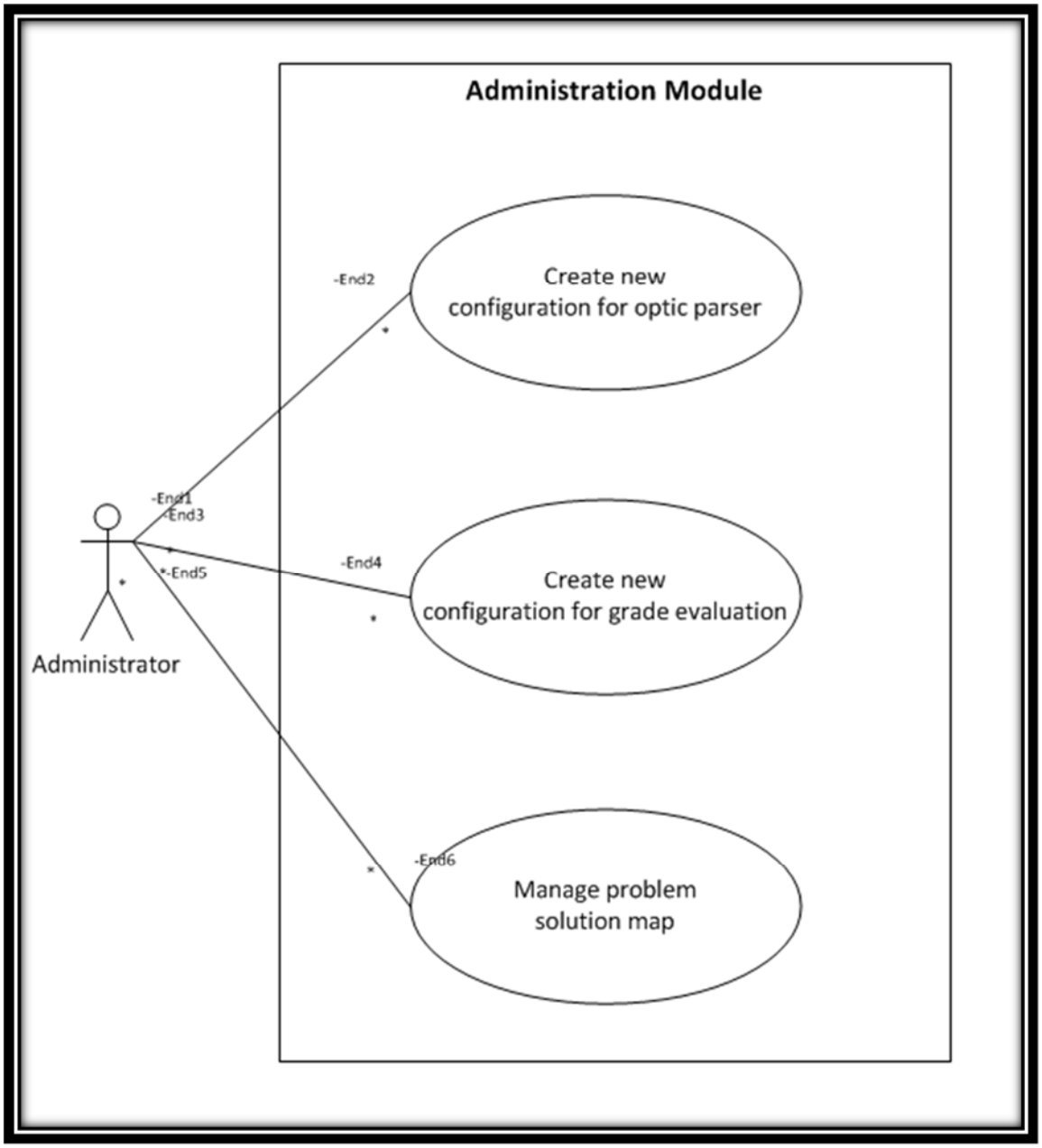

Figure 3.8. Use case diagram of administration module

3.5.3 Evaluation Module Use Cases

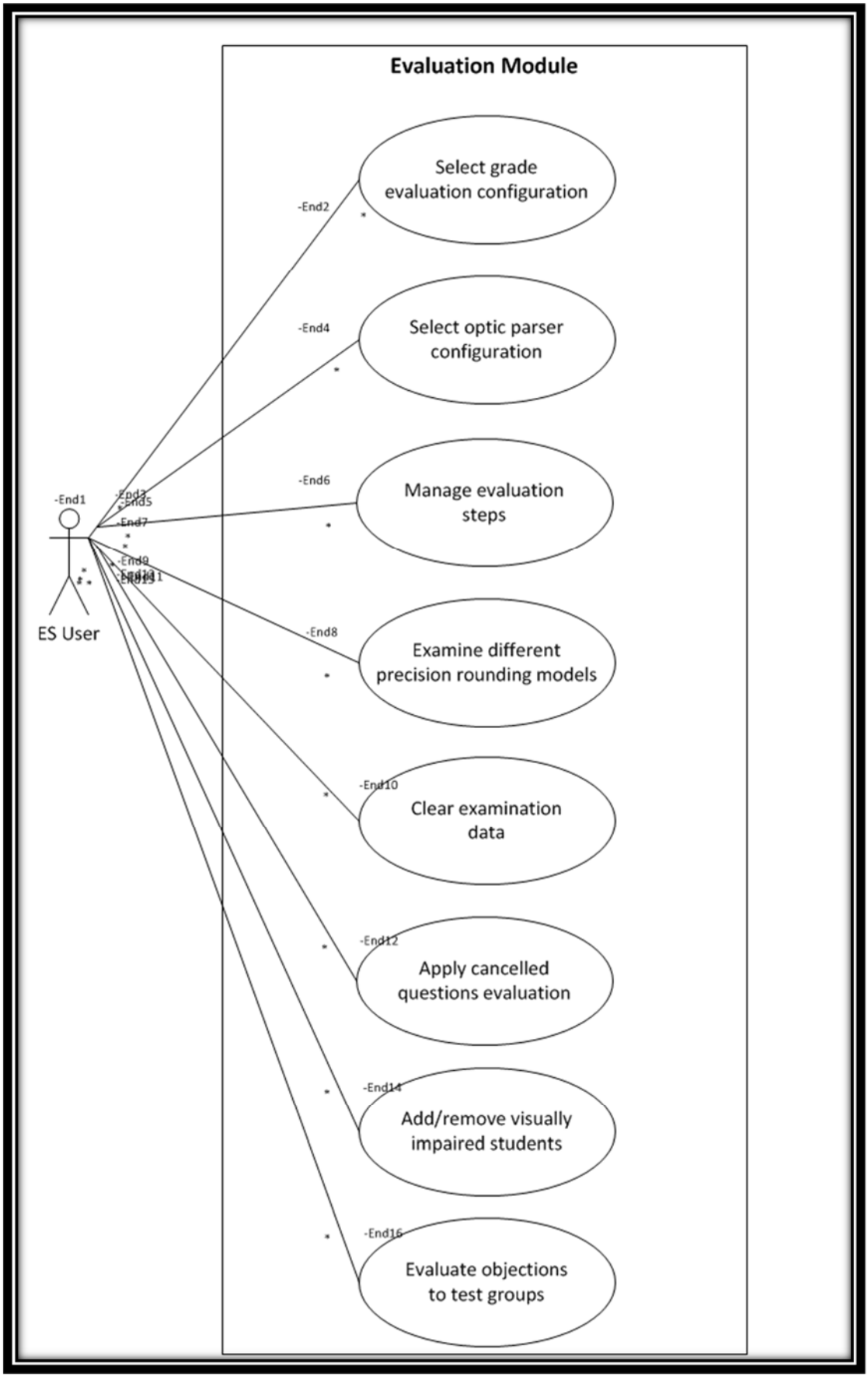

Figure 3.9. Use case diagram of evaluation module

3.5.4 TRU Module Use Cases

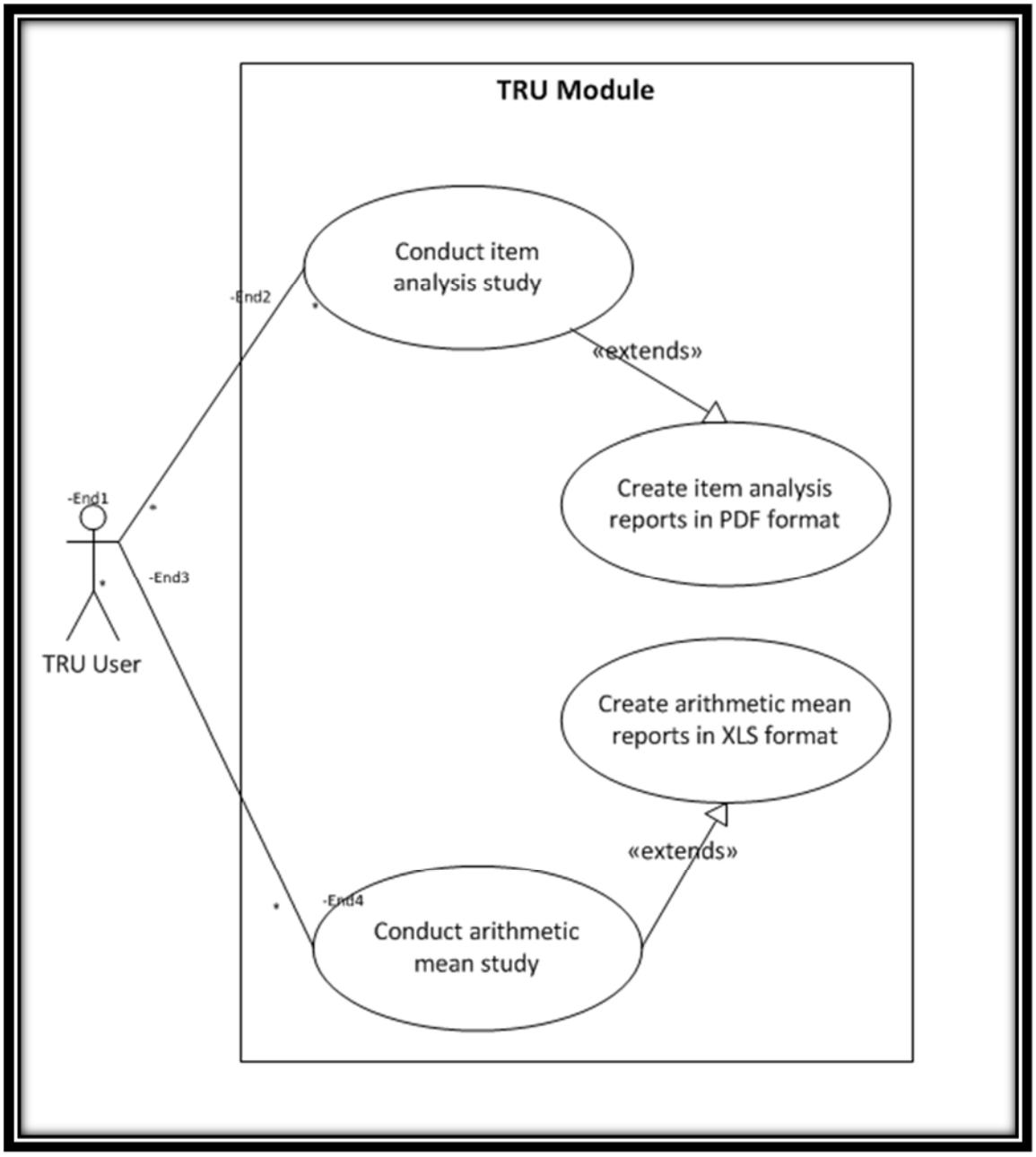

Figure 3.10. Use case diagram of TRU module

3.6 Experimental Results

Before we started to make experiments about measuring system effectiveness, usability, and reliability; we had created a test collection including sessions of some previous examinations. In order to create a variety of examinations, we selected an institutional examination, a big scale Open Education Faculty (OEF) examination, and a European Union based examination. For each of the examinations, the experiments are carried on the first sessions of the exams for the first scanning. The table below illustrates some statistics about these examinations.

Table 3.5. Statistics of exams in test collection

Exp. No	Exam Type	Total Sessions	Experimented Session	Application Centers	Students
1	Institutional	1	1	2	16773
2	EU Region	4	1	13	1003
3	Big scale OEF	4	1	91	1112151

In order to measure the effectiveness and the reliability of OMR tagging and exam evaluation, we uploaded the data files of evaluation to the system. Once the files had been uploaded, the problems at the scanning process of OMR were detected. Afterwards, we tagged the problems with suitable solutions and then moved to the evaluation stages. At this moment, we can report the accuracy of the OMR systems in Anadolu University. We have evaluated the accuracy by dividing the total number of answer sheets to the number of answer sheets not throwing any exception that the system should handle. The accuracy values calculated for each exam is shown in Table 7. The accuracies are nearly perfect for the OMR system in Anadolu University, but the errors are still not negligible and they must be controlled for 100% correctness in evaluation.

$$\text{Accuracy} = \frac{\text{Number of sheets having no critical errors}}{\text{Total number of answer sheets}}$$

Table 3.6. Calculated accuracies for exams in test collection

Exp. No	Total Answer Sheets	Total Errors	Accuracy
1	16773	9	0.99946
2	1003	14	0.98604
3	1112151	1614	0.99854

After transferring data files to the system, problems detected in recognition process were matched with suitable solutions via the tagging tool. Then, we applied the evaluation procedures in order to achieve the final course grades of students participated in the examinations. In the 3rd experiment, we did not include all of the application centers in order to reduce the time to check the answer sheets physically. After the calculation, the grades were compared with the official values announced to the students. In table 8, the statistics of grade comparison are shown.

Table 3.7. Grade comparison for exams in test collection

Exp. No	Evaluated Students	Official Grades	Matching Grades	Success
1	16775	33550	33550	100%
2	1003	2867	2867	100%
3	312853	931247	931236	99.998%

The analysis of the results introduced that the system works without any unexpected results for small and medium scale exams. In big scale exams, there exist little differences due to missing answer sheets. Since the missing sheets were not taken into consideration while experimenting with the system, these differences were actually expected.

Finally, our experiments encouraged us to develop and deploy the new system to the live environment in near future. The grades were successfully calculated, and the overall time and effort for evaluation were reduced. The improvements described in next chapter will result in deployment of a new system for AUES.

4 CONCLUSIONS AND FUTURE WORK

This thesis covers the development of automation about the evaluation for the examination services of Anadolu University. Taking the workflow and the workload of exam evaluation stages into consideration, different modules are created and integrated for different types of user groups.

In order to measure the correctness of efficiency of our system, we evaluated 3 past examinations by using the system. The evaluation data files are used as the test collection and all the procedures in evaluation stages are applied for this test collection. The parser and evaluation configurations are declared as the real values for those examinations. Our measurements and experiments approved that the effort in this thesis will result in the startup of a new, reliable, efficient and improvable evaluation system.

In near future, we are planning to deploy the system to the live environment. Before the deployment, we are going to schedule some tasks and improvements that need to be completed. Firstly, we are going to expand the authentication mechanism so that the authentication will be done with user credentials. In other words, the user group authentication will be replaced by user in-group authentication.

After the replacement of the authentication mechanism, we are planning to log every event in the system with the user details. This improvement will be useful when tracking the history of actions. Furthermore, it will increase the security level in general.

The deployment of the system will result in new performance tests, functionality updates according to user demands and bug fixes. Our aim is to meet the requirements in a few iterations and releases.

REFERENCES

[1] Open Education. (n.d.) Retrieved June 14, 2013, from http://www.anadolu.edu.tr/en/uzaktan-egitim-sistemi

[2] Distance Education. (n.d.) Retrieved June 14, 2013, from https://en.wikipedia.org/wiki/Distance_education

[3] Lever-Duffy, Judy; McDonald, Jean B (March 2007). Teaching and Learning with Technology. Ana A. Ciereszko, Al P. Mizell (3rd ed.). Allyn & Bacon. p. 377. ISBN 0‐205‐51191‐0. Retrieved 23 January 2011.

[4] Honeyman, M; Miller, G (December 1993). "Agriculture distance education: A valid alternative for higher education?". Proceedings of the 20th Annual National Agricultural Education Research Meeting: 67–73.

[5] Latchem, C., Özkul, A., Aydin, C. H. and Mutlu, M. E. 2006. The Open Education System, Andolu University, Turkey: E‐transformation in a mega‐university. Open Learning, 21(3): 221‐236.

[6] Optical Mark Recognition. (n.d.) Retrieved June 14, 2013, from http://en.wikipedia.org/wiki/Optical_mark_recognition

[7] Palmer, Roger C. (1989, Sept) The Basics of Automatic Identification [Electronic version]. Canadian Datasystems, 21 (9), 30-33

[8] Optical Character Recognition (n.d.) Retrieved June 14, 2013, from http://en.wikipedia.org/wiki/Optical_character_recognition

[9] Srihari S N (1992) High-Performance Reading Machines. Proceedings of the IEEE, 80(7): 1120-32.

[10] Bergeron B P (1998) Optical Mark Recognition: Tallying information from filled- in bubbles. Postgraduate Medicine, 104(2): 23-24.

[11] Dillman D A (2000) Mail and Internet Surveys: The Tailored Design Method. Pp.418-19. John Wiley & Sons, New York.

[12] McAlpine, Mhairi, (2002), A summary of methods of item analysis, Bluepaper, number 2, CAA Centre, for HEFCE.

[13] Arithmetic Mean. (n.d.) Retrieved June 16, 2013, from http://en.wikipedia.org/wiki/Arithmetic_mean

[14] Standard Deviation. (n.d.) Retrieved June 16, 2013, from http://en.wikipedia.org/wiki/Standard_deviation

[15] Skewness. (n.d.) Retrieved June 16, 2013, from http://en.wikipedia.org/wiki/Skewness

[16] Kurtosis. (n.d.) Retrieved June 16, 2013, from http://en.wikipedia.org/wiki/Kurtosis