Detecting cheats in online student assessments using Data Mining

José-Alberto Hernández[a], Alberto Ochoa[b,c] , Jaime Muñoz[d] & Genadiy Burlak[a]

[a.] Programa Doctoral de Ingeniería Eléctrica, Centro de Investigación en Ingeniería y Ciencias Aplicadas; UAEM. Av. Universidad #1000; C.P. 62209 Cuernavaca, México, jose_hernandez@uaem.mx
[b.] Programa de Ingeniería en Computación, UAIE; Universidad Autónoma de Zacatecas.
[c.] Institute (Postdoctorale Program), State University of Campinas; Postal Box 6176, 13084-971 Radamaelli – SP, Brazil.
[d.] Universidad Autónoma de Aguascalientes.

Abstract: We can find several online assessment applications, Windows oriented or Web based, licensed or gnu free software, proprietary or standardized. All of them executing basic questions and test interoperability stages: providing assessment items, training and/or evaluation, and the assignment of a grade. Tons of information resulting of this educational process is stored into databases, including starting times, local or remote IP addresses, finishing times and, the student's behavior: frequency of visits, attempts to be trained, and preliminary grades for specific subjects, demographics and perceptions about subject being evaluated. We propose the use of data mining to identify students (persons) that commit cheat in online assessments (cyber cheats) and identify patterns to detect and avoid this practice.

Keywords: online testing, internet frauds, cyber cheats, data mining, student behavior

Introduction

It is a basic question: Why to use the online assessment? We can answer the following. Online assessment is a common practice around the world; more than 271,000,000 links at Internet search engine (similar Google.com [10]) support this statement. However, is it really a good tool to assess persons? Can we trust on results? Can we trust on students?

We can say online assessments are useful to evaluate the students' knowledge; they are used around the world for schools -since elementary to higher education institutions- and recognized training centers of very important companies like the Cisco Academy [2].

Solving second and third question we are not sure, in traditional tests "copying from another student during a test" and "using banned crib notes or cheat sheets during a test" is categorized by the students as cheating [5] and several research [20-21] have shown is common practice. Scenario changes drastically when exams can be done remotely through the Internet [22, 25]. One of the basic problems to solve is to know: who is there?

Is it acceptable to us as a society to tacitly accept cheating as a fact of life and not be so shocked when it comes to light? Students don't stop at graduation, as we have seen in recent scandals in business and journalism. And cheating or cutting corners in one's professional or personal life can cause real damage—both to oneself and to others. We need to care about it [5].

Our discussion focus on the student's behavior under the online assessment environment, and our proposed model to help organizations to detect and to prevent cheats in online assessments. First we analyze different student personalities, stress situations generated by online assessments, and the common practices used by students to make cheat to obtain a better grade on these exams. Later we present our DMDC (Data Mining to Detect Cheats) model based on what we call the summary of best practices; here we analyze the designed database schema to register the student's information. We analyze the key variables for modeling and their possible values. We explain the use of Weka [28] to carry out data mining to find behavior patterns that fits suspect profiles to detect cheats in online assessments. Finally, we discuss the data preliminary obtained by applying of the proposed model and summarize our conclusions.

1. Internet frauds and student cheating in online assessments

The term "Internet fraud" refers generally to any type of fraud scheme that uses one or more components of the Internet - such as chat rooms, e-mail, message boards, or Web sites - to present fraudulent solicitations to prospective victims, to conduct fraudulent transactions, or to transmit the proceeds of fraud to financial institutions or to other connected with the scheme [17]

For our purposes, a fraud is a deception made for personal gain and permeates different areas of life [10] business, art, archeology, science and education. We can classify cheating in online assessments inside the last category, as a kind of (AKO) internet fraud which purpose is to conduct fraudulent transactions, for example to modify grade information in database (DB), to steal answers for questions, to substitute respondent, to copy from another student or cheat sheets, in single words to "commit cheat" to obtain a "better grade" in an online assessment. Considering that he Internet is the media of interaction to commit this kind of cheats, we introduce the term cyber cheat.

2. Analysis of student behavior

Have you ever commit cheat in an exam? And if true: Why? A survey conducted by Donald McCabe at the Center for Academic Integrity [20], asked that question and discovered the top five reasons why students cheat: lazy / didn't study or prepare, to pass a class or improve a grade, external pressure to succeed, didn't know answers, time pressure / too much work. In a sample of 1,800 students at nine state universities in United States of America, seventy percent of students admitted to cheating on exams [21]. Is of interest to analyze the student behavior, we part of the basic principle none cheating, but this statement is not true at all on what we see day to day on class rooms; we know or at least perceive some students mastering the art of cheating. Genderman [14] describes in his work Academic dishonesty and the Community College, four factors associated with dishonesty:

2.1. Individual Characteristics. Academic achievement, age, social activities, major and gender.

Grade Point Average (GPA). Students with lower GPA [30] were more likely to cheat on examination, Crown and Spiller [6] found "a significant negative relationship between cheating and GPA".

Age/personality. Younger students, traditional college students, and underclassmen are more likely to engage in cheating and other forms of academic dishonesty [6,21,31].

Social activities. such as membership in a fraternity/sorority, frequent partying, and increased extracurricular involvement have also been related to higher levels of dishonesty [6,21].

Major. Several studies have indicated that business majors are more likely to cheat than non-business students and that business majors have more tolerant attitudes toward dishonesty [6, 24].

Gender. The relationship between gender and dishonesty is less clear. Studies indicating that males are more likely to cheat are common, as are studies indicating no significant differences in gender [6, 31].

2.2. Peer group influences.

According to Crown & Spiller [6], studies have consistently indicated that students are more likely to cheat if they observe other students cheating or if they perceive that cheating is commonplace or acceptable among peers.

2.3. Instructor influences.

A number of studies have indicated that the environment within the classroom or examination setting, as established by the instructor, can have significant impacts on cheating [6, 24, 31].

2.4. Institutional policies.

Effective communication of policies and increased student awareness of penalties and enforcement tend to reduce dishonest behavior [1, 6, 20].

3. Data Mining to Detect Cheats (DMDC) in online assessments

Data mining is a knowledge discovery process to reveal patterns and relationships in large and complex data sets [7], refers to extracting or "mining" knowledge from large amounts of data [12]. Moreover, data mining can be used to predict an outcome for a given entity.

Data mining has been successfully used to analyze student behavior [18, 27], and to detect user cheats in credit cards [9] and insurance companies. We propose the use of Knowledge (interesting patterns) Discovery in Databases (KDD) a non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [8] for detecting student cheats in online exams. Figure 1 shows required modules: a Data Warehouse facility, a Data Mining Engine (DME), the Knowledge base (Training Data, Verification and evaluation), and the Model of pattern recognition.

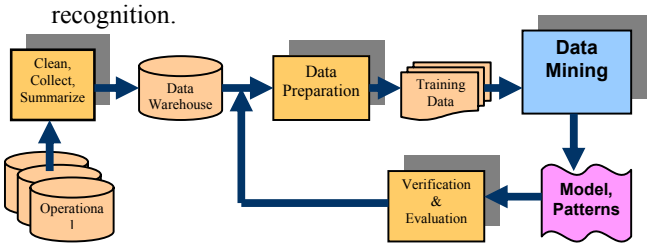


Figure 1. KDD schema and data mining

Many people treat data mining as a synonym of KDD. Alternatively, others view data mining as simply an essential step in the process of knowledge discovery in databases. We will use the second approach.

3.1. The Data Warehouse

Delmater and Hancock [8] wrote: "The science underlying predictive modeling is a mixture of mathematics, computer science, and domain expertise." The typical steps taken by researchers to make statistical assumptions about the population are not necessary. However, understanding the

database in which data reside and the data characteristics (structured and unstructured) are essential to successful data mining [18]

Data mining works best in exploratory analysis scenarios that have no preconceived assumptions [29] Research questions do not begin with "what *if;*" instead, they begin with "what *is*" [19].

In this section we analyze relevant classes, their key variables or attributes, and expected values. We will start by defining the Key variables identification based on Suggested variables for the Class Students are based on the works of Genderman [14] and Smyth M. L. et al [26] that will be stored the DBMS.

Class Students
GPA. A, B, C, D
Age. Younger, average, Above average
Gender. Male, female
Semester. 1-10 (Freshmen, sophomore, last semester)
Employment. Part Time, Full-time, None
Enrolment. Part time, Full-time
Major. Business, Education, Fine Arts and Humanities, Health professions, Math Sciences/ IT, Social Science, Nursing
Personality. Leader, traditional/average, underclass
Social activities. A lot of, average/some, none
Influenced by peers. True/False
Awareness of penalties. True/False

Class StudentPerception
Subject perception. Interesting, Unimportant
Professor. Involved, Indifferent
Instructor vigilance. High, Medium, Low
Unfair exam. True, False
Confusing exam. True, False
Quiz Test. True, False

Class SubjectEnvironment
Chat access. True, False
Spacing of students. True, False
Multiple choice exam. True, False
Multiple versions exam. True, False

Class Results of evaluation
IPAddress. Internet Address
Date and Time. Variables used to verify coherence of transaction times
Grade. Average.
Rating. Sum of +/- (Tr/Ta), (+) good answer, (-) bad answer, Tr. time of response, Ta (time available)
Level. Complexity of the test
TestPassed. True/False
TestInterrupted. True/False

Class HistoryLog
IdTransaction, IdStudent, IdCareer. Identifiers to keep track of transactions
IPaddress. Known, Possible, Suspicious
Action. Register, authentification, send questions, finish of test, send results of test, sign-off, and error.

Once we identified key variables, we establish relationships among classes as shown in the DB schema see Figure 2. The classes Careers, Subjects and Topics are used to manage tests.

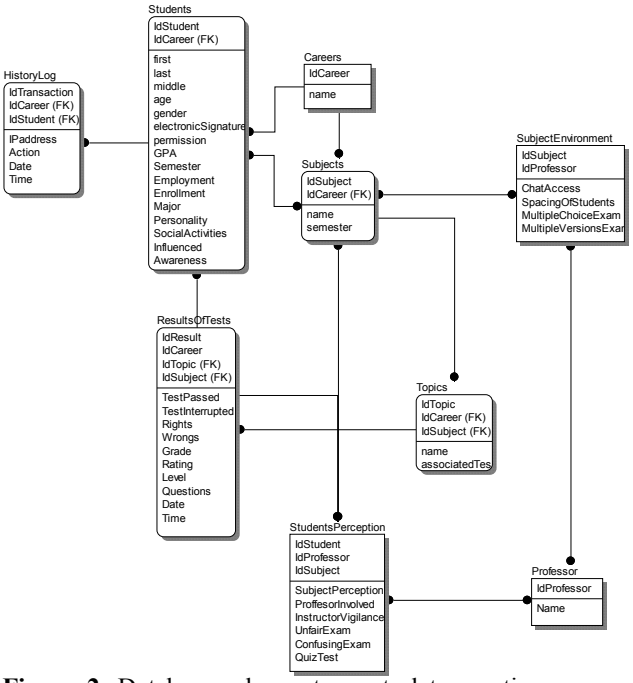


Figure 2. Database schema to control transactions including biometrical security, see "electronicSignature" field in table Students.

3.1.1 Information Sources

DB raw data will be obtained from students and professors through online surveys and the preferred online testing system.

Online Surveys
Information about student's behavior, demographics, and student's perception about subject and professor style are obtained by using this approach. Information about Subject Environment is obtained from the professor.

Proprietary Online Testing System (OTS)
Information about careers, subjects, topics, results of test and log files must be supplied by the selected OTS. In our case we use our made at home OTS. This system use XML learning objects based on IMS QTI Version 2.1 [16], is multiplatform (developed on Java), ciphers communication between the Client and the Server and allows students to be tested from remote places or in local area network [4]. Our OTS shows the test's questions and the answers in a random way. Each question has assigned a time to be answered (depending on the difficulty level), and do not permit to go back to the previous questions. These features are recommended to avoid cheating [22]. Figure 3 shows the OTS performance schema between the Client (student(s)) and the Server (examiner) [3].

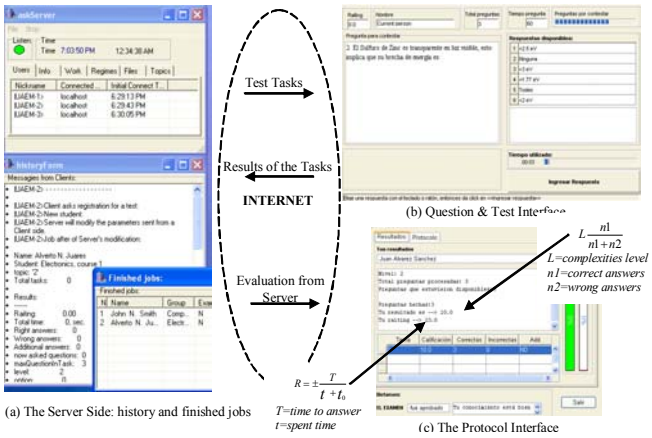


Figure 3. The Server is attending several clients

The OTS has been enhanced since past December 2005 by using the MySQL [23] database management system, one of the leading free-software DBMS, to deal with the arrangement and security of users personal information.

3.2. Knowledge base

Data base contains knowledge that allows the inference mechanisms obtain conclusions [15]. Such knowledge can include concept hierarchies, used to organize attributes or attribute values into different levels of abstraction [12].

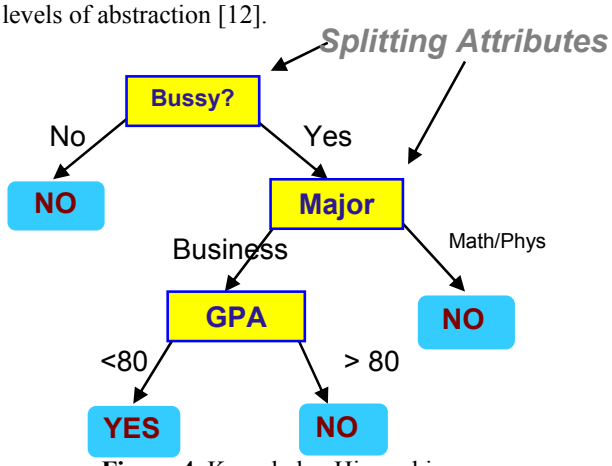


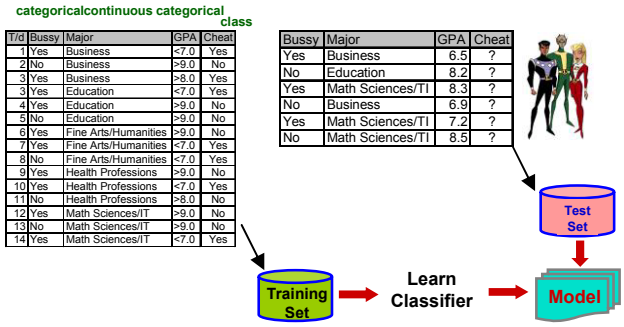
Figure 4. Knowledge Hierarchies

This knowledge is used to guide the search, or evaluate the interestingness resulting patterns (see Figure 4).

3.3. Data mining engine (DME)

Ideally DME consists of a set of modules for tasks such as characterization, classification, cluster analysis, and evolution and deviation analysis.

We will use Weka as data mining engine. (Weka is a collection of machine learning algorithms for data mining tasks.) The algorithms can either be applied directly to a dataset or be called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes [32]. "Weka is unique because it's easy to use and understand, and provides a comprehensive environment for testing methods against other existing methods" [13].



T/d	Bussy	Major	GPA	Cheat
1	Yes	Business	<7.0	Yes
2	No	Business	>9.0	No
3	Yes	Business	>8.0	Yes
3	Yes	Education	<7.0	Yes
4	Yes	Education	>9.0	No
5	No	Education	>9.0	No
6	Yes	Fine Arts/Humanities	>9.0	No
7	Yes	Fine Arts/Humanities	<7.0	Yes
8	No	Fine Arts/Humanities	<7.0	Yes
9	Yes	Health Professions	>9.0	No
10	Yes	Health Professions	<7.0	Yes
11	No	Health Professions	>8.0	No
12	Yes	Math Sciences/IT	>9.0	No
13	No	Math Sciences/IT	>9.0	No
14	Yes	Math Sciences/IT	<7.0	Yes

Bussy	Major	GPA	Cheat
Yes	Business	6.5	?
No	Education	8.2	?
Yes	Math Sciences/TI	8.3	?
No	Business	6.9	?
Yes	Math Sciences/TI	7.2	?
No	Math Sciences/TI	8.5	?

Figure 5. The classification process

Figure 5 shows the process of classification while in Figure 6 the generation of clusters is shown.

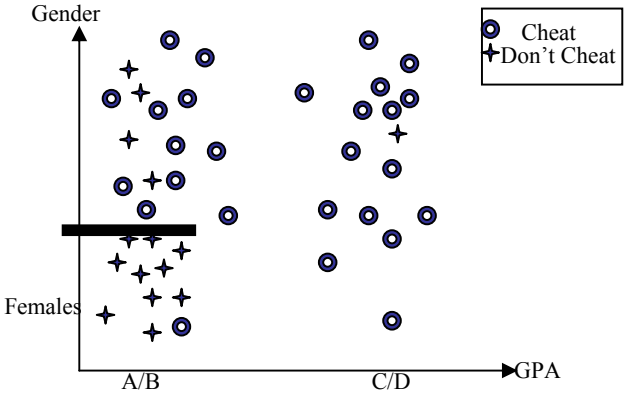


Figure 6. What is really happening?

In a fraud detection system, it is important to define the performance metrics carefully. Several fraud detection techniques use metrics like the detection rate, false alarm rate, and average time of detection. The typical fraud detection techniques attempt to maximize accuracy rate and minimize false alarm rate [33].

3.4. Pattern evaluation

The focus of this component is search towards interesting patterns. It may use interestingness thresholds to filter out discovered patterns [12]. Kohonen nets are useful for discerning patterns and groups within a feature space [18]. Researchers may use Kohonen nets to learn about the data before building other models. They have great value in understanding who takes what clusters of courses or what groups of students tend to have similar course-taking patterns.

4. Implementation of the DMDC Model

We propose the next five steps to data mining databases containing responses of online assessments, surveys and historical information to detect cheats in online exams:

4.1. Getting Data from students. Information from students must be collected from the historical data files and surveys.

4.2. The evaluation process. Carry out the student assessments using the OTS. To help reduce testing anxiety, you may consider offering practice quizzes with detailed feedback that do not count (or count very little) towards the students' grades. This provides a non-stressful way of practicing test-taking skills [22].

4.3. Obtaining Feedback. At end of each exam the student will be is asked for feedback about exam, and also about the professor and examination conditions. Professor will fill respective online form. Data from student, professor and institutions will be is stored on the DB.

4.4. Creation of the .arrf Data File. Necessary data for Weka system [28] will be obtained from DB, by using SQL statements and export facilities for further process (see Figure 7).

```
@RELATION students

@ATTRIBUTE name         string
@ATTRIBUTE edad         NUMERIC
@ATTRIBUTE sexo         {F,M,?}
@ATTRIBUTE isWorking    {0,1}
@ATTRIBUTE idstudent    NUMERIC
@ATTRIBUTE idCareer     {1,2,3,4,5,6,7}
@ATTRIBUTE permission   {0,1,?}
@ATTRIBUTE idSubject    {1,2,3,4,5,6,7,?}
@ATTRIBUTE idTopic      {1,2,3,4,5,6,7,?}
@ATTRIBUTE testPassed   {0,1}
@ATTRIBUTE testInterrupted {0,1}
@ATTRIBUTE rights       NUMERIC
@ATTRIBUTE wrongs       NUMERIC
@ATTRIBUTE grade        NUMERIC
@ATTRIBUTE rating       NUMERIC
@ATTRIBUTE level        {0,1,2,3,4,?}
@ATTRIBUTE questions    NUMERIC
@ATTRIBUTE daterecord   DATE "dd/MM/yyyy HH:mm:ss"
@ATTRIBUTE iscontroltest {0,1}

@DATA
Francisco_Martinez_Lopez,25,M,0,2,1,1,1,1,0,0,12,3,8,7.77,1,15,"08/12/2005 19:45:33",0
Rodrigo_De_Maria,22,M,0,72370,1,1,1,1,0,0,10,5,6.66667,30.07,2,15,"08/12/2005 18:44:43",(
Rodrigo_De_Maria,22,M,0,72370,1,1,1,1,1,0,11,4,7.33333,50.7792,2,15,"08/12/2005 18:52:37"
Rodrigo_De_Maria,22,M,0,72370,1,1,1,1,1,0,11,4,7.33333,35.5577,2,15,"08/12/2005 18:59:10"
Israel_Navarro,21,M,1,72529,1,1,1,1,0,0,2,13,1.33333,-24.3192,2,15,"08/12/2005 18:45:17",
```

Figure 7. Weka file (.arrf extension)

4.5. Data mining process. This process includes: reading the .arrf file in the Weka Explorer system, proceeding to classify, visualize clusters and discover associations in the data (see Figure 8). Start the hidden patterns finding, remember to keep mind open (No prejudices).

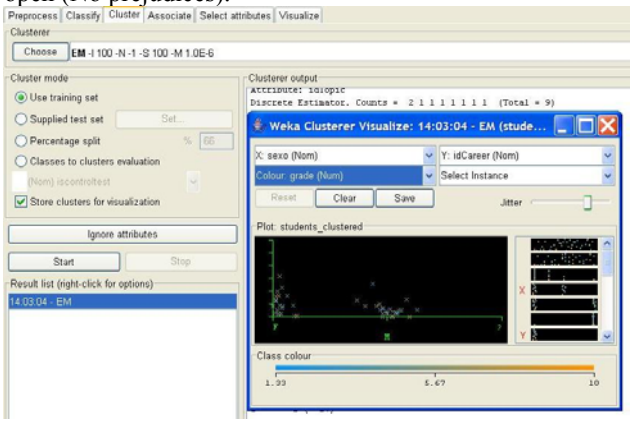


Figure 8. Weka options and the Clusterer Visualizer

5. Preliminary results and discussion

Second week of December 2005 we did our first attempt to implement the proposed methodology. We prepare the material for the final exam of Update on Information Systems subject, a 7th semester curricula for the career of Computers and Managerial Systems. Complete analysis of results will be published elsewhere.

5.1. Setup. At the beginning of the semester students were asked about demographics and a personality profile was prepared.

5.2. Training. On December 1st students were trained on the use of the OTS on class.

5.3. Evaluation. On December 8th students were evaluated using a fifteen, random, multiple choice test, each question had assigned a response time, and once selected the answer, students could not go back to check answers. Exam was automatically created from a sixty question database. The place for testing was the classroom and one of the computers was used a Server. First attempt was on training mode, second attempt was to assign a grade. In both cases students were supervised by their Professor.

5.4. Feedback. Students were asked about perceptions about exam. They mention tool was good and user friendly, but they felt under stress and there was not enough time to think about correct answers.

5.5. Data mining.

Information required to use Weka was prepared from the MySQL database, and the resulting .arrf was analyzed. Data mining process did not show any suspicious pattern to determine student online cheating. We believe two factors contribute to this result:

Exam environment was very controlled. Server and clients (students) were working in the same LAN (same range of IP Address), the employed OTS avoids undesirable practices, and professor was always present in exam.

Information to analyze was not enough. The students' class under analysis includes 97 persons. Data mining requires more information to produce confident results.

Conclusions

In this paper we have studied the Data Mining (DM) as a tool to detect the student cheats in online assessments. Since the increasingly new technologies evolve, students are mastering cheating in online tests. This allows us to define this fact as cyber cheating due to the use of the Internet as the interaction media.

We consider as very important the use of DM to identify the student cyber cheating in online assessments. We believe that the Data mining can be used successfully to find the student cheats in online

tests since enough information is provided. Best results can be obtained from information received from remote places.

We propose a five- step- methodology, involving Weka software as data mining tool, to identify "suspicious" student's behavior or patterns which we can classified as cyber cheating.

To perform a better DM analysis about student's cheating in online assessments we have to pay attention not only the student's perceptions and behavior, but besides the professor's teaching style yet. The information received from such analysis can be important in discussing and improving the testing environments and finally the general institution's philosophy.

Further work

Testing of the proposed technology in our research center and other institutions to measure the effectiveness of methodology and make required adjustments.

Compare the results obtained with Weka as DM engine with results obtained using other tools like Matlab and SPSS.

Implementation of electronic signature module based on Biometrical information to increase the security through the Internet.

Acknowledgments

We want to thank to J. Romero and S.P. Verma who made several helpful comments. Part of this work is supported by CONACYT project 47220.

References

[1] Aaron, R. M. (1992). "Student academic dishonesty: Are collegiate institutions addressing the issue?" NASPA Journal, 29(2), 107-113. (EJ 442 669)

[2] Academy Connection – Trainning Resources (2005) In http://www.cisco.com/web/learning/netacad/index.html December 28th, 2005

[3] Burlak, Gennadiy N. Hernández, José-Alberto. Zamudio-Lara. A. (2005) "The Application Of Online Testing For Educational Process In Client-Server System". CONGRESS: IADIS International Conference, Lisbon, Portugal at October 19-22, 2005. Vol. 2. p. 389-392. ISBN: 972-8924-04-6

[4] Burlak, Gennadiy N., Ochoa-Zezzatti Alberto, Hernández, José-Alberto (2005) "The application of learning objects and Client-Server technology in online testing to measure basic knowledge level". CONGRESS: X. Simpósio de Informática V Mostra de Software Academico, Uruguaiana, RS Brasil at October 19-22, 2005. ISSN 0103-1155

[5] Carnegie Perspectives (2005) "Justice or Just Us?" In http://www.carnegiefoundation.org/perspectives/sub.asp?key=245&subkey=577 December 28th, 2005

[6] Crown, D. F., & Spiller, M. S. (1998). "Learning from the literature on collegiate cheating: A review of empirical research". Journal of Business Ethics, 17, 683-700

[7] De Veaux, R. (2000) "Data Mining: What's New, What's Not." Presentation at a Data Mining Workshop, Long Beach, Calif.

[8] Delmater, R., and Hancock (2001), M. "Data Mining Explained: A Manager's Guide to Customer- Centric Business Intelligence". Boston: Digital Press

[9] Di Fata, Giussepe (2005) "Distributed Data Mining ACAI'05/SEKT'05 Advanced course on Knowledge Discovery" in www.ktschool.org/wiki/presentations/difatta/ difatta-DDM-NOVA.ppt January 11th, 2005

[10] Fraud (2005) In http://en.wikipedia.org/wiki/Fraud December 27th, 2005

11 [10] Google (2005) In http://www.google.com December 27th, 2005

[12] Han, Jiawei. Kamber, Micheline (2001) "Data Mining Concepts and Techniques". Academic Press San Diego, CA 92101-4495, USA

[13] FRST (2006). Foundation for Research Science and Technology http://www.frst.govt.nz/publications/tech-reports/downloads/index/Issue_29.pdf Consulted in January 12th, 2006

[14] Genderman R. Dean (2000) "Academic Dishonesty and the Community College. ERIC Digest" In http://www.ericdigests.org/2001-3/college.htm December 27th, 2005

[15] Giarratano, Joseph. Riley, Gary. (2001) "Sistemas Expertos Principios y Programación" Tercera edición. Internacional Thomson Editores, S.A. de C.V. p. 596

[16] IMS Global Learning Consortium (2005) "IMS Question & Test Interoperability Specification", In: http://www.imsglobal.org/question/, July 2005

[17] Internet Fraud (2005) In http://www.usdoj.gov/criminal/fraud/Internet.htm December 27th, 2005

[18] Luan, Jing & James Derrick. (2003) "Data Mining and Its applications in Higher Education", EBSCO Publishing 2003 Consulted online December 2005

[19] Luan, Jing. Terrence, Willet. (2000) "Data Mining and Knowledge Management: A System Analysis for Establishing a Tiered Knowledge Management Model". Report. Cabrillo College Office of Institutional Research 6500 Soquel Dive Aptos, CA, USA

[20] McCabe, D. L., & Trevino, L. K. (1996). "What we know about cheating in college: Longitudinal trends and recent developments". Change, 28(1), 28-33. (EJ 520 088)

[21] McCabe, D. L., & Trevino, L. K. (1997). "Individual and contextual influences on academic dishonesty: A multi-campus investigation. Research in Higher Education", 38(3), 379-396. (EJ 547 655)

[22] Morris, Terry A. (2005) "Cheating and plagiarism in the information age" in http://terrymorris.net/tohe2004/ CPIA_morris.pdf consulted in December 30th, 2005
[23] MySQL in http://www.mysql,org Consulted in January 2006
[24] Roig, M., & Ballew, C. (1994). "Attitudes toward cheating of self and others by college students and professors". Psychological Record, 44(1), 3-12.
[25] Rove, Neil C. (2004) "Cheating in Online Assessment: Beyond Plagiarism in Online Journal of Distance Learning Administration", Volume VII, Number II, Summer 2004
State University of West Georgia, Distance Education Center Consulted In http://www.westga.edu/~distance/ojdla/ summer72/rowe72.html December 30th, 2005
[26] Smyth, M. Lynnette, and James R. Davis. (Summer 2003) "An examination of student cheating in the two-year college." Community College Review 31.1 17(17). InfoTrac OneFile. Thomson Gale. Univ. Autonoma del Estado de Morelos. 30 December 2005 <http://find.galegroup.com/ips/infomark.do?&contentSet=IAC-Documents&type=retrieve&tabID=T003&prodId=IPS &docId=A107200754&source=gale&srcprod=ITOF &userGroupName=uadedm&version=1.0>
[27] Van Horn, Royal. (1998) "Data Mining For Research and Evaluation" Phi-Delta-Kappan Technology section Nov. 1998 251-252. 256
[28] WEKA 3 (2005)– Data Mining Software in Java In http://www.cs.waikato.ac.nz/ml/weka/ December 21st, 2005
[29] Westphal, C., and Blaxton , T. (1998) "Data Mining Solutions: Methods and Tools for Solving Real-
World Problems". New York: Wiley Computer Publishing.
[30] What is a GPA? (2005) Inhttp://educationusa.state.gov/ undergrad/about/system.htm December 28th, 2005
[31] Whitley, B.E., Jr. (1998). "Factors associated with cheating among college students": A review. Research in Higher Education, 39(3), 235-274. (EJ 567 552)
[32] Witten, Ian H. & Frank, Eibe. (2005) "Data Mining: Practical machine learning tools and techniques", 2nd Edition, Morgan Kaufmann, San Francisco
[33] Yufeng Kou, Chang-Tien Lu, Sirirat Sirwongwattana, Yo-Ping Huang (2004) "Survey of Fraud-Detection Techniques" In http://europa.nvc.cs.vt.edu/~ctlu/ Publication/ICNSC-04-KLSH.pdf December 27th, 2005