

PLAGIARISM DETECTION AND PREVENTION – EXPERIENCES FROM THE OPEN UNIVERSITY OF CYPRUS

C. Rodosthenous¹, P. Christoforou¹, M. Epiphaniou¹, G. Matheou¹, S. Mavrotheris¹, C. Christodoulides¹

¹Open University of Cyprus (CYPRUS)

christos.rodosthenous@ouc.ac.cy, petros.christoforou@ouc.ac.cy, mepiphaniou@ouc.ac.cy, georgia.matheou@ouc.ac.cy, stathis.mavrotheris@ouc.ac.cy, christodoulides@ouc.ac.cy

Abstract

The introduction of eLearning led to a number of advancements in student assessment and evaluation. Conventional methods of coursework submission, evaluation and feedback were substituted by online workflows and were enhanced with plagiarism detection tools to assist educators in evaluating large numbers of coursework. The increasing number of eLearning students along with the huge amount of freely accessible information on the Internet, require tools and services for educators, to assist the grading of students' coursework and check the originality of their work. In this paper, we present the methodology used for introducing and delivering a plagiarism detection service in the Open University of Cyprus eLearning platform and integrating it with the coursework submission process. Additionally, we describe the challenges faced during and after the implementation and the actions taken to train educators in using the service in a proper way. We conclude by presenting an evaluation of the service, both for its educational and technical soundness, along with statistics of its usage during the last four years.

Keywords: distance education, eLearning, plagiarism detection, student assessment, evaluation.

1. INTRODUCTION

The Open University of Cyprus (OUC) is a distance education public institution that was founded in 2002 and its first students were admitted in 2006. During the first year of its operation, OUC was offering the educational content mainly through the mail delivery system while coursework was submitted by students using the email service. After identifying the difficulties and constraints of that process, OUC chose to become a fully online university by introducing a state of the art eLearning Platform and incorporating all of the university's educational processes such as delivering educational content, conducting online meetings, submitting electronic coursework, evaluating students and providing feedback. Nowadays, the eClass eLearning Platform is the only educational medium used in the OUC to facilitate the educational process. One of the main processes handled is coursework submission. The increase of the number of courses along with the increase of students led to a major explosion of submitted coursework that needed grading and checking for plagiarism. To cover this need, the OUC introduced a plagiarism detection service.

This paper is organized as follows: First, the authors discuss what plagiarism is and describe the educational need at the OUC. Next, the methodology used for introducing a plagiarism detection service to eClass eLearning Platform is presented followed by usage statistics. Finally, authors present their conclusions and future work, introduction OUC, open education, eclass, coursework submission

2. WHAT IS PLAGIARISM

According to Carroll [1] "plagiarism is when you submit someone else's work as your own work". Neville [2] defines plagiarism as "one of a number of practices deemed by universities to constitute cheating, or in university-speak: 'a lack of academic integrity'". There are many other definitions, but most of them are similar and have the same meaning described earlier. Plagiarism occurs in a number of forms, like self-plagiarism [3], direct plagiarism [4], paraphrase plagiarism [4][5] and mosaic plagiarism [6]. An extended report on plagiarism forms can be found in work of Ercegovic [5], Bretag [6] and Jones [7].

There is also numerous bibliography related to the factors that lead to plagiarism. In a recent work of Dennis [8], the authors concluded that the first reason for students' cheating is the inability to do the coursework, followed by lack of time and difficulty to keep up. According to Wilkinson [9], some of the most common reasons for plagiarism are that students do not manage their time correctly, they do not understand the rules of referring and they have easy access to internet sources.

During the last decade, a number of plagiarism detection tools were developed to help students and educators to online check the originality of coursework by comparing it to a centralized database of uploaded papers and online sources. According to Vie [10], there is a discussion that these tools cannot be used to offer a complete solution to the academic problem of plagiarism, since they are not always very reliable detecting plagiarism.

3. EDUCATIONAL NEED

In each Academic Year, the OUC offers a number of Programmes of Study. Each of these has a number of courses, both mandatory and optional, for students to enrol in. A course consists of one or more groups, depending on the number of students enrolled in. Each group has an educator and can have up to 25 students. The educator is responsible for students' tutoring, counselling and for providing feedback on their coursework and academic progress. The OUC's educational methodology states that written coursework is a basic component of students' evaluation and a prerequisite for successful completion of each course [11]. Coursework in the form of assignments counts for up to 50% of every course's final grade. An average grade of 50% in all mandatory coursework is required for every student to be qualified to attend the final examination.

For these reasons, OUC's educators were always very concerned about the quality and originality of submitted coursework. During the first three years of the OUC's operation, the plagiarism check was done 'by hand'. Each educator was checking coursework submitted by its students, using available course material, search engines and free anti-plagiarism tools. During that period, the number of students and educators grew rapidly. In addition, a number of incidents occurred, where students were copying students' coursework from previous years. More specifically, students belonging to different groups of the same course were submitting near identical assignments and educators could not identify this form of plagiarism since they did not have access to that group. There were also numerous reports on online services who offered "OUC's assignment services", that allowed students to get assignments prepared by others against a fee.

Manual checking of student's coursework became almost impossible. Educators needed access to coursework submitted in previous years and access to all groups' coursework within the same course in order to be able to check for possible plagiarism. The university also needed a way to ensure that all educators were actually checking coursework against plagiarism. It was obvious to everyone involved in the educational process that an automated plagiarism detection service was a good option for plagiarism prevention and detection, a service that would automatically provide the information needed for an educator to decide whether a student's coursework is original or not.

For handling severe cases of plagiarism, the Student Matters Disciplinary Committee is the responsible body. Members of that body are responsible for verifying the findings of educators and provide the necessary penalties to the students. Plagiarism of any form is not tolerated at the OUC and the measures taken against the students doing so are severe.

4. DEPLOYING A PLAGIARISM DETECTION SERVICE

The deployment of a plagiarism detection service is a multistep process, since it involves a series of prerequisites and checks that need to be completed for the service to be deployed in the eClass eLearning Platform. As reported in [12], deploying a new service comprises of four phases: the initiation phase, the requirements analysis phase, the service deployment and integration phase and the evaluation phase.

4.1. Initiation phase

In 2010, the OUC's eLearning team and a number of faculty members identified the need for a service that could aid educators in checking students' coursework and thesis submissions for plagiarism. A

feasibility study was prepared that included factors like the tech infrastructure needed, the relevance to the educational methodology model followed in the organization, rough estimations of human resources needs, budget and of course the ability to integrate the service to the coursework submission process followed by students through the eClass eLearning Platform. Case studies from other Universities, mostly in Europe and the United States of America, were also reviewed, and a number of cases were identified that could be used as best practices.

The above were documented in the service initiation report. This report was then forwarded to the university's administrative board for approval. The approval of the document marked the initiation of the second phase, where a project team was appointed to gather the service requirements.

4.2. Requirements analysis phase

The project team proceeded with conducting interviews of the stakeholders. These interviews included both the faculty members and the Academic Affairs Committee. The outcome, was the need for a service that could assist the academic staff in detecting similarities in coursework from the internet, previously submitted coursework, books, academic articles and journals from various publishers and translated text. A number of academics also suggested a service for detecting source code plagiarism and for allowing students to check their coursework for plagiarism before submitting it for evaluation.

A number of technical requirements were also identified; like the need for a web service that did not require local installation on educators' devices, support for all the popular file formats used in coursework submission like word processor documents, portable document format etc., support for both Greek and English language content in files and finally security and privacy settings for secure coursework submission.

The project team prepared the requirement analysis documents, which included the needs described previously and launched an RFI (Request for Information) procedure for identifying available solutions. This procedure succeeded in involving a number of service providers who gave feedback on the requirements and proposed possible solutions that supported these. From the responses, the project team understood that Greek language support was a major obstacle, since most services did not support proper checking of documents with Greek content. In addition, the majority of the service providers only offered their system with a SaaS model [13]. This fact required extra precautions to verify that the OUC will be the Intellectual Property Rights (IPR) holder and that content will be securely stored on the service provider's infrastructure.

The project team proceeded with a thorough investigation of the methods used by the plagiarism detection service to integrate with eClass eLearning Platform. This resulted to a number of requirements like:

1. Service providers must have a plugin that will be added to the current coursework submission plugin of the eClass eLearning Platform and not substitute the default one with a new one.
2. Integration with the eClass eLearning Platform must allow educators to view a percentage of the plagiarized text detected in the gradebook for easy checking.
3. Students must be presented with a message before submitting, notifying them that their coursework will be checked for plagiarism automatically.
4. The total amount of time needed for the service to return results must not exceed 48 hours.

The eLearning team requested a demo of each service, so that the effectiveness and functionality could be checked in detail.

4.3. Testing and demonstration

For the tests, three major plagiarism detection service providers were used and were integrated to the test environment of the eClass eLearning Platform. In addition, a number of assignments were uploaded to populate the plagiarism detection database. For the tests to be conducted properly, test datasets were created with documents that included:

- T1: content from English and Greek news sites
- T2: content from open repositories with thesis and dissertations in Greek language
- T3: content from the OUC's institutional repository¹
- T4: content from assignments added to the plagiarism detection database
- T5: content from publishers that the OUC has access to
- T6: content from books used for teaching in OUC courses
- T7: source-code from sites with examples (e.g., Microsoft Developer Network, php.net, etc.)

Each test dataset was uploaded in the test environment and the results of the detection process were documented and presented in the table below.

Table 1 - Comparison of Plagiarism Detection Services

	T1	T2	T3	T4	T5	T6	T7
Ephorus ²	56%	47%	85%	90%	46%	35%	48%
Turnitin ³	88%	63%	79%	92%	61%	53%	37%

From the three services tested, only two were able to properly display the results of plagiarized parts of the document in the Greek language. From these two services, only one could be integrated with the eClass eLearning Platform using the default assignment plugin. For that service, the results for the detection of plagiarized text from the seven areas presented were good, except from the detection of content from publishers and teaching books. After further investigation, the poor results were justified since most publishers had their content available in formats that could not be directly accessible from the service indexing engine and most of the content indexed was in English language.

During the tests, a number of cases were detected, where the plagiarism service could not provide proper results. These cases included translated text, text in pictures or video, source-code and proper source citations. Translated text is really an issue for plagiarism detection services, since there is not an accurate one-to-one translation of a text passage.

The project team finalized the requirement specification document and suggested the preparation of an agreement with the selected service provider (Ephorus). The fact that the service was provided as a SaaS, required special terms in the agreement to protect user privacy and intellectual property rights [14].

4.4. Service deployment and integration

The completion of the agreement signalled the beginning of the next phase for the project, the deployment of the service and the integration of it. The project team prepared a detailed project plan comprised of four distinct phases:

1. Activation of the service and content migration.
2. Integration of the service with the coursework submission workflow of the eClass eLearning Platform.
3. Preparation of training material and educator training.
4. Pilot phase initiation.

¹ <http://kypseli.ouc.ac.cy>

² <https://www.ephorus.com/>

³ <http://turnitin.com/>

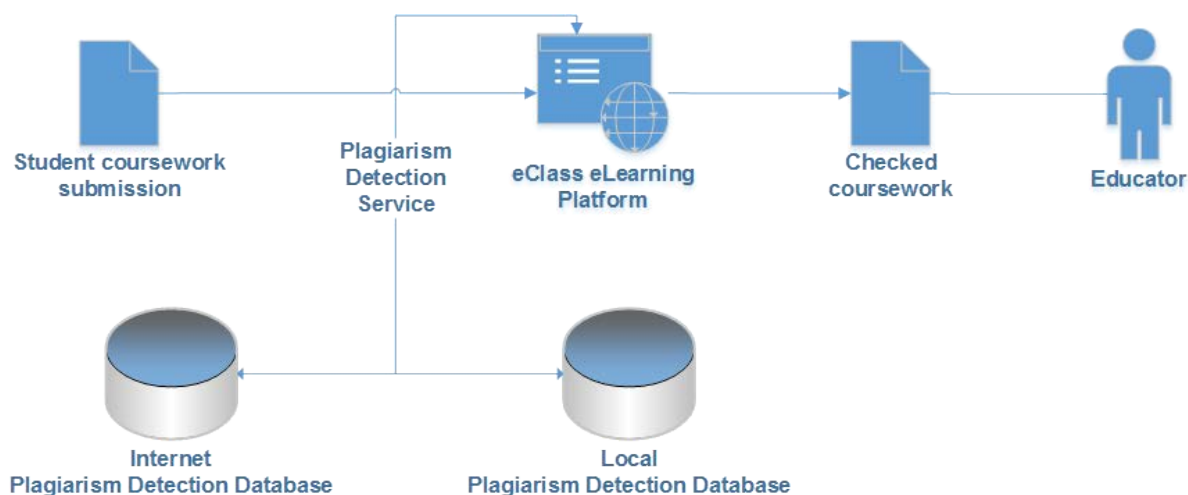


Figure 1 - Schematic of the integration of the Plagiarism Detection Service with the eClass eLearning Platform

Up to 2010, the OUC had a large volume of coursework stored from previous year students. This coursework was uploaded to the plagiarism detection database, so that to be used as reference material. A number of thesis and books were also uploaded to populate the database.

Furthermore, a plagiarism plugin was installed to extend the coursework submission workflow. This plugin allowed educators to view the plagiarized content detection percentage. It also enabled educators to view both summarized and detailed detection results with changes detected from the original text as well. These reports could be printed, exported and emailed to other educators for consultancy purposes.

For educators to properly use the service, training material was created with detailed documents and videos on how to check an assignment using the service interface. Moreover, a training session was organized with participation from faculty members from all Programmes of Study. All participants had the chance to discuss the academic perspectives of plagiarism, university rules and penalties, and the culture that underlies beneath plagiarism. Moreover, live demos of the system were prepared, with people participating and asking questions. During the training sessions, it was made clear by everyone that the service will be used for aiding educators to detect plagiarism and under no circumstances, this service would be used to substitute the actual work needed by educators' to detect plagiarism. Moreover, most participants understood that the introduction of the service could also be used as a prevention mechanism for students submitting not original coursework.

The plagiarism detection service was initially deployed for a six-month pilot. During that period, users had the chance to test it and get a better understanding of the functionality and benefits of the new service. Users also had the chance to report and request extra functionality and features. These requests were hierarchized and some of them were accepted for addition in future versions of the integration plugin.

5. EVALUATION OF THE SERVICE

The plagiarism detection service is deployed in the Open University of Cyprus since 2011. The university policy states that every submitted assignment has to be checked for plagiarism. The usage of the service is continually evaluated by the academic community of the OUC. Its usage and results are monitored and forwarded to the overseeing body of the university. In this section, the authors present statistics for each Academic Year since the deployment of the service. The fact that the OUC is an online University and students are obliged to use the eClass eLearning Platform for submitting their coursework, allows us to have detailed statistics of the service usage. These statistics include, number and type of sources detected, similarity percentage for checked assignment and user engagement [15].

5.1. METRICS

The following table presents figures for the OUC that are important in understanding the reasons that led to the need for selection and deployment of a plagiarism detection service. From the numbers below, it is clear that the amount of coursework submitted has increased rapidly in each Academic Year.

Table 2 - Organization in Numbers

	2011-2012	2012-2013	2013-2014	2014-2015
Number of Programmes of Study	9	14	17	21
Number of courses in eClass	96	149	234	289
Number of students	2445	3557	4172	4727
Number of faculty	18	20	19	18
Number of educators	115	135	230	320
Number of assignment activities in eClass	386	794	1344	1439
Number of coursework submissions	14071	30263	29253	30208

When an assignment is submitted, it is automatically checked for plagiarism. The outcome of this check is a percentage of similarity of the content of that assignment with various sources already indexed in the service database. The following figure (Fig. 2) presents for each Academic Year, the distribution of the similarity detection percentage.

These numbers, also include content that is part of the student's coursework, like the coursework title and instructions, pre-defined coursework templates, multiple choice question templates etc. that are detected as plagiarized text but are not. In addition, cases of resubmitting an assignment are also sometimes detected as plagiarism, since educators forget to withdraw the first submission from the reference database. These are some of the "weaknesses" a plagiarism detection tool has.

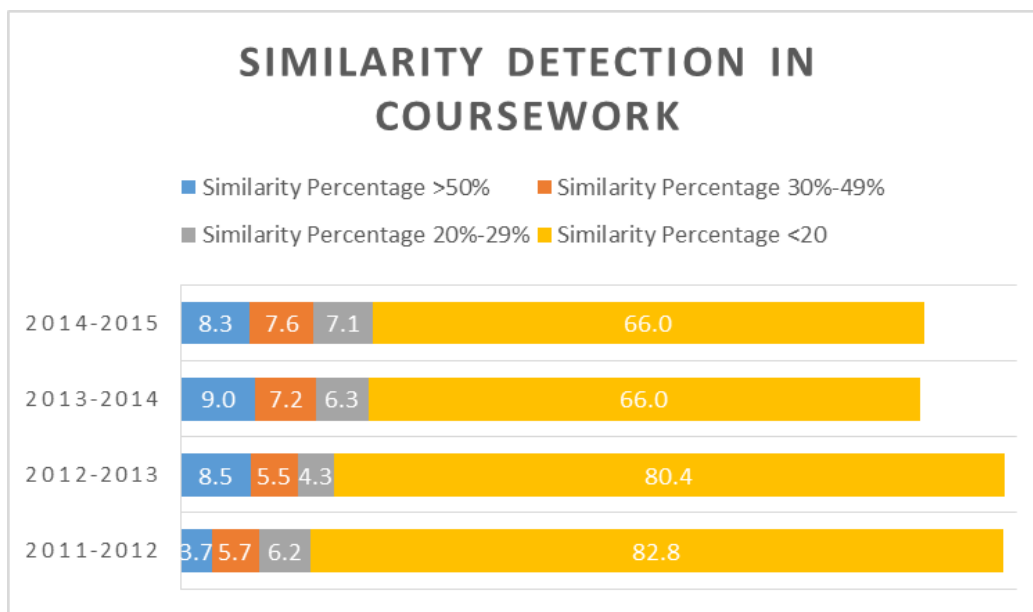


Figure 2 - Similarity detection rates

Through the years, the plagiarism detection database has grown. It currently has more than 125,000 submitted documents that are used as reference material for other submitted coursework. It is also populated with course material (books, exercises, notes etc.). Fig. 3 presents the sources from which similarity is detected. A local source includes previously submitted coursework, course material and any other form of material that is locally stored in the local plagiarism detection database. Internet sources include news sites, blogs, and open repositories etc., which are indexed in the global

plagiarism detection database. The results show a significant increase in the detection of similarities from local sources instead of internet sites.

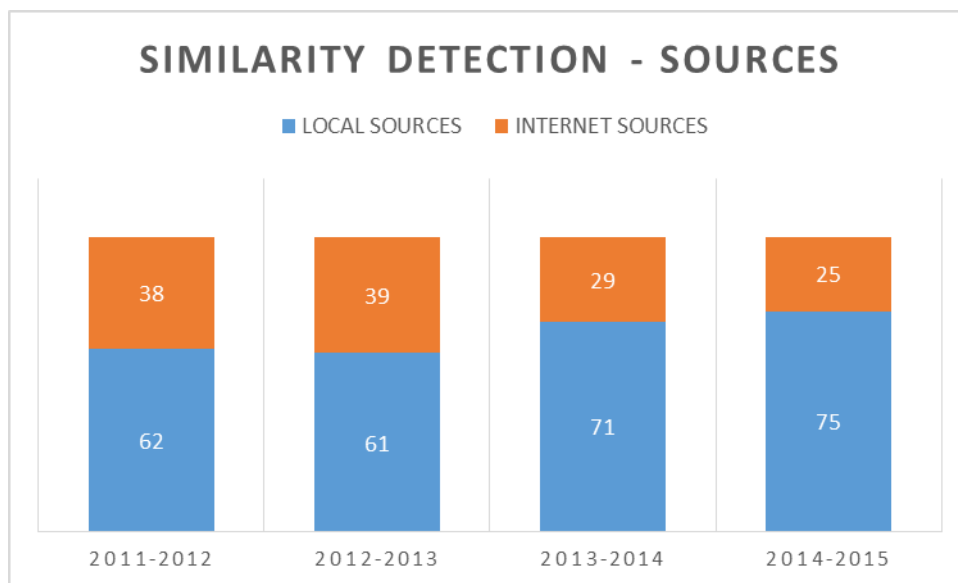


Figure 3 - Similarity detection - sources distribution

One of the reasons for deploying a plagiarism detection service was to aid educators in grading students' coursework. Since the deployment of the service, the average grading time for coursework is decreased. Table 3 presents the increase of assignments and coursework submission in eClass eLearning Platform, and the decrease in the average grading time in each Academic Year.

Table 3 - Average grading time for coursework

	2011-2012	2012-2013	2013-2014	2014-2015
Number of assignment activities in eClass	386	794	1344	1439
Number of coursework submissions	14071	30263	29253	30208
Average grading time (in days)	17	11	14	13

6. CONCLUSIONS & FUTURE WORK

The authors of this work aim in helping other institutions interested in deploying similar services, by presenting their experiences and there up to date usage statistics. The introduction of a plagiarism detection tool in a university is far from being a cure for the phenomenon of plagiarism. It is in fact a step towards the direction of building a healthy scientific culture and academic honesty for future scientists and to act as a prevention method for students to present not original work. The results presented here reflect the complexity of the problem in an online distance education University where the amount of submitted coursework is huge, the deadlines are tight and educators and students do not have physical presence as it stands with conventional universities. The authors of this paper discuss the process of deploying a plagiarism detection service and the factors needed to be checked during the process. For the future, we plan to build a network with Greek Universities that will use a common plagiarism detection database. This effort is far from easy, since there are many policy issues that we need to overcome.

REFERENCES

- [1] Carroll, J., & Oxford Centre for Staff Development (2002). A handbook for deterring plagiarism in higher education. Oxford: Oxford Centre for Staff and Learning Development.
- [2] Neville, C. (2010). *The Complete Guide to Referencing and Avoiding Plagiarism*. McGraw-Hill Education
- [3] Scanlon, P. M. (2007). Song from myself: an anatomy of self-plagiarism. *Plagiarism: Cross-Disciplinary Studies in Plagiarism, Fabrication, and Falsification*, 57.
- [4] Klausman, J. (1999). Teaching about Plagiarism in the Age of the Internet. *Teaching English in the Two Year College*, 27(2), 209-212.
- [5] Ercegovac, Z., & Richardson, J. V. (2004). Academic dishonesty, plagiarism included, in the digital age: A literature review. *College & Research Libraries*, 65(4), 301-318.
- [6] Bretag, T. & Mahmud, S. (2009). A model for determining student plagiarism: Electronic detection and academic judgement. *Journal of University Teaching & Learning Practice*, 6(1)
- [7] Jones L., Taylor R., Irvin S., Faircloth L. (2001). *Academic Dishonesty, Cheating, and Plagiarism*. Humanities & Communication Department Florida Institute of Technology Melbourne, Florida
- [8] Dennis, L. A. (2005). *Student Attitudes to Plagiarism and Collusion within Computer Science*. University of Nottingham. Available at <http://eprints.nottingham.ac.uk/319/> [Accessed 20 May 2015]
- [9] Wilkinson, J. *Staff and Student Perceptions of Plagiarism and Cheating*. Vol. 20. International Society for Exploring Teaching and Learning, 2008. Print.
- [10] Vie, S. (2013). A Pedagogy of Resistance Toward Plagiarism Detection Technologies. *Computers and Composition*, 30(1), 3-15.
- [11] Open University of Cyprus (2013). Internal rules of study. Retrieved from <http://www.ouc.ac.cy/web/guest/ysmf/library> on 20/05/2015
- [12] Christoforou, P., Rodosthenous, C., Epiphaniou, M., Mavrotheris, S., Christodoulides, C. (2015). Deploying new service in the Open University of Cyprus eLearning Platform – Our experiences. In 9th International Technology, Education and Development Conference, Madrid, Spain
- [13] Messier, R. (2014). *Software as a Service, In Collaboration with Cloud Computing*. Edited by Ric Messier, Syngress, Boston, pages 33-55, ISBN 9780124170407, <http://dx.doi.org/10.1016/B978-0-12-417040-7.00003-4>.
- [14] Foster, A. L. (2002). Plagiarism-detection tool creates legal quandary. *The Chronicle of Higher Education*, 48(36), A37-A38.
- [15] Epiphaniou M., Rodosthenous, C., Christoforou, P., M., Mavrotheris, S., Christodoulides, C. (2015). Engaging users in Distance Education - The case of the Open University of Cyprus. In 9th International Technology, Education and Development Conference, Madrid, Spain