Zashcholkin K.V.
Odessa Polytechnic State University

Ivanova O.M.
Odessa Polytechnic State University

Kuznietsov M.O.
Odessa Polytechnic State University

Sulima Y.Y.
Separated Structural Subdivision "Odessa Technical Applied College
Odessa National Academy of Food Technologies"

APPROACH TO THE AUTOMATED TEXT BORROWING ASSESSMENT BASED ON THE CLASSIFICATION OF SOURCES AND DESTINATIONS IN E-LEARNING SYSTEMS

The article is devoted to one of the aspects of eLearning process - the analysis of the impact of the volume and nature of textual borrowing on the quality and the final assessment of academic works. A form of knowledge control, based on the assessment of academic work, the results of which are presented in the form of a set of text files, is considered. The disadvantages of the existing e-learning systems in terms of providing such control are noted. The paper specifies the disadvantages of the existing approaches to automated analysis and evaluation of the legitimate text borrowing (without signs of plagiarism) in students' works. Within the framework of the existing approaches the presence of such kinds of borrowing either does not influence on a students' work mark or reduces it if the total amount of borrowing exceeds some threshold. The proposed approach is based on the numerical appraisal of borrowing effect on the quality of students' works and the final mark. It is different from the existing ones as it is founded on the differentiation of text sources and location of borrowing in students' works (hereinafter sources and destinations). The principal thesis of the approach is as follows: a source of borrowing from one hand and its destination from the other hand are able to have either positive or negative influence on the quality of students' works. Within the framework of the proposed approach the classification is performed and as a result each of the borrowing is referred to the two created classes: class-source and class-destination. The numerical assessments assigned to each of the classes as well as the amount of borrowing of each of the classes are the arguments of the resultant function which indicates the borrowing influence on the quality of students' works. The paper represents the proposed approach advantages and possibilities of its practical implementation.

Key words: *e-learning, text borrowing, automated assessment of text borrowing, borrowing search in students' works, learning management systems.*

Introduction and statement of problem. Out-of-class tasks are an important constituent of training process frequently used in the University system of education. Such kind of tasks is natural for graduate and course works and partially laboratory works of engineering specialties. The course of doing such kind of tasks consists of the following stages: getting a task; its immediate doing; creating a report on the task done; probably defending the results and at last its final evaluation. As a result of the task done a student forms the text information totality: the notes or report of the work carried out. This text information is typically demonstrated in the form of a file or multiple files represented in one of the popular electronic formats of text documents: doc, docx, odt, pdf. The mentioned files are given to an instructor or automated knowledge management system for examination. The instructor analyses the information, if necessary he (she) asks the student for some explanation as to the details of the task done, and marks it.

The student's tasks of the described type are characterized by the usage of text borrowing [1]. The borrowing can be both of legitimate character (containing all the necessary references to sources made in a proper way) and illegitimate one (possessing the plagiarism) [2]. Hereinafter in the given paper an analysis of only the legitimate text borrowing without plagiarism signs is being considered.

The traditional approaches to the students' work evaluation do not consider text borrowing used in work as a factor effecting on the work quality. In the authors' opinion it is a disadvantage of the existing approaches. The goal and way of the borrowing usage along with the other factors characterize the level of students' knowledge and skills. In evaluating the task done by an instructor (without automation means application) the borrowing factor effects on the final mark. However if the instructor takes into account this factor he (she) has to examine in detail the links between the borrowing, methods of its representation in the text and student's explanations on the contents of the task done. The procedure of such kind of verification of student's knowledge and correctness of borrowing usage takes a lot of time (wherein one should take into account that the time required for verification is limited). Under these conditions the decision about the borrowing effect on the task quality is made by the instructor on the basis of incomplete data and dependent on the instructor's experience.

Analysis of recent research and publications. The borrowing factor should be obviously taken into consideration in using the computer-aided estimation system including in Learning Management Systems (LMS). However the existing LMSs [3; 4] possess only two scenarios of borrowing processing:

− *scenario 1*: legitimate borrowing is considered as a neutral factor and its presence does not influence on the mark;

− *scenario 2*: legitimate borrowing is considered as a neutral factor however in exceeding some threshold of the total amount of borrowing the task is transferred to the class of illegitimate work [5–7].

The example of the second scenario implementation is the system Turnitin [8], which at present is the standard de facto in the field of plagiarism verification and the verification process control. This system refers the works containing a large amount of legitimate borrowing to the class of illegitimate ones [9]. According to the Turnitin's classification such kind of works are referred to the type "Aggregator" and defined as "Includes proper citation to sources but the paper contains almost no original work" [10]. Turnitin in his survey [11] estimates the frequency of occurrence of this types of works as 4.4 according to 10-point scale.

Thus within the framework of both of the first scenario and the second one the presence of the legitimate text borrowing does not effect on the assessment. The second scenario just permits to define a work as legitimate or illegitimate depending on the exceeding of stated threshold of the total amount of borrowing in text.

On this basis we can argue that the existing approaches to automated assessment of students' tasks (and software systems realizing them):

a) in the area of their functioning, which is performed without instructor's participation (automatic part), solve only the trivial problem of selection of works containing a large portion of legitimate borrowing;

b) concentrate in the nonautomatic part (implemented by an instructor) the analysis how borrowing influences on the work quality without offering means of its (analysis) formalization and methods of (analysis) performance.

The goal of the paper. In the authors' opinion the borrowing influence assessment process on the students' task quality can be substantially formalized and transferred to the automatic area of the functioning of automated estimation systems. In the connection with this fact *the goal of the given paper is as follows* – to improve the approach to analysis of text borrowing in students' tasks by assessing in numerical way the borrowing effect on work quality and its final mark.

The proposed approach to text borrowing analysis. The approach to the text borrowing analysis offered in the given paper is as follows: to consider different types of borrowing as heterogeneous objects, which being present in the text, influence differently on the mark of a task. To classify different types of borrowing the two new characteristics are distinguished: kind of the borrowing source and borrowing place (borrowing destination) in the assessed work. According to these two characteristics classification as well as the characteristic of amount of borrowing, the numerical assessment of the degree of borrowing effect on the task mark is offered. The proposed approach is determined by the following basic thesis.

The *first thesis* determines that text borrowing used in students' tasks possesses different nature of origin, various roles and purposes. Depending on the mentioned characteristics the borrowing can be: negative phenomena and in this case they make the mark lower; of neutral character and do not effect on the mark of a task; considered as a positive phenomena and on the contrary make the mark higher [12].

The *second thesis* permits to find out the character and origin of borrowing by classifying them according to the attribute of borrowing source. The basic classes have been distinguished in order to carry out the classification:

− CS_1 – the statement of task and instructions how to do it;

− CS_2 – methodological literature, which covers the issue;

− CS_3 – scholarly literature, which covers the issue;

− CS_4 – the results obtained from the engineering experience generalizing systems of such type as StackOverflow [13], MathOverflow [14];

− CS_5 – borrowing represented as answers received at topical Internet forums on the corresponding fields;

− CS_6 – borrowing from the web sites representing the documentation on technology, equipment, means, which are characteristic for the field of the task performed;

− CS_7 – text fragments represented as the solutions of analogous problems solved by students before and obtained from the Internet.

The list of classes mentioned above is not comprehensive and closed and can be added depending on the peculiarity of task and estimation priority.

The *third thesis* deals with the numerical assessment of each class of the borrowing sources according to some scale with the view to differentiate the borrowing sources and indicate the degree of their effect on the task quality and future mark [15].

The *fourth thesis* allows to indicate the borrowing role in the evaluated task by classifying them according to their place (location) in the text of task (borrowing destination). In order to create such kind of classification the structural model of the estimated task is necessary with the view of indicating its structural elements and links between them. By structural elements we mean e.g. the chapters or smaller structural units of the task.

The *fifth thesis* represents the numerical assessment of each of the destination class according to some scale with the view to differentiate the structural elements of the considered task in which the borrowing is located (placed).

The *sixth thesis* proposes the numerical assessment of each of the borrowing found in the task with considering the numerical assessment of source classes and destination classes of borrowing as well as the relative amount of borrowing in the total volume of task.

In accordance with the thesis mentioned above each of the borrowing B_i is characterized with the three parameters: A_{Si} – numerical assessment of the borrowing source; A_{Di} – numerical assessment of its destination; V_i – relative amount of borrowing. Where the amount V_i is indicated as the ratio of the number of borrowing symbols Bi to the total amount of symbols in the task. In the case of including the borrowing

in the preset classes of sources and destinations the assessments A_{Si} and A_{Di} are equal to the assessments of the corresponding classes. If the borrowing cannot be referred to one of the preset classes then these assessments are assigned to the borrowing manually.

Technique of the proposed approach usage. The sequence of actions, which are necessary to assess the borrowing influence on the task quality is the basis of the proposed technique. The technique is implemented with the help of software called hereinafter automated borrowing assessment system (ABAS). The technique consists of two parts. The first is performed once in initializing ABAS and is its initial starting adjustment. The second part is applied in each of the student's tasks being considered.

The part of the technique performed once (ABAS initiation) consists of the following stages.

Stage 1. ABAS proposes the instructor to input the source classes in the system and carry out their formal description. This description is to contain a characteristic attitude permitting to refer the source to the definite class. Such kinds of attitudes can be: web site addresses, authors' names, book and article titles. As a results of this stage a set of source classes $CS = \{CS_1, CS_2, \ldots, CS_n\}$, where $CS_i = <desciption_i>$, $desciption_i$ is the formal description of i-source class, is formed in ABAS database.

Stage 2. ABAS offers the instructor to give a numerical assessment to each of the source classes. The instructor independently states the scale of assessment or chooses one of some predetermined scales offered by the system. The assessment assigned to each of the source classes expresses the degree of positive, neutral or negative influence of borrowing of the given class on the task quality. As a result of this stage the numerical assessment A_{CSi} is added to each of the source classes; the class description has the form $CS_i = <desciption_i, A_{CSi}>$ in ABAS database.

Stage 3. ABAS proposes the instructor to input the borrowing destination classes (distinctive structural elements of student's task) into the system and perform their formal description. This description is to contain the attributes permitting to include a destination in the specific class. The titles of task chapters or key words of the chapter titles are offered to be used as such kinds of attributes. As a result of this stage a set of destination classes $CD = \{CD_1, CD_2, \ldots, CD_m\}$, where $CD_i = <desciption_i>$, $desciption_i$ is formal description of i-class of destinations, is formed in ABAS database.

Stage 4. ABAS offers the instructor to give the numerical assessment to each of the destination classes in the stated or predetermined scale. This assessment

expresses the degree of borrowing pertinence in the given structural element of the task. As a result of this stage a numerical assessment A_{CDi} is added to each of the destination classes; the class description has the form $CD_i = <desciption_i, A_{CDi}>$ in ABAS database.

Stage 5. ABAS offers the instructor to state (or choose of the variants proposed by the system) the equation of borrowing assessment function:

$$A_B = f(A_S, A_D, V), \qquad (1)$$

where A_S – borrowing source assessment; A_D – borrowing destination assessment; V – relative amount of assessment.

Stage 6. ABAS offers the instructor to state (or choose of the variants offered by the system) the equation of the function of the resultant borrowing assessment in student's task:

$$A_W = f(A_{B1}, A_{B2}, \dots , A_{Bp}). \qquad (2)$$

This function arguments are assessments A_{Bi} of each of the borrowing used in the task.

The part of the technique applied for each of the student's task consists of the following stages.

Stage 1. ABAS enters the student's work into the plagiarism search system, which indicates whether there is any illegitimate borrowing in the work. If there is some the measure set by the System setups are taken: the work is automatically rejected or given to an instructor with the list of illegitimate borrowing sources.

Stage 2. ABAS performs the classification of legitimate borrowing containing in student's work according to its (borrowing) source attitude (Fig. 1). If the source of some borrowing cannot be referred to one of the formed classes CS_i then ABAS offers the instructor to create a class for this source and numerically assess it without including it in any of classes.

Stage 3. ABAS divides the text of work into structural elements (borrowing destinations) and forms the structural model of work. The instructor has the possibility to interfere this process and divide the work into the structural elements in detail.
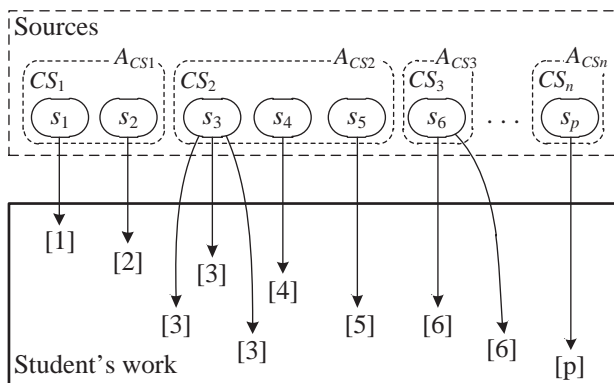

Fig. 1. Example of the borrowing source classification

Stage 4. Depending on the location in the text of work the borrowing is referred to this or that destination CD_i class (Fig. 2). If the destination of some borrowing cannot be included in one of the stated classes then ABAS offers the instructor to create a class for this destination and numerically assesses it without referring it to any of classes.
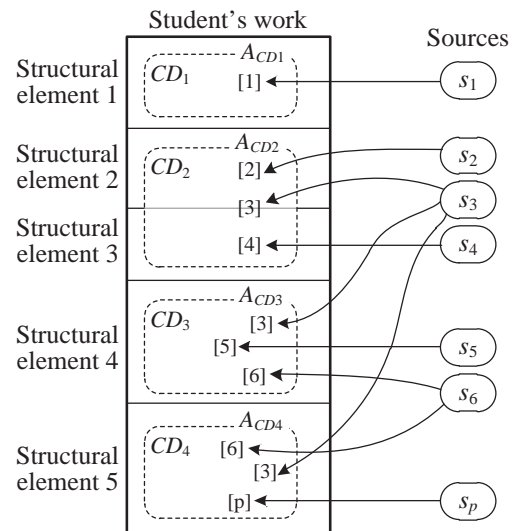

Fig. 2. Example of the borrowing destination classification

Stage 5. For each of the borrowing the assessment function (1) value is calculated.

Stage 6. The value of the resultant assessment function (2) of borrowing in student's task is calculated.

Stage 7. A report of the initial data and assessment results is visualized for the instructor.

The advantages of the software implementation of the student's task examination offered in the given paper are as follows:

a) the additional information of the nature of borrowing used in students' tasks, links between borrowing and ways of its representation in text is provided;

b) the final information about the influence of the borrowing factor on the task quality is represented in the numerical form;

c) the period of time necessary for analyzing the work is decreased on account of introducing the borrowing analysis in the automatic part of LMS functioning.

Conclusions. The approach proposed in the paper considers the legitimate text borrowing in students' works as a factor effecting on the work quality and their marks. This effect is conditioned by the fact

that borrowing can possess the different nature of origin, the variety of roles and purposes in works. The proposed approach and technique of its application permit to carry out the numerical assessment of the borrowing effect on the student's work quality.

Theoretical offers of the given paper are just the preliminary results of the research dedicated to numerical borrowing assessment based on the borrowing source and destination classification. The further research in this field will concern the following issues:

− the types of mathematical functions, which are required for the description of local (1) and resultant (2) functions of borrowing assessment; the situations in which these or those of the types of functions should be applied;

− the type of scale, which should be used in representing the results of assessment of borrowing effect on student's work quality; result representation in the way apprehensible for the instructor;

− calculation of reduction of time required for the student's work estimation as a result of usage of the offered approach implementing software;

− the differences between evaluation results obtained without using the automated assessment system and the ones if the automated assessment system is used.

The research of these issues will permit to give the recommendations on the effective usage of the proposed approach and implement the represented software under the conditions of training process.

References:
1. Carroll J. Handbook for Deterring Plagiarism in Higher Education. Oxford : OCSLD, 2014.
2. Howard R., Robillard A. Pluralizing Plagiarism: Identities, Contexts, Pedagogies. Portsmouth : Heinemann, 2008.
3. Clark R., Mayere R. E-learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning. New Jersey : Wiley, 2016.
4. Sofia D., Jose D., Leontios H.: Towards an Intelligent Learning Management System Under Blended Learning: Trends, Profiles and Modeling Perspectives. Switzerland, Cham : Springer, 2014.
5. Pecorari D. Academic Writing and Plagiarism: A Linguistic Analysis. London : Continuum, 2008.
6. Halak B., El-Hajjar M. Plagiarism detection and prevention techniques in engineering education. *11th European Workshop on Microelectronics Education, EWME*. 2016. P. 1-3, DOI: 10.1109/EWME.2016.7496465.
7. Agrawal M., Sharma D., A state of art on source code plagiarism detection. *2nd International Conference on Next Generation Computing Technologies, NGCT*. 2016. P. 236–241. DOI: 10.1109/NGCT.2016.7877421.
8. Turnitin Homepage. URL: http://turnitin.com.
9. Fan Z., Fujimoto T. Design and application of the prevention model based on the examination of academic plagiarism. *International Conference on Computational Science and Computational Intelligence, CSCI*. 2020. P. 1383-1388, DPI: 10.1109/CSCI51800.2020.00257.
10. Turnitin, The Plagiarism Spectrum. URL: http://turnitin.com/assets/en_us/media/plagiarism-spectrum.
11. The Plagiarism Spectrum: Tagging 10 Types of Unoriginal Work.
12. Иванова Е.Н., Защелкин К.В. Дифференцированный подход к организации анализа текстовых заимствований в рамках автоматизированной системы контроля знаний. *Проблемы информатики и моделирования, ПИМ-2016* : тезисы международной научно-технической конференции. Одесса, 2016. С. 38.
13. Stackoverflow. URL: https://stackoverflow.com.
14. Mathoverflow. URL: https://mathoverflow.net.
15. Іванова О.М., Защолкін К.В. Автоматизований аналіз текстових запозичень з диференціацією їх джерел та локалізації для задач E-learning. *Інфокомунікації – сучасність та майбутнє*: матеріали міжнародної науково-практичної конференції. Одеса, 2016. С. 132–134.

Защолкін К.В., Іванова О.М., Кузнєцов М.О., Суліма Ю.Ю.
ПІДХІД ДО АВТОМАТИЗОВАНОГО ОЦІНЮВАННЯ ТЕКСТОВИХ ЗАПОЗИЧЕНЬ, БАЗОВАНИЙ НА КЛАСИФІКАЦІЇ ЇХ ДЖЕРЕЛ ТА ПРИЙМАЧІВ У СИСТЕМАХ E-LEARNING

Статтю присвячено одному з аспектів процесу функціонування систем електронного навчання (e-learning) – аналізу впливу характеру та обсягу легітимних текстових запозичень на якість та кінцеву оцінку навчальних робіт. Розглянуто форму контролю знань, базовану на оцінці навчальної роботи, результати виконання якої представлені у вигляді сукупності текстових файлів. Визначено недоліки систем електронного навчання щодо забезпечення такого контролю. Зокрема, розглянуто недоліки наявних підходів до автоматизованого аналізу та оцінювання легітимних текстових запозичень (таких, що не містять ознак плагіату) у навчальних роботах. У межах наявних підходів такі

запозичення або не впливають на оцінку роботи, або знижують цю оцінку в разі перевищення деякого порога сумарного обсягу запозичень. Запропоновано підхід, що ґрунтується на чисельному оцінюванні впливу запозичень на якість навчальної роботи та її кінцеву оцінку. Запропонований підхід відрізняється від наявних тим, що в його основу покладено диференціацію джерел та приймачів запозичень. Основним положенням підходу є те, що джерело запозичення, з одного боку, та приймач запозичення (місце його локалізації у тексті роботи), з іншого боку, можуть визначати як позитивний, так і негативний вплив на якість роботи. У межах запропонованого підходу здійснюється класифікація, у результаті якої кожне запозичення уналежнять до елементів таких двох виділених класів, як елементи класу-джерела та елементи класу-приймача. Для кожного із зазначених класів задається чисельна оцінка, яка визначає ступінь впливу запозичення на якість роботи. Ці оцінки, а також обсяг запозичень кожного класу є аргументами результувальної функції, що визначає вплив запозичень на якість навчальної роботи. У статті показано переваги запропонованого підходу та наведено інформацію про можливості його практичної реалізації.

Ключові слова: e-learning, текстові запозичення, автоматизована оцінка текстових запозичень, пошук запозичень в навчальних роботах, системи керування навчанням.