

Rewriting, academic fraud and falsification: analyzing cases of academic practice in St. Petersburg, Russia

Rekhtina L.S., Pankratova L.S., Sokolov N.V.

Saint Petersburg State University
Saint-Petersburg, Russia
lisabet-09@mail.ru, l.s.pankratova@spbu.ru,
nvsspb@yandex.ru

Karimov R.N.

Institute of Geography of National Academy of Sciences
Baku, Azerbaijan
rovshan_karimov@yahoo.com

Samedov F.R., Yunisov J.K.

Research Institute of Aerospace Information
Baku, Azerbaijan
farid.samedov@mail.ru, cyunisov@yandex.ru

Abstract—The article discusses phenomenon of plagiarism forms in academic end education practice and its detection and reaction system.

Index Terms—rewriting, academic fraud, falsification higher education.

I. INTRODUCTION

Information technologies are an essential part of the education system. E-learning is gradually developed and implemented in education process, not only as a conveyor of knowledge and information, but also as a communication buffer between faculty, administration, and students. This system should save resources and optimize learning process, reduce time budget of education agents and formalize communication between them. The article considers some complex issues in the work of certain elements of the implemented E-learning system and science.

Due to the overproduction of texts, as well as the demand to quickly produce new texts, the probability of plagiarism in a paper increases. A wide spread of the Internet, information technologies, and communication tools give to a vast audience access to a huge amount of information, which also increases the probability of textual borrowing in academic papers. Thus, according to various studies, the percentage of plagiarism in student papers increased from 23% to 90% during the period from 1941 to 2001 [1]. For example, the number of students in Russia from 1940 to 2015 has increased from 478,000 [2] to 14,191,000 people [3].

Scientists, participants and managers of major research projects, supervisors and even students are facing the need to check texts for plagiarism, as they are no longer confident in their own texts [4]. Previously, textual borrowing was possible to find

out only by manually comparing the two texts - the source and its replication. The information society offers a wide range of tools to facilitate this task. But the attitude to plagiarism has become more severe and serious. Simple, even automatic, comparison is no longer sufficient to define plagiarism in a text. There is a need for a serious, professional expertise of each individual case of textual borrowing to identify whether it is plagiarism or not. Increasing proportion of plagiarism toughens and complicates the procedure of inspection, and that requires more costs. Requirements for texts and their expertise are considerably growing. Now it is necessary not only to compare two texts, but also to conduct semantic analysis of the papers. That is poorly automated and requires manual comparison and the work of experts.

II. RESEARCH METHODS

This article is based on the thorough analysis of 9 typical cases of plagiarism detection in student papers (there are more than 30 cases overall) in Saint Petersburg, Russia. These are papers submitted by the students of the final year of undergraduate program in Sociology on their major courses.

First of all, these cases illustrate different examples when the work of one student is stolen by the others and presented as their own paper. In 2011 there was a case when the successful work of one student was brought in a week three times by his/her co-students with minimal changes. But lately, cases of complete repeat of someone's work are rare. As a rule, the original text of the first paper is edited: parts of the text are put in different order, and words are replaced by synonyms. Among such cases the compilation of several works is a usual practice. These texts may belong to their co-students or senior students who studied earlier.

Primary this applies to text assignments (course programs and analytical reports) and practical tasks.

Plagiarism from the Internet sources (total or partial borrowing and their compilation) is almost a global problem. Such kind of plagiarism could be found in every second job. Most often it is detected in the section of a paper that presents methodological and theoretical bases of the work. This is a common case, but quite problematic for a check as plenty of materials are used from sources that are not indexed (social media, books or journals from private, closed databases or archives).

We consider separately cases of practical works downloaded from the Internet or fabricated, as there are no any plagiarism detection technologies for such papers. They are usually checked manually and it is effective only if part of the material is in the public domain.

III. RESULTS

Rewriting is a natural component of the information explosion. It allows enormously increase the amount of information while reducing its cost and resource intensity [5]. The academic environment does not involve the use of rewriting as a source of new original texts. Rewriting is a skill that is necessary for abstracting of scientific sources, but it is not allowed in the final, original (especially research) texts. In the situation of “the tyranny of the moment” a student dwells on rewriting rather than trying to continue to work creatively with information from the initial source. This “tyranny of the moment” [6]. is felt not only by students who do not have time (motivation, desire) to work thoroughly with the original paper, but also by teachers (lecturers) who do not have time to educate, develop skills of independent, critical and creative writing in the class. Moreover, students are already immersed in an environment where rewriting is a commonplace. For this generation it is obvious that many of the tasks do not require the creation of a new text, it is enough just compiling it from materials previously created by other authors.

Plagiarism detection of students’ papers gradually becomes a mandatory procedure in the educational practice in the post-soviet space. In Russia a lot of universities have introduced requirements for the verification of the final qualifying papers for plagiarism. In Azerbaijan in 2016 began the development of a unified system for plagiarism checks of the texts in the spheres of education and science [7].

The main specialized system of plagiarism detection recognized by the Russian scientific and academic community is called “antiplagiat.ru”. It appeared in year 2007 and is still in the process of development. This is a basic tool to check texts. This system (a free version) allows you to create a personal profile where it is possible to download text files, and then the system searches for a match with its database of papers, educational materials, and public sites. These systems hide the algorithm of the work, but presumably the “antiplagiat.ru” system uses “w-

shingling based on N-grams with an additional normalization of texts in accordance with dictionary of synonyms, and finding of plagiarism locally has a trigger effect to protect against false positives” [8].

In several Russian universities are introduced e-learning and internal communication systems right now. In Saint Petersburg State University, it is the system named Blackboard (<https://bb.spbu.ru>), which consists of three platforms - Blackboard Course Delivery, Blackboard Content Management, Blackboard Community Engagement. To check downloaded texts Blackboard uses such component as SafeAssign. The Higher School of Economics University uses the LMS system (Learning Management System - <http://lms.hse.ru>) which allows to download texts in an extended version of the “antiplagiat.ru” service. Modification and enhancement of applications comprising these systems still require a lot of time. Their users constantly address to support division and are forced to bypass the limitations of the systems in order to complete their professional tasks. For instance, key problems of the Blackboard are following: basic functionality, not enough flexibility, clunky interface, technical difficulty or inability to organize work in groups/in pairs, sensitivity to the bibliographic data and common phrase.

As practice shows, the automatic plagiarism detection of papers is not enough. On the first stage of the check all works were tested by “antiplagiat.ru” system which is the leading one among existing in Russian. Such services warn that they are looking for the plagiarism, meaning matches in the text. But that is not enough for plagiarism detection to know what percentage of matches were found by the program. It is essential to understand what is behind found textual borrowing. This may be a list of references, properly executed quotes, set expressions and common phrases that simply cannot be formulated otherwise. That means that a match found by automated tools may be not plagiarism. It is important to explain this nuance to students. Each year students raise the question about the acceptable percentage of plagiarism in the work. Whilst according to the documents they signed before their admission to the university state that writing works (in accordance with curriculum) prepared by student must not contain plagiarism. Works may contain a certain percentage of such textual borrowing, as quotations and illustrations if they are correctly formatted. Apparently it is up to 20% of textual borrowing that is allowed. Experience has shown that level of understanding of the differences of these phenomena among students is very low.

So, on the second stage there is a necessity to carry out further check based on the results provided by automated services. If the software allows you to view a report, you need to go through it and see each case matches. Internet is a very mobile system and in the case of the anti-plagiarism report the links, which should show the alleged source, often do not work. Then teachers have to carry out manual checks of the matches, using public search engines (like “Google”, “Yandex”). This procedure requires form a lecturer time and resources, research skills and

knowledge of different software systems that allow fixing the fact of plagiarism.

Thus, the first level of such “forced” plagiarism is a rewrite without references. It seems to be a natural reaction of the generation to a task that is considered daunting and excessive. (That is the result of “the tyranny of the moment” when resources are limited.)

Let us consider the example of a good student paper containing a rewriting without causing any damage to its content. It is a work of group of students which is, according to the “antiplagiat.ru” service, original on 96.43%, meaning the amount of borrowing in the text is 3.57%. The work regarded issues of identifying and detecting signs of social identity in everyday life. In this study there were used valuable and correct scientific sources expanding and enriching the conceptual and basis of the paper. And the references were listed. The problem was in the lack of quotation marks (“_”) on the both sides of the borrowed text. The necessity to use quotation marks was explained to the students and they solved the problem by adding them shortly afterwards. (Although there were cases when other students’ teams could not fulfill this requirement because they did not manage to detect borrowing text in their own paper and arrange everything properly.) This case can be seen as a rewriting which allows students to complete the task successfully. In fact, while preparing a research proposal there is absolutely no need to create new theoretical text, since it will be written worse by students then in the original version used by them. But academic rules exclude the possibility for students to use only someone else’s text. Students are required to complete a very intensive, consuming a lot of resources tasks, meaning to prepare their own academic products under the conditions of resource shortage [9] and integration in the environment where there are obvious easy solutions in the form of rewriting, for example.

Interestingly this paper was also submitted by other students from the same course as their own original work. The students who tried to cheat submitted this paper before the real authors. Actually it was one of the previous (unfinished) versions of the final paper that was given to them by real authors as an example of the task. This case, as well as others, shows that the students do not have a sense of the value of the original author’s work and their own creative work.

The second level of academic dishonesty is an attempt to create academic papers from other texts trying to imitate the scientific activities. Thus, the fabrication and falsification of meanings of academic activities happen. If rewriting does not lead to loss of previously created meanings, such compilation of texts leads to destruction of these meanings. Unfortunately, rewriting often transforms into plagiarism when there is not only absence of quotation marks of a text which is appropriate to be used, but instead a consciously created paper from fragments of other texts. An example of such case is a student’s essay (8 000 characters) devoted to the problem of homeless children. According to the “antiplagiat.ru” service the originality of this

text is just 9.88%, everything else is different pieces of other texts. At the second stage of plagiarism detection it turned out that this essay is an example of dual plagiarism. The essay was a downloaded text of student paper from the Internet, which is in turn, was a compilation of parts of texts written by other authors (“patchwork text”). There were only a few references to the sources used. This is an example of academic fraud [10]. But a passing glance given by a teacher at the paper may not raise any questions as the text has logic, and text pieces fit together, and the problem of homelessness and research approaches to it are well presented.

The third example illustrates a case where a set of “pieces” of a text are combined chaotically to form a meaningless, false and sophisticated text. The task was to prepare a research design to study some social issues of migration. The originality of the text according to the “antiplagiat.ru” service is 67%. The second stage of the check has shown that the submitted paper consists of different pieces of various texts mixed together. There were some references in the paper which were not related to the borrowing from the source text. The submitted paper contained references that were falsified. Drawbacks of this paper: borrowed pieces in the text are copied with misprints and errors; the text is replete with trivial phrases (at the level of common sense). There were selected most common and expatiative arguments for the compilation. They do not indicate a problem, do not justify the relevance of the work, and do not allow justify the sequence of researcher’s activities. The extracts from statistical compendiums are given without any reference to them. Interestingly, the manual search with the help of “Yandex” (search engine) showed that almost the whole paper consists of pieces from other texts. A high percentage of originality shown by the “antiplagiat.ru” service could be explained by the fact that the database of this service did not contain some original sources.

The third level is the partial or complete falsification of the academic process and academic product. It is criminal activity that must be stopped in the early stages of education, as it can completely discredit and devalue the whole scientific spheres. Below are presented examples related to the problem of falsification of papers/practical exercises by students. In one case, the students falsified the interview – they role played the interview which was published in “LiveJournal” (service for journals and blogs). Another example of falsification of the same task: some tracks that were not recorded during the interview were “glued” in it. It was possible to identify this because of the difference in recording quality, specificity of students’ conversation style. It was clearly heard that the text is read on the recording and there are places where the text was “glued”. This is very time consuming practice. Falsification of the interview takes more time and resources than the fulfillment of the task (to conduct an interview). Another instance is when the interview was completely borrowed from the article of the Russian internet media “BaltInfo.ru”. Students changed only names of the interviewee and the city of origin.

If rewriting and academic fraud can be checked with the help of technical means, falsification of practical exercise may be identified only by manual comparison of two files. But this requires creation of a database that contains all students' papers that could be compared with each other. In Russian universities there have been already introduced the system of communication between teachers and students which should solve this problem. It is a social network for internal use at the university supplemented with special software. Students submit all their papers only through this system, and lecturers check them their (upload reviews and comments). Thus, communication between students and lecturers become more formalized; implemented more strict time management - if a submission deadline has passed, a student will not be able to download a paper; formed an internal database of all student papers. When the same paper is uploaded into the database, the system should automatically identify and report such cases. But in fact, such databases are just in the process of formation at the moment. The database check file attributes and, so far, is not good in identifying repeating parts of the file or compilation within it. Internal systems of plagiarism detection have not proved to be effective yet. They create an additional burden on a lecturer for its maintenance, and thus taking his/her attention from checking of students' papers. Another effect of such systems concerns bibliographical references which are seen by the system as borrowings. Thus, bibliographical references increase the total percentage of plagiarism detected by the system. Students are determined to minimize the percentage of "plagiarism" by reducing a list of bibliographic references, as they do not understand the difference between plagiarism and borrowing from a text. There is a major controversy - the system is designed to check and teach students how to work with texts, to follow copyright laws and to save hyperlinked apparatus of science, but in fact it achieves the opposite effect as it destroys the habit of students to fix all the sources used. Another alternative was introduced in Azerbaijan in 2016. It is a development and implementation of a general system of plagiarism detection in texts for universities and research centers. Yet it is too early to evaluate its effectiveness [11]. Such idea has its benefits, as texts and papers can be spread between students not only from different courses, but also universities.

IV. CONCLUSION

Informational and software technologies for borrowing detection in papers designed to maintain and preserve the norms and rules of academic work with texts become more complex. The development of these systems requires very serious resources, as well as development of their interaction with other information systems (libraries, search engines). The development and refinement of these systems, and expertise in case of serious precedents requires work of a number of specialists: programmers who are capable of implementing the sophisticated algorithms, lawyers that deal with the legal aspects, marketing consultants [12], sociologists, economists, and certainly linguists that are able to assess and understand the features of word usage

and carry out semantic analysis of the text [13]. Complexity of these systems is already very high and their potential is not sufficient to eliminate the problem of manual plagiarism detection of the texts and difficult professional expertise. Rewriting technologies, plagiarism and falsification of academic papers are ahead of the development of these systems. When the "antiplagiat.ru" service was released a counter service "antiplagiat-killer.ru" appeared. That is a kind of arms race. New communication technologies have created endless amounts of information that are available to a large number of consumers. That requires constant replenishment of this information. Rewriting is an answer to this demand. The need to link new texts with original sources and to suppress theft of papers makes it necessary to develop new systems of sophisticated search and detection of borrowing. And they, in turn, force to look for new ways to "create" materials and/or opportunities to bypass the check. The dialectics of this process makes to constantly increase the resource intensity of these systems, thus distracting resources from the main goal - teaching students the norms of academic writing and development of their writing skills. The era when a student brought someone else's work ends with the emergence of the opportunity to compare student papers with papers from previous years. But there are technologies that are beginning to produce texts from materials that are not indexed by databases of plagiarism detection systems.

The practices of plagiarism are dangerous, because they destroy the frame of hyperlinking in science. But total detection of plagiarism and rigid copyright protection leads to the situation when information is "locked" in the narrow frames of citations. And that is even more dangerous and harmful for science. This anti-plagiarism war shadows the main purpose of the educational process. Students are no longer taught how to write a text or to conduct an interview, no one cares about that. What is important is whether a student is caught or not caught plagiarizing. And not to be caught plagiarizing is enough for a successful continuation of a career.

V. DISCUSSION

Plagiarism in academic texts is a social phenomenon which is a reaction of the actors of educational process to the spread of modern information technologies. Problem of plagiarism occurs in the following cases: 1) students do not understand the boundaries of legal and illegal usage of someone else's information product, 2) students do not understand the purposes of educational tasks assigned to them or believe these purposes are irrelevant, 3) students believe that rewriting technologies, fraud and falsifications are reliable and safe. But plagiarism is widely distributed in student texts also because of lecturers, meaning that: 1) lecturers are not interested in the process of creating a student text and control only the result, 2) lecturers give tasks that are not relevant to the modern level of development of information technologies and provoke plagiarism, 3) lecturers do not provide the declared level of control of student papers and texts. There is third party responsible for the problem of plagiarism it is academic

management which: 1) reduces the resources needed for the effective and close cooperation between lecturers and students, 2) formalizes the academic interaction, 3) tends to massification and simplification of educational processes. Plagiarism is the result of economy of resource from three sides - students, lecturers and management. The social problem is the latent criminal convention of these three groups on the admissibility of plagiarism in academic texts within the margin error of informational and technological tools that are used to monitor it at any given moment. But these tools do not solve the problem, but only reduce amount of attention paid to primary tasks of education and science. Moreover, they create alternative problem of artificial restrictions on the free dissemination of information that is necessary for the academic environment. Only the transition to social and information tools will allow moving the focus of control from overprotection of information to the motivation of its operators. The aim is to reduce benefits that come from the exploitation of someone else's intellectual labor. Solution of the plagiarism problem is in the field of social technologies, meaning reducing the benefits of using fraud and falsification in academic sphere.

The key point is the redistribution of educational procedures in accordance with the modern status of the student and correction of academic status itself. For centuries, students have created academic texts in order to achieve the participation in the elite club of intellectuals that are free from heavy physical labor for the sake of doing science. Modern society is close to the total displacement of manual labor by intellectual, but the latter has completely lost elitism. Overcoming the lack of information allows us to solve most of the intellectual problems without creating an original text. This means that the competent rewriting that implies respect for the law and science ethics, as well as the preservation of the original source of meaning, and provision of access to it via hyperlinks, is the necessary and sufficient level of competence for a large part of students, future professional practice of which does not imply intellectual creativity. Creative education demands significantly more resources than the one dominant today in the mass market. Teaching to write a text is much more expensive than doing rewriting. Therefore, social

solution of the plagiarism problem implies restoring of the elite status of the creative intellectual labor and education.

VI. REFERENCES

- [1] L. A. Jensen, J. J. Arnett, S. S. Feldman & E. Cauffman. "It's wrong, but everybody does it: Academic dishonesty among high school and college students", *Contemporary Educational Psychology*, 27, pp. 210, 2002.
- [2] G.V. Balashov. "The scale of higher education in Russia", *Economics of education*, issue 2, pp. 125-130, 2014.
- [3] "Russia 2015" Statistical Handbook URL: http://www.gks.ru/free_doc/doc_2015/rus15.pdf (the date of access 30/05/2016)
- [4] N. S. Nikiforov, S. V. Lavrinenko. "Tests in education and their application in the electronic environment", *Application of Information and Communication Technologies (AICT)*, 2015 9th International Conference on 14-16 Oct. 2015, pp. 482 – 485.
- [5] R. Hilliard. "Rewriting" in *Writing for Television, Radio, and New Media*, Cengage Learning, 2014, pp.176-178.
- [6] T. H. Eriksen. *Tyranny of the moment. Fast and slow time in the information age*, London: Pluto, 2001.
- [7] Azerbaijan intensifies combating plagiarism in education. Veliev. G. Baku network: International policy and security // URL: <http://www.trend.az/azerbaijan/society/2516303.html> (the date of access 30/05/2016)
- [8] E.S. Chirkin. Use of Antiplagiat in education *Vestnik Tambov University. Series: Natural and Technical Sciences Issue number 6 (2) / volume 18/2013*
- [9] L.S. Rekhtina, N.V. Sokolov. "A one-dimensional man in Gutenberg galaxy: the professionalization of the young scientist", 11 *Materials of the international scientific-practical conference "Modern scientific potential" - February 28 - March 7, 2015.*
- [10] M.A. Eckstein. *Combating academic fraud: Toward a culture of integrity.* International Institute for Educational Planning. UNESCO, 2003, p. 101.
- [11] E. Ostapenko. "Azerbaijan begins to deal with academic plagiarism", URL: <http://Inews.az/society/20140611052049665.html> (the date of access 03/06/2016).
- [12] J. Bailey. "Why Marketers Should Care About Copyright and Plagiarism. *Plagiarism Today*", URL: <https://www.plagiarismtoday.com/2012/09/13/why-marketers-should-care-about-copyright-plagiarism/> (the date of access 05/06/2016).
- [13] A. Kravchenko., V.I. Kureichik, D. Zaporozhets. "Translation of natural language queries to structured data sources", *Application of Information and Communication Technologies (AICT)*, 2015 9th International Conference on 14-16 Oct. 2015. P 57-5