



An intelligent testing system development based on the shingle algorithm for assessing humanities students' academic achievements

Saule Brimzhanova¹ · Sabyrzhan Atanov² · Khuralay Moldamurat³ · Botagoz Baymuhambetova⁴ · Karlygash Brimzhanova⁵ · Aitkul Seitmetova⁶

Received: 9 February 2022 / Accepted: 12 April 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Computer-based testing of humanities students has some inconveniences and difficulties, where the whole learning process is practically based on communicative methods. In this regard, one needs such a testing system, which would allow one to ask open-ended questions, and students would be able to enter detailed answers. Despite the popularity of using the shingle algorithm in determining plagiarism, few researchers have attempted to use it in assessing the academic achievements of students. In this regard, the aim of this study was to develop an intelligent testing system based on the shingle algorithm in assessing the academic achievements of humanities students. Taking into account that during testing humanities students will formulate answers of their own understanding, the developed system should be able to determine the degree of their identity to the correct answer. At the same time, answers with a high degree of correspondence to the answer stored in the dictionary should also be entered in the database as one of the variants of the correct answer. The shingle algorithm, stemming, and MD5 hashing algorithms were used to achieve this goal. The performance of the algorithm was evaluated in terms of degree of matching (S), completeness (P), F-measure and performance (t). The experiment involved 120 humanities students in 2–3 courses at the age of 18–20 years, including 80 girls and 40 boys. It was found that the effectiveness of the developed algorithm is achieved at the optimal time $t=77\%$ and the degree of compliance of the final grade $F=77\%$. In this case, the final score of the F-measure fully reflects the result at the proportion of truthfulness equal to 0.5 and is directly proportional to the degree of compliance (S) and completeness (P) of use. It is found that a high value of the matching degree (S) is achieved with a smaller shingle length, while with a larger shingle length the matching degree decreases, thus, the probability of finding the same phrase in two documents increases. In addition, with smaller shingle lengths, the time spent calculating checksums is longer, and with larger shingle lengths, the time spent calculating checksums is shorter. Calculations

Extended author information available on the last page of the article

showed that the optimal shingle algorithm efficiency was at the length of the shingle $N=5$ of the average data processing time. The results of this study show that the developed algorithm can be included in pedagogical practice in order to objectively assess the learning achievements of humanities students, taking into account their communicative and cognitive abilities. In the future, the developed algorithm can also be used in other areas requiring text analysis, in particular for checking plagiarism.

Keywords Efficiency · Intelligent system · Knowledge assessment · Open testing · Quality · Shingle algorithm · Testing

1 Introduction

An intelligent testing system development based on the shingle algorithm is primarily associated with global digitalization, the COVID-19 pandemic, which determined the widespread introduction of information and communication technologies (hereinafter referred to as ICTs) in the education system, as well as the difficulties and inconveniences of assessing humanities students' learning outcomes based on computer testing (Onyema et al., 2020). At the same time, despite certain challenges, computer testing makes it possible to quickly determine, evaluate, analyze and systematize the knowledge of a large number of people (Seree et al., 2021).

In pedagogical and methodological literature, the concept "learning achievements" is understood as a result of learning, which indicates the progress of a person in the learning activity, the desire for self-improvement and self-development (Napora, 2021). According to this, learning achievement tests are sometimes also called tests of motivation for success (Serrano et al., 2022).

Therefore, in the study, the concept of "learning achievements" is the success of the learner achieved as a result of mastering knowledge, skills and abilities in academic subjects.

Science provides enough evidence that computer testing is especially effective in assessing technical students' achievements rather than those of humanities students (Kovalev et al., 2017; Melnyk et al., 2019). This is explained by the fact that in technical specialties the solution of any problem is always reduced to a specific value, which facilitates the process of computer testing (Mejia et al., 2019). In the humanities, it is difficult to reduce the solution of a problem to a precise answer. This is due to the specifics of humanities disciplines, where the entire learning process is practically based on communicative methods (Lo & Tang, 2020). Application of which is focused on forming a culture of speech, the ability to conduct a constructive dialogue, skills of persuasion, argumentation, so that in the future humanities graduates could be in demand in the labor market and most effectively build their professional activities (Cabaleiro-Cerviño & Vera, 2020; Uvarina et al., 2017). Therefore, very often in humanities disciplines questions are debatable, to which there may be several correct answers, depending on the subjective opinion of the student, his/her worldview, etc. Except for primitive questions like "What?", "Where?" and "When?"

(Benlahcene et al., 2020). However, the control of learning achievements of humanities students cannot be reduced to how accurately students remember significant dates, specific facts or wording of any definitions (Putra et al., 2020).

Modern studies show that tests based on open-ended questions ("Why?", "How?") ensure complete assessment of humanities students' academic achievements (Akcanca & Cerrah Ozsevgec, 2018; Gardiner, 2020). For this purpose, many different modern computer-based testing systems are used in the educational processes of universities, helping to improve the quality of computer-based testing of students' academic achievements (Liu & Geertshuis, 2016).

The analysis of computer systems used in the educational process to assess the progress of humanities students has shown that the most popular in Kazakhstan universities are computer systems such as Moodle and INDIGO.

The popularity of these computer systems is due to the fact that they have a flexible system of test configuration, which allows one to use different types of questions, make a random selection of questions, which reduces the possibility of "cheating" and contributes to a more objective assessment of student knowledge. In addition, they set up tests with the ability to set test time, use drawings, multimedia, hints, the ability to return to the answer question and fix it, etc. These systems allow one to structure tests by level of difficulty, which allows one to individually assign specific tests to specific students depending on their level of academic achievement (Saw et al., 2018).

In the systems themselves, it is possible to use the following types of test questions: choice questions (answer from a set of options), matching questions, questions to establish the correct sequence and the questions of the open form or addition, questions with multiple choice of correct answers, multi-level tests (Suradi et al., 2018).

In this case, despite the variety of questions used in computer testing, they are all based on a simple algorithm that determines "right or wrong" given answer, in accordance with the standard answers that have been entered in the database. The use of this method reveals obvious shortcomings in the form of inability to take into account the communicative and cognitive skills of students taken into account when testing offline, which contradicts the very goals of teaching humanities professionals (Eunice & Cosmas, 2020).

These shortcomings of computerized testing of humanities students became particularly acute during the COVID-19 pandemic. Studies have reported a decrease in students' academic achievements (Radu et al., 2020). Among the negative factors affecting the results of students' learning achievements are the use of technological applications in the educational process, which did not meet the needs of students, as well as the non-compliance of computer testing systems to the learning strategies of professional training specialists (Rahman, 2021).

The above circumstances prompted the authors to develop an intelligent testing system based on the shingle algorithm in assessing the learning achievements of humanities students. The choice of the method of shingles is due to its effective property of checking the texts in the document for identity (Veisi et al., 2022). Consequently, the modification of the shingle algorithm will allow one to determine the degree of correspondence of the expanded answers of humanities students to the

correct answer. At the same time, given that students' answers will be formulated in terms of their own understanding, applying different phrases and sentences, the algorithm should be able to recognize synonyms and not to remove those particles that can distort the correctness of the answer. Accordingly, the repository database of correct answers should not contain short standardized answers.

1.1 Setting objectives

Difficulties in assessing and analyzing humanities students' knowledge determined the **purpose of the study**: to develop an effective intelligent testing system based on the shingle method to assess humanities students' learning outcomes.

In this regard, the following **tasks have been set**:

- 1) to create an intelligent testing system based on the shingle algorithm which finds correct words in the text;
- 2) to evaluate and analyze the results obtained, as well as to compare them with various system parameters;
- 3) to determine the highest system effectiveness based on the shingle algorithm.

1.2 Literature review

Today researchers around the world thoroughly consider the use of "artificial intelligence" (AI) in education (Taguma et al., 2018). AI role in learning is considered as one of the varieties of human interaction with the computer; and the possibilities are revealed, which are aimed at creating adaptive learning systems that simulate the operational teacher-student dialogue (LaFreniere & Shannon, 2021). Basically, in scientific studies, the AI use in student knowledge assessment is presented in the form of a model containing two modules: a student and diagnostic tools. The first module includes student information: previous academic achievements; skills, interests, and needs while the second module contains a test, as a result of which the information in the student module is updated and changed (Blaylock, 2019; Dhavala, 2018; Kurzaeva & Chusavitina, 2019). It is known that each student masters learning material in their own way depending on the abilities, cognitive development and many other factors (Ashok, 2020; Baimukhambetova et al., 2016; Iliichuk & Vorobets, 2020). Therefore, the use of AI in learning is effective as it is more adapted to the individual psychological characteristics and development level of the student (Ly et al., 2021).

In this context, artificial neural networks (ANN) are widely used to model and monitor students' cognitive development, as well as to identify students with common characteristics (Petchtone & Chaijaroen, 2012). Thus, the Cogito software has been designed to assess students' critical thinking skills, to determine academic achievement patterns, and compare them with intellectual development models (Zgurovsky & Zaychenko, 2016).

The NEFCLASS system developed based on artificial neural networks (ANN) has confirmed its effectiveness in predicting applicants' progress. The system was

tested by comparing the results of students' entrance exams with their learning achievements (Ivanova & Zlatanov, 2019b). Moreover, based on neural networks, an algorithm has been developed to predict the joint coping of stressful situations between parents and students in order to further maintain their emotional stability in the learning process (Dialani, 2019). Particularly important for the quality of higher education is the research on the use of neural network to study the needs of students in the educational process, in order to adapt the educational program of the university to them (LaFreniere & Shannon, 2021).

The advantage of fuzzy logic, which models human decision-making, is the most in-demand in the education field to reliably demonstrate expert knowledge (Ivanova & Zlatanov, 2019a; Pillay, 2020). A valuable tool to address these problems is genetic programming based on the biological evolution principles, namely the Darwinian survival of the fittest. In addition, genetic programming in education is widely used in robotics (Okewu et al., 2021). Reliability of algorithms based on the genetic approach is provided by the possibility to apply analogues of already existing solutions, which guarantee error-free, unbiased evaluation of students' learning achievements (Airoboman & Ogujor, 2020). At the same time, the use of algorithms in computer testing of students' learning achievements, regardless of the specifics of the academic discipline is carried out only by matching "knowledge image" (educational information is primary). All other information is not considered. Any information to be used in the test tasks must be represented by a certain amount, calculated using the conditional unit of educational information (Xu, 2015).

A student performance assessment system developed based on Bloom's taxonomy has been recognized in education. It contains three large components: processing survey results, correlating students by their performance, and systematizing correct answers. The system effectiveness was demonstrated by the performance of 100 high school students in mathematics; the results were confirmed by cognitive science experts (Nykänen, 2006).

The use of a fuzzy system in higher education makes it possible not only to predict student performance but also improve the quality of education by preventing great learning gaps and providing timely assistance to students when needed. This has been confirmed by the study conducted among mathematics students at the Tampere University of Technology (Finland) (Rivera et al., 2018).

Among the numerous automated academic performance assessment systems, there is a special subgroup of information systems based on the Rasch model, which are widely used in education (Bond, 2010). However, when processing psychological tests, the accuracy is not high, but it is possible to process the results on a large number of scales.

The main characteristics of systems developed based on the Rasch model are as follows:

1. Qualitative indicators are measured by quantitative methods, which makes it possible to use a large number of statistical processing methods.
2. The complexity of test items is not determined by the respondent sample and the assessment of the individual psychological qualities of respondents does not depend on the test items given.

3. The combination of data (respondent—test item) is not critical.

In education, Rasch-based systems basically consist of two main components. These are the level of student achievement and the degree of test item complexity that is aimed at identifying the relationships between these parameters (Chang et al., 2021). However, there are quite serious shortcomings of such systems, which include the need for the ability to perform complex statistical calculations, as well as the lack of test validity and reliability. Given these difficulties, the researchers modified the Rasch model, which provides virtually objective estimates of element parameters (Steinfeld & Robitzsch, 2021).

In education, systems developed based on the shingle algorithm are quite common. They are mainly used to check a text for plagiarism (Rzheuskyi et al., 2019). The idea of this method is to fix the step of various sequences of closely spaced words. Fixing various word sequences occurs in several stages: text canonization; splitting the text into shingles; determination of checksums; search for identical documents.

The most popular modification of this method is the super shingle method (Sharapova & Sharapov, 2019), which makes it possible to very quickly detect a borrowed document. It is used in such famous search engines as google or nigma (Brimzhanova et al., 2019). Other super shingle method variations are methods using the IMatchsignature and keyword methods. The operation of anti-virus software and intrusion detection systems is mainly based on the first group of methods. The algorithm lies in the fact that when viewing any file or document, IMatch fingerprints are determined. When there is a signature match, the text being checked is considered plagiarism. At the same time, this method has a serious drawback that has been noted by many users: a minor change in the text leads to unstable indicators (Shakhovska & Stepashko, 2017). The keyword method is based on lexical principles, that is, the use of pre-developed keywords. This method is very effective in the search for text borrowing despite the complexity of its implementation (Hrkút et al., 2019).

In modern research to detect borrowings it is proposed to use algorithms based on the classical principles of information retrieval, such as the use of Jaccard's similarity function (Anuradha et al., 2021), the method of hierarchical clustering (Lutsenko, 2017), the use of which can achieve good results even in texts that use synonyms and the presence of spelling errors.

Thus, the analysis of international studies devoted to the intelligent system development for assessing students' achievements showed a wide range of different systems and algorithms. However, despite numerous studies in this area, many aspects have not been properly studied. In particular, there are limitations on the shingle algorithm application, which is mainly used to determine the uniqueness of documents. In this regard, to expand the capabilities of this algorithm and to increase the efficiency of computer testing in humanities education, it was decided to develop an intelligent system for assessing and monitoring student knowledge based on the shingle algorithm.

2 Methods and materials

The study took place from December 2019 to June 2020. The algorithm was tested at [BLINDED] University and [BLINDED]. The experiment used a random sample of humanities students, where the only selection criterion was to study in humanities education programs, in which open-ended questions were required in testing. As a result, the following students took part in the study: second-year and third-year humanities students (rising teachers, lawyers, philologists) taking their winter and summer examinations. A total of 120 respondents took part in the study. Their average age was 18–20 years old; there were 80 girls and 40 boys. The sample selection criterion was humanities profile which requires testing based on open-ended questions.

The empirical study was carried out in three stages:

- Stage 1 - an intelligent testing system development based on the shingle method to assess humanities students' academic achievement. The main methodological approach chosen in the study is the method of fuzzy logic, which allows one to find data based on the incomplete matching of texts and assessment of their relevance—a quantitative criterion of similarity.
- Stage 2 - algorithm effectiveness evaluation. At this stage, calculations were made to determine the importance of the final F-measure on the example of a single test item. A series of experiments (100 tests) was carried out. The tests were checked both automatically and manually by the examiner. Next, the developed algorithm was tested in the context of various system parameters (shingle length). For this, two test suites were created. The first test suite was created based on the data included in the knowledge base of the intelligent system at $Pr_1 = 268$ words. The second test suite was entered by a user at $Pr = 237$ words. When comparing the test suites, different system parameters were run.
- Stage 3 - determination of the highest algorithm effectiveness. At this stage, an experimental comparison of the test samples was made based on a single test item and a series of different questions for 120 users. The optimal hashing time, the F-measure precision, and the shingle length were revealed.

For this, the following methods were used:

1. The shingle algorithm is the shingles composed of word subsequences. When implementing this method, a text modified according to certain rules (text normalization, token normalization, splitting the text into sentences, generation of shingles, sequential comparison, and determination of results) is put into the system.
2. Stemming algorithm—used in the normalization process to simplify and expand text recognition.
3. Modified shingle algorithm (author's method)—to modify the code of shingle algorithm canonization stage, in order to preserve those particles that can distort the correctness of the answer and recognition of word synonyms.

4. The algorithm effectiveness was assessed based on the calculation of three parameters (Gu et al., 2022): P—completeness and S—degree of matching are introduced as a measure of efficiency. The completeness value is the ratio of the number of correct words probability to the number of words stored in the database. The matching degree value is the ratio of the probability of correct words to the sum of the number of words stored in the database and entered by the user. Determining the importance between the degree of correspondence and completeness, the study introduced F-measure—the final score, where F-measure is the arithmetic mean between the degree of correspondence (2) and completeness (1). In the evaluation of the algorithm effectiveness, the F-measure is introduced to give equal weight to the selected indicators S and P. It brings together the indicators (S) and (P) and falls equally when the values of (S) and (P) decrease (Fig. 1).

The indicators were calculated according to the following formulas:

$$P = \frac{Pr_3 * 100\%}{Pr_1}, \quad (1)$$

Where Pr_3 is the probability of correct words; Pr_1 is the total number of words stored in the database.

$$S = \frac{2 * Pr_3 * 100\%}{Pr_1 + Pr_2}, \quad (2)$$

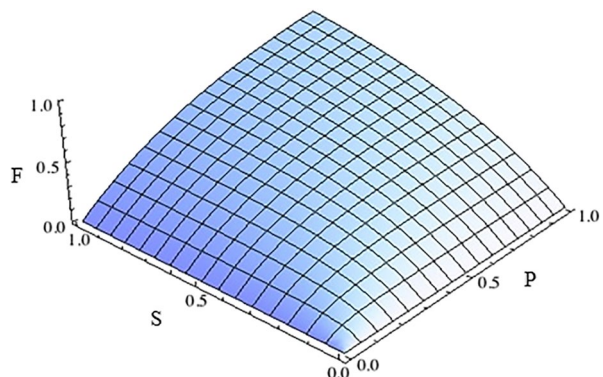
where Pr_2 is the number of words entered by the user.

The F-measure was calculated by the formula:

$$F = \frac{S + P}{2}. \quad (3)$$

4. To determine the F-measure importance degree, additional indicators were introduced: F_{exp} , which is the experimental measure, and D, which is the proportion of truthfulness, where $D = [0, 0.1, \dots, 1]$. To check the values, the experimental measure F_{exp} that determines the priority of precision and recall was calculated by the formula:

Fig. 1 Assessment of algorithm performance based on the parameters F-measure, P (completeness) and S (degree of correspondence)



$$F_{\text{exp.}} = D * P + (1 - D) * S. \quad (4)$$

5. To analyze the results of running various system parameters $N = [1, 2, \dots, 10]$, where N is the shingle length, formulas (1), (2), (3) were used.
6. To determine the highest algorithm effectiveness, each step of the shingle was presented as a signature followed by hashing. The hashing time was calculated according to the formula:

$$t = \frac{t_n * 100\%}{t_1}, \quad (5)$$

where $n = [1, \dots, 10]$ is the shingle length, $t_1 = 0.03104$ is the worst time, and as a percentage $t_1 = 100\%$.

7. Empirical data processing, generalization and systematization of the results were performed in MATLABR2021b.
8. JavaScript, XAMPP, PHP, and Perl were used for the development of server and client software.

2.1 Restrictions and ethical standards

The proposed system of computer-based testing meets all the requirements of ethical standards, namely, non-biased testing, transparency, accountability of developers for the results of the testing algorithm, compliance with copyright, citation and borrowing rules.

A limitation of the proposed algorithm is the inability to use it for too large answers, such as dictation, essays, outlining entire chapters of the book, etc., which will lead to increased processing time, which in testing will not be effective.

3 Research results

1. In accordance with the research objectives, an algorithm was developed based on the shingle method, which finds the probability of correct words in the text. Figure 2 shows the principle of the shingle algorithm operation.

The shingle algorithm is implemented according to the following rules:

1. **Text normalization.** Converting the text to lower case using the keyword method, which involves removing all unnecessary separators and symbols from the text;
2. **Word normalization based on stemming.** Word stems are highlighted as the endings change. The process of finding the word stem is as follows (See Figs. 3 and 4).
3. **Splitting the text into sentences.** The correctness of answers is checked by selecting a subsequence of overlapping words (Fig. 5).

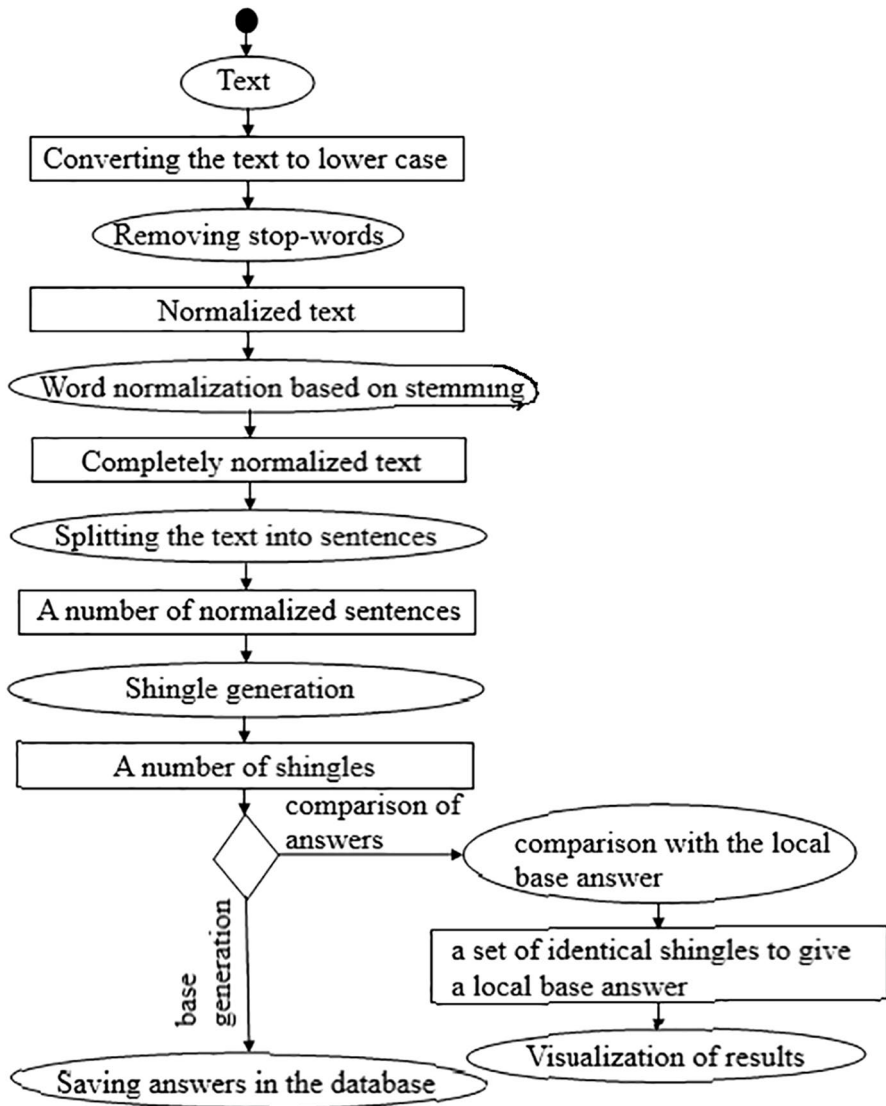


Fig. 2 The shingle algorithm operation flowchart

4. **Shingle generation.** The checksum for each shingle is calculated using the CRC32 function and the answer is converted into a one-dimensional user array.
5. **Comparing, saving answers, and visualizing the result.**

A one-dimensional user answer array is compared with a one-dimensional array of the correct answer from the dictionary based on the checksum values of

Before normalization:

This is the main board of the system block. The highways connecting the processor with the main storage (called buses) are located on it.

.....

.....

Microprocessor set of microchips is called a chipset.

Fig. 3 User-entered text

After normalization:

The main board system block locates the highways connecting the processor with the main storage (called buses).

.....

.....

Microprocessor set of microchips is called a chipset.

Fig. 4 Transformed text as a result of word normalization

Initial sentence:

The main board of the system block doesn't locate the highways connecting the processor with main storage called bus.

.....

microprocessor set of microchips call a chipset

Sh1= 'main board of the system block' // *splitting the text into shingles*

Sh2= 'board of the system block'

Sh268 = 'microchips call a chipset'

Fig. 5 Splitting the text into sentences

these arrays. The correspondence ratio of these values determines answer correctness (test results) (Fig. 6).

2. The algorithm effectiveness was assessed both automatically and manually by the examiner. As a result, it was experimentally proven (Table 1) that at $D = 0$, $F_{\text{exp.}}$ is directly proportional to the precision ($F_{\text{exp.}} = S$) at, $D = 1$, $F_{\text{exp.}}$ is directly pro-

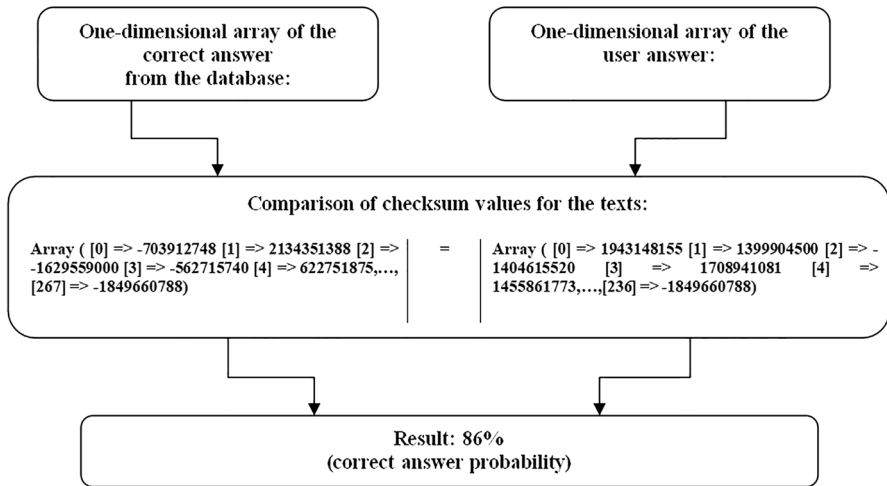


Fig. 6 Comparing, saving answers and visualizing the result

Table 1 F – measure importance degree

D	S	P	F	$F_{\text{exp.}}$
0	90%	86%	88%	90%
0.1	87%	83%	85%	86%
0.2	84%	79%	82%	83%
0.3	81%	76%	78%	80%
0.4	78%	73%	75%	76%
0.5	75%	69%	72%	72%
0.6	71%	66%	68%	67%
0.7	68%	63%	65%	64%
0.8	65%	59%	62%	60%
0.9	62%	56%	59%	57%
1	59%	52%	56%	52%

portional to the recall ($F_{\text{exp.}} = P$), when $D = 0.5$, $F_{\text{exp.}}$ is directly proportional to the F -measure ($F_{\text{exp.}} = F$).

Thus, the precision and recall of the algorithm turned out to be optimal at equal ratios. That is, these two indicators make an equal contribution to the reliability of the experiment results.

The experiment showed that the proportion of truthfulness $D=0.5$ fully reflects the result at the intersection of F -measure and $F_{\text{exp.}}$ (the experimental measure) (Fig. 7).

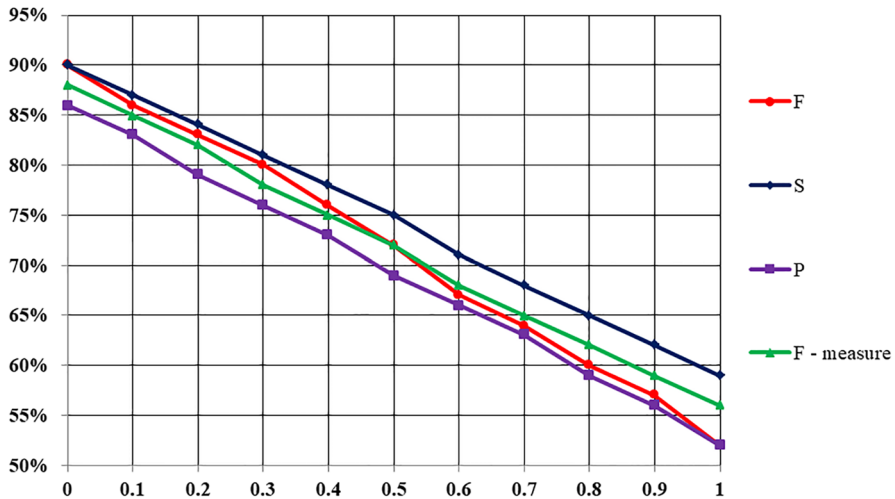


Fig. 7 Importance of F—measure

Based on the experimental data, the importance degree of F was confirmed after introducing additional values: F_{exp} . (the experimental measure) and D (the proportion of truthfulness).

The runs with different system parameters (shingle length) showed that a shorter shingle length $N=1$ ensures higher precision (S) while a longer shingle length $N=10$ reduces it. Thus, the probability of finding the same phrase in two documents increases.

A further decrease in the shingle length (N) in the tests does not lead to a significant decrease in precision but leads to an increase in recall as most matching phrases do not contain more than 3 words in a row (Table 2).

Determination of the highest system efficiency based on the shingle algorithm showed (Table 3) that high precision $F=90\%$ is achieved at the shortest shingle

Table 2 Runs with different system parameters

N	S	P	$F - \text{measure}$
$N=1$	93%	89%	91%
$N=2$	90%	86%	88%
$N=3$	87%	82%	85%
$N=4$	84%	79%	81%
$N=5$	81%	76%	78%
$N=6$	78%	72%	75%
$N=7$	74%	69%	71%
$N=8$	71%	66%	68%
$N=9$	68%	62%	65%
$N=10$	65%	59%	62%

Table 3 Shingle algorithm effectiveness

N	Pr_1	Pr_2	Pr_3	F-measure	t (sec. / %)
$N=1$	268	237	236	90%	0.03104 (100%)
$N=2$	268	237	227	87%	0.02918 (94%)
$N=3$	268	237	218	84%	0.02732 (88%)
$N=4$	268	237	209	80%	0.02576 (83%)
$N=5$	268	237	200	77%	0.02390 (77%)
$N=6$	268	237	191	74%	0.02204 (71%)
$N=7$	268	237	182	70%	0.02018(65%)
$N=8$	268	237	173	67%	0.01831(59%)
$N=9$	268	237	164	64%	0.01645 (53%)
$N=10$	268	237	156	61%	0.01459(47%)

length ($N=1$). In this case, the time spent on calculating the checksums is inefficient. Conversely, low precision $F=61\%$ is observed with the longest shingle length ($N=10$), and the checksum calculation time decreases ($t_{10}=47\%$).

When $N=[6,...,9]$, the time needed to calculate checksums is shorter and the precision is less efficient. However, its computational complexity and computation time increase sharply when the shingle length decreases. It follows that the highest shingle algorithm effectiveness showing the most optimal time and precision $F=77\%$, is achieved at the shingle length $N=5$ and time $t=77\%$. Thus, the algorithm itself is simple, but effective.

For greater clarity, the shingle algorithm effectiveness is shown in Fig. 8 in accordance with the time and precision indicators.

Calculations based on a series of different questions for 120 users (Table 4).

Comparison of the results obtained based on the test samples using a single test item and a series of different questions with a network version of Tables 3 and 4 showed that in both cases, the optimal and more effective operation can be noted at $N=5$. This means that in the context of the network version, the algorithm is more effective.

The table shows that computational complexity reduction and maintaining acceptable precision required 12–6 min. This increase in effectiveness is associated not only with the optimal computation but primarily with a decrease in network traffic (Fig. 9).

Table 5 shows an example of the algorithm for converting a student's answer to a question.

4 Discussion

The analysis of the results showed the effectiveness of the developed intelligent testing system based on the shingle method in the assessment of humanities students' academic achievements. In fact, the numerical assessment of the algorithm quality showed that F-measure fully reflects the result at the proportion of truthfulness

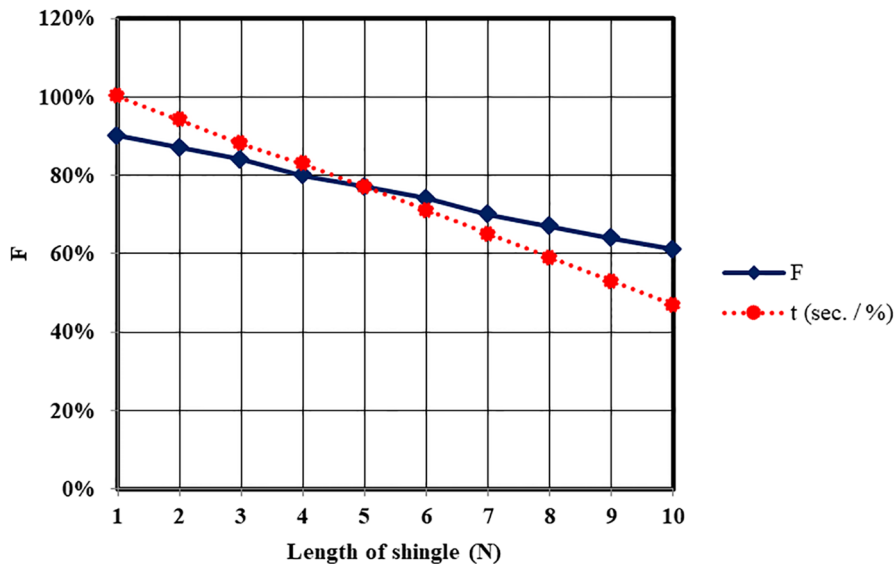


Fig. 8 Shingle algorithm effectiveness

Table 4 The shingle algorithm effectiveness at a series of different questions

N	$Pr_{1[1,...,n]}$	$Pr_{2[1,...,n]}$	$Pr_{3[1,...,n]}$	F	t (min. / %)
$N=1$	[57,...,562]	[54,...,520]	[50,...,513]	92%	12 (100%)
$N=2$	[57,...,562]	[54,...,520]	[48,...,503]	89%	11.4 (95%)
$N=3$	[57,...,562]	[54,...,520]	[46,...,493]	86%	10.8 (90%)
$N=4$	[57,...,562]	[54,...,520]	[44,...,483]	83%	10.2 (85%)
$N=5$	[57,...,562]	[54,...,520]	[42,...,473]	80%	9.6 (80%)
$N=6$	[57,...,562]	[54,...,520]	[40,...,463]	77%	9(75%)
$N=7$	[57,...,562]	[54,...,520]	[38,...,453]	74%	8.4(70%)
$N=8$	[57,...,562]	[54,...,520]	[36,...,443]	71%	7.8(65%)
$N=9$	[57,...,562]	[54,...,520]	[34,...,433]	68%	7.2 (60%)
$N=10$	[57,...,562]	[54,...,520]	[32,...,423]	65%	6.6 (55%)

$D=0.5$ and is directly related to the precision (S) and recall (P). Due to the fact that the algorithm evaluation mechanism (precision, recall and F-measure) relied on additional values (the experimental measure F_{exp} and the proportion of truthfulness D), it was confirmed that the factors of precision and recall make an equal contribution to the reliability of the experimental results, which shows the accuracy of F .

These calculations have confirmed their reliability and validity regardless of the area of study from a computational perspective of the algorithm quality when the basis of verification is the comparison of the text with the correct answers known in advance (Blandón Andrade & Zapata Jaramillo, 2021; Ganter et al., 2019). The introduction of the F-measure harmonizes precision and recall giving them the same

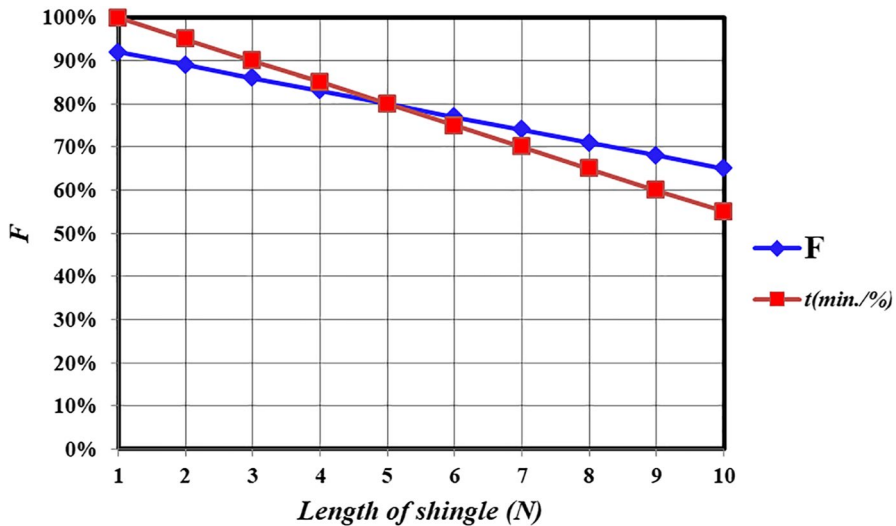


Fig. 9 The shingle algorithm effectiveness at a series of different questions

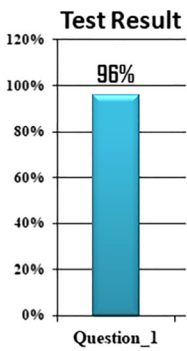
weight. A series of experiments (100 tests) was performed to evaluate the approach. The tests were checked both automatically and manually by the examiner.

The research confirms the results of numerous studies that a shorter shingle length contributes to higher precision, and vice versa, a longer shingle length reduces it (Ceglarek, 2013; Chen et al., 2016; Osman et al., 2012a, b). However, in these studies, the shingle algorithm was used to detect plagiarism in texts. Therefore, the shorter the shingle length, the more accurately the algorithm identifies borrowing in documents. In our case, we checked the probability of correct answers in the texts.

According to the results of our study, the highest efficiency of the intelligent testing system based on the shingle algorithm is achieved at the shingle length $N=5$ as it provides for the variability of the correct answers. At the same time, it should be noted that a further decrease in the shingle length (N) in the tests did not reduce precision but increased the recall value. The insignificant decrease in the precision of correct answers and the increase in the ratio of the probability of correct words to the number of words stored in the database can be explained by the fact that most matching phrases do not contain more than 3 words. Such assumptions are confirmed by the data obtained based on the "shingle algorithm" use when processing course design assignments, in which short phrases (of 3–6 words) were not detected at a certain shingle length (Sotnikov, 2014).

The evaluation and analysis of the algorithm effectiveness showed that the algorithm strength is the hashing time, which was determined based on the test samples using a single test item and a series of different questions with a network version. As a result, it was found that a shorter shingle length contributes to precision, and the time spent on checksum calculations increases and therefore becomes inefficient (the worst). Conversely, longer shingles reduce precision, but the checksum calculation time decreases, and therefore it is considered the most efficient (the best).

Table 5 Question_1: How did the abolition of serfdom in 1861 affect the socio-economic and political development of Russia?

Lecture text	Student's answer to question_1 before text conversion:				
<p>The beginning of the twentieth century was the reign of Emperor Nicholas II, who came to the throne in 1894. Russia at that time was a country with a medium level of capitalism development. Abolition of serfdom in 1861, the reforms of the 60-70s did not go unnoticed: capitalist industry grew rapidly (first place in the world), new industries (oil production, chemical, machine building) and new industrial areas (primarily the Donbass-Krivoy Rog) emerged. Railroads linked the center with the outskirts and stimulated the development of the country. Large banks connected with industry emerged.</p>	<p>The abolition of serfdom in 1861 played an enormous role in raising the level of capitalism in the country. Industry, especially mechanical engineering, began to develop at a rapid pace. New industries appeared: the chemical industry, and oil production increased. Railroads appeared that connected remote settlements with the capital. Banks appeared, which financed the development of industry.</p>				
<p>Student's answer to question_1 after conversion:</p> <p>The abolition of serfdom in 1861 played an enormous role in increasing the level of capitalism development in the country. Industry, especially mechanical engineering, began to develop at a rapid pace. New industries appeared (chemical), and oil production increased. Railroads appeared which connected remote settlements to the capital. Banks that financed industry development appeared.</p>	<div><p>Result on test (score in points):</p><table border="1"><caption>Test Result Data</caption><thead><tr><th>Question</th><th>Score (%)</th></tr></thead><tbody><tr><td>Question_1</td><td>96%</td></tr></tbody></table></div> <p>Result on question_1: 96% (probability of correct answer)</p>	Question	Score (%)	Question_1	96%
Question	Score (%)				
Question_1	96%				

The same results were obtained in another study, which relied on the method based on adding character repetition in order to search for similar documents. Different system parameters (shingle lengths) were run, and it was found that the worst hashing time is observed at a shorter shingle length (Azgomi et al., 2014).

In our study, a comparison of test samples on the example of a single test item and a series of different questions for 120 users showed that the most effective shingle algorithm operation is observed at the shingle length $N=5$, the optimal time $t=77\%$, and precision $F=77\%$. However, when implementing the network version, the increase in effectiveness was associated not only with the optimal calculations, but primarily with a decrease in network traffic.

Thus, the distinctive feature of the developed computer testing algorithm from the basic algorithm of shingles, is its ability not to remove those particles that can distort the correctness of the answer and the ability to recognize synonyms in words, in whole sentences.

At the same time, it should be noted that this algorithm also has disadvantages, as in all existing algorithms for testing today. This is a limitation in applying tasks that require too large answers, such as dictation, essays, entire chapters of the book, etc., which will lead to an increase in processing time, which in testing will not be effective.

However, these types of assessment of learning achievements mainly relate to creative tasks, and their implementation requires more time from the student. Therefore, teaching didactics recommends using this type of control of learning achievements in the format of students' independent work. At the same time, the consequences for the educational assessment were accepted positively by humanities students. The results of the final assessment for the test were particularly delightful, where in almost all answers students answered in their own words, while expressing their own opinion on the question posed, using various word combinations, phrases, and sentences. With conventional testing systems, they would have had to memorize the answers, or it would have resulted in a lower overall test score.

A comparative analysis of the study results with similar previous studies showed that, in performance parameters, the developed algorithm is similar to the results of studying the development of the NeuroLD algorithm based on the calculation of weighted Levenshtein distance and Kohonen neural networks (Poguda, 2016). The algorithm is also able to recognize synonyms, has text filtering, combines closed and open types of tasks and allows students to choose an answer from the offered ones or give a free-form answer. However, it also has limitations in processing large texts. At the same time, the developed intelligent system based on the shingle algorithm is much easier to use.

In this regard, the developed system of computer testing based on the shingle algorithm has been implemented and is already effectively used in the educational processes of universities in the Republic of Kazakhstan, which were the experimental base of the study, as well as in several children's educational development centers. It should be noted that when implemented in the educational process, its use does not require installation on a local computer, a standard browser is enough. Users note the simplicity and the ability to use different types of tasks in testing.

5 Conclusions

The research has expanded the possibilities of using the shingle algorithm in education. In particular, it is used not only to check the borrowing of two documents but also to assess humanities students' learning outcomes during the examination sessions. This algorithm allows students to give detailed answers to open-ended questions.

It has been experimentally proven that F-measure fully reflects the result with a proportion of truthfulness equal to 0.5 and is directly proportional to the precision (S) and recall (P) of use. The most efficient operation of the shingle algorithm was revealed at the optimal time $t=77\%$ and precision $F=77\%$.

It has been substantiated that at a shorter shingle length, the time spent on calculating checksums is greater, and at a longer shingle length, checksums are calculated faster. The optimal shingle algorithm effectiveness was found at the length of the shingle $N=5$ of the average data processing time.

Thus, the effectiveness of testing procedure optimization has been established. It increases the quality of assessment, which makes it possible to quickly and impartially identify the level of knowledge, skills, and abilities, as well as to form introspection and self-control skills.

Authors' contribution Conceptualization: Saule Brimzhanova, Khuralay Moldamurat, Karlygash Brimzhanova; Methodology: Sabyrzhan Atanov, Botagoz Baymuhambetova, Aitkul Seitmetova; Formal analysis and investigation: Saule Brimzhanova, Karlygash Brimzhanova; Writing—original draft preparation: Saule Brimzhanova, Karlygash Brimzhanova; Writing—review and editing: Sabyrzhan Atanov, Khuralay Moldamurat, Botagoz Baymuhambetova, Aitkul Seitmetova; Funding acquisition: Sabyrzhan Atanov; Resources: Khuralay Moldamurat, Aitkul Seitmetova; Supervision: Botagoz Baymuhambetova.

Data, materials and/or code availability Data will be available on request.

Declarations

Ethics approval The study was conducted in accordance with the ethical principles approved by the Ethics Committee of Kostanay Regional University named after A. Baitursynov and Kostanay Academy of the Ministry of Internal Affairs of the Republic of Kazakhstan named after Sh. Kabyibaev.

Consent All participants gave their written informed consent.

Competing interests The authors have no competing interests to declare that are relevant to the content of this article.

References

- Aioboman, A. E., & Ogur, E. A. (2020). Reliability optimization on power systems network using genetic algorithm. *Journal of Electrical, Control and Technological Research*, 2, 18–29. <https://doi.org/10.37121/jectr.vol2.119>
- Akcanca, N., & CerrahOzsevec, L. (2018). Effect of activities prepared by different teaching techniques on scientific creativity levels of prospective pre-school teachers. *European Journal of Educational Research*, 7(1), 71–86. <https://doi.org/10.12973/eu-jer.7.1.71>

- Anuradha, A., Prasad, K. S. R., & Madala, S. R. (2021). Multi-label classification system that automatically tags users' questions to enhance user experience. *NVEO-Natural Volatiles & Essential Oils Journal*, 8(6), 1281–1288.
- Ashok, P. (2020). Use of artificial intelligence in education. *The Management Accountant*, 55(5), 64–67.
- Azgomi, H., GhasemiMahsayeh, M., Mohammadi, M., & Moradi, M. (2014). A method for finding similar document relying on adding repetition of symbols in length based filtering. *Indian Journal of Scientific Research*, 2(1), 81–84.
- Baimukhambetova, B. S., Nauryzbaevna, E. K., Li, E. D., Sapieva, M. S., & Pchelkina, T. R. (2016). Preparation of students of pedagogical universities to guide the project-research activity of pupils. *Mathematics Education*, 11(7), 2177–2185.
- Benlahcene, A., Lashari, S. A., Lashari, T. A., Shehzad, M. W., & Deli, W. (2020). Exploring the perception of students using student-centered learning approach in a Malaysian Public University. *International Journal of Higher Education*, 9(1), 204–217. <https://doi.org/10.5430/ijhe.v9n1p204>
- Blandón Andrade, J. C., & Zapata Jaramillo, C. M. (2021). Gate-based rules for extracting attribute values. *Computación y Sistemas*, 25(4), 851–862. <https://doi.org/10.13053/CyS-25-4-3493>
- Blaylock, J. (2019). *The top five changes that occur with AI in Education*. Analytics insight. Retrieved December 29, 2021 from <https://www.analyticsinsight.net/the-top-5-changes-that-occur-with-ai-in-education/>. Accessed 2 Nov 2021
- Bond, T. (2010). Rasch model. In *The Corsini encyclopedia of psychology*. <https://doi.org/10.1002/9780470479216.corpsy0773>
- Brimzhanova, S. S., Atanov, S. K., Khuralay, M., Kalmanova, D. M., & Tabys, T. (2019). Problems of detecting fuzzy duplicates. In *ICEMIS '19: Proceedings of the 5th International Conference on Engineering and MIS* (vol. 23, pp. 1–5). ACM. <https://doi.org/10.1145/3330431.3330455>
- Cabaleiro-Cerviño, G., & Vera, C. (2020). The impact of educational technologies in higher education. *GIST-Education and Learning Research Journal*, 20, 155–169.
- Ceglarek, D. (2013). Evaluation of the SHAPD2 algorithm efficiency in plagiarism detection tasks. In *2013 The International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAECE)*. IEEE. <https://doi.org/10.1109/TAECE.2013.6557319>
- Chang, C., Lee, Sh., Wu, C., & Liu, C. (2021). Using word semantic concepts for plagiarism detection in text documents. *Information Retrieval Journal*, 24, 298–321. <https://doi.org/10.1007/s10791-021-09394-4>
- Chen, Z., Cui, B., & Zhang, Z. (2016). The development and application of intelligent testing system for relay protection based on IEC61850. In *6th International Conference on Advanced Design and Manufacturing Engineering (ICADME 2016)* (pp. 448–454). Atlantis Press.
- Dhaval, S. (2018). *Artificial intelligence for education*. UNESCO MGIEP. Retrieved December 29, 2021, from <https://unesdoc.unesco.org/ark:/48223/pf0000372277>. Accessed 2 Nov 2021
- Dialani, P. (2019). *Use of AI and VR in the Indian Education Sector*. Analytics Insight. Retrieved December 29, 2021 from <https://www.analyticsinsight.net/use-of-ai-and-vr-in-the-indian-education-sector/>. Accessed 2 Nov 2021
- Eunice, M., & Cosmas, M. (2020). An analysis of factors affecting utilisation of Moodle learning management system by open and distance learning students at the University of Eswatini. *American Journal of Social Sciences and Humanities*, 5(1), 17–32. <https://doi.org/10.20448/801.51.17>
- Ganter, B., Rudolph, S., & Stumme, G. (2019). Explaining data with formal concept analysis. In M. Krötzsch & D. Stepanova (Eds.), *Reasoning web. Explainable artificial intelligence* (pp. 153–195). Springer. https://doi.org/10.1007/978-3-030-31423-1_5
- Gardiner, P. (2020). Learning to think together: Creativity, interdisciplinary collaboration and epistemic control. *Thinking Skills and Creativity*, 38, 100749. <https://doi.org/10.1016/j.tsc.2020.100749>
- Gu, B., Kargar, S., & Nawab, F. (2022). *Efficient dynamic clustering: Capturing patterns from historical cluster evolution*. arXiv preprint [arXiv:2203.00812](https://arxiv.org/abs/2203.00812). <https://doi.org/10.48550/arXiv.2203.00812>
- Hrkút, P., Ďuračík, M., Mikušová, M., Callejas-Cuervo, M., & Zukowska, J. (2019). Increasing K-means clustering algorithm effectivity for using in source code plagiarism detection. In *International Conference on Smart Technologies, Systems and Applications* (pp. 120–131). Springer. https://doi.org/10.1007/978-3-030-46785-2_10
- Ilichuk, L., & Vorobets, O. (2020). Psychological and pedagogical features of students' adaptation to studying at higher education institutions. *Journal of Vasyl Stefanyk Precarpathian National University*, 7(1), 184–191. <https://doi.org/10.15330/jpnu.7.1.184-191>

- Ivanova, V., & Zlatanov, B. (2019a). Application of fuzzy logic in online test evaluation in English as a foreign language at university level. In *AIP Conference Proceedings* (Vol. 2172, p. 040009). AIP Publishing LLC. <https://doi.org/10.1063/1.5133519>
- Ivanova, V., & Zlatanov, B. (2019b). Implementation of fuzzy functions aimed at fairer grading of students' tests. *Educational Science*, 9(3), 214–219. <https://doi.org/10.3390/educsci9030214>
- Kovalev, I. V., Loginov, Y. Y., & Okuneva, T. G. (2017). Education quality monitoring of students of technical and economic specialties. *The European Proceedings of Social & Behavioural Sciences*, 581–588.
- Kurzaeva, L., & Chusavitina, G. (2019). Methodology of developing an ontological model for the formation of content and assessment of ICT competencies of university students. *Russian Journal of Resources, Conservation and Recycling*, 6(4), 1–10. <https://doi.org/10.15862/21INOR419>
- LaFreniere, J. R., & Shannon, K. L. (2021). Examining potential mediators between parents' relational maintenance with college-age children and students' perceived stress. *Communication Quarterly*, 69(5), 479–500. <https://doi.org/10.1080/01463373.2021.1960875>
- Liu, Q., & Geertshuis, S. (2016). Explorations in learning management system adoption. In *The Fifth International Conference on E-Learning and E-Technologies in Education* (p. 1–13). ICEEE2016.
- Lo, W. Y. W., & Tang, H. H. H. (2020). Chasing phantoms? Innovation policy, higher education and the pursuit of a knowledge economy in Hong Kong. *Journal of Higher Education Policy and Management*, 42(2), 178–193. <https://doi.org/10.1080/1360080X.2019.1687268>
- Lutsenko, E. V. (2017). Data volume-invariant fuzzy multiclass generalization of the F-measure of C. J. van Rijsbergen models' reliability in ASC-analysis and the Eidos system. *Polythematic network electronic scientific journal of Kuban State Agrarian University (KSAU) (Scientific Journal of KSAU)*, 2(126), 1–32.
- Ly, T., Kern, R., Pathak, K., Wolk, K., & Burnell, E. (2021). Contextual soft dropout method in training of artificial neural networks. In N. T. Nguyen, S. Chittayasothorn, D. Niyato, & B. Trawiński (Eds.), *Intelligent information and database systems. ACIDS 2021. Lecture notes in computer science*. (Vol. 12672). Springer. https://doi.org/10.1007/978-3-030-73280-6_55
- Mejia, J., Quiazon, J., & Ganitano, A. (2019). *Effectiveness of Technology-Based Learning vs Traditional-Based Learning on Mathematics*. De La Salle-College of Saint Benilde. <https://doi.org/10.13140/RG.2.2.29860.81282>
- Melnyk, V., Borysova, N., & Melnyk, K. (2019). Development of computer system for adaptive knowledge testing in humanitarian disciplines. In *Conference: IV All-Ukrainian scientific-practical conference "Prospective directions of modern electronics, information and computer systems. MEICS-2019"* (pp. 27–28). Dnipro, Ukraine.
- Napora, E. (2021). Selected predictors of educational achievements in students. *Archives of Psychiatry and Psychotherapy*, 23(2), 61–72. <https://doi.org/10.12740/APP/131936>
- Nykänen, O. (2006). Inducing fuzzy models for student classification. *Educational Technology & Society*, 9(2), 223–234.
- Okewu, E., Adewole, P., Misra, S., Maskeliunas, R., & Damasevicius, R. (2021). Artificial neural networks for educational data mining in higher education: A systematic literature review. *Applied Artificial Intelligence*, 35(13), 983–1021. <https://doi.org/10.1080/08839514.2021.19228471>
- Onyema, E. M., Eucheria, N. C., Obafemi, F. A., Sen, S., Atonye, F. G., Sharma, A., & Alsayed, A. O. (2020). Impact of Coronavirus pandemic on education. *Journal of Education and Practice*, 11(13), 108–121. <https://doi.org/10.7176/JEP/11-13-12>
- Osman, A. H., Salim, N., Binwahlan, M. S., Alteeb, R., & Abuobieda, A. (2012a). An improved plagiarism detection scheme based on semantic role labeling. *Applied Soft Computing*, 12(5), 1493–1502. <https://doi.org/10.1016/j.asoc.2011.12.021>
- Osman, A. H., Salim, N., Kumar, Y. J., & Abuobieda, A. (2012b). Fuzzy semantic plagiarism detection. In *International Conference on Advanced Machine Learning Technologies and Applications* (pp. 543–553). Springer. https://doi.org/10.1007/978-3-642-35326-0_54
- Petchtone, P., & Chaijaroen, S. (2012). The development of web-based learning environments model to enhance cognitive skills and critical thinking for undergraduate students. *Procedia-Social and Behavioral Sciences*, 46, 5900–5904. <https://doi.org/10.1016/j.sbspro.2012.08.001>
- Pillay, N. (2020). The impact of genetic programming in education. *Genetic Programming and Evolvable Machines*, 21, 87–97. <https://doi.org/10.1007/s10710-019-09362-4>
- Poguda, A. A. (2016). *Models and algorithms of knowledge control in humanities disciplines*. Ph.D. thesis. Tomsk Publishing House.

- Putra, A. B. N. R., Mukhadis, A., Ulfatin, N., & Subandi, M. S. (2020). Innovation of learning model SA'I with web-application to improve teaching efficacy and teaching skills for prospective vocational education teachers. In *2020 6th International Conference on Education and Technology (ICET)* (pp. 36–41). IEEE. <https://doi.org/10.1109/ICET51153.2020.9276609>
- Radu, M. C., Schnakovsky, C., Hergheliegiu, E., Ciubotariu, V. A., & Cristea, I. (2020). The impact of the COVID-19 pandemic on the quality of educational process: A student survey. *International Journal of Environmental Research and Public Health*, 17(21), 7770. <https://doi.org/10.3390/ijerp17217770>
- Rahman, A. (2021). The impact of Covid-19 pandemic on students' learning outcome in higher education. *AL-ISHLAH: Jurnal Pendidikan*, 13(2), 1425–1431. <https://doi.org/10.35445/alishlah.v13i2.974>
- Rivera, A., Tapia-Leon, M., & Luján-Mora, S. (2018). Recommendation systems in education: A systematic mapping study advances in intelligent systems and computing. In Á. Rocha & T. Guarda (Eds.), *Proceedings of the International Conference on Information Technology & Systems (ICITS 2018). ICITS 2018. Advances in Intelligent Systems and Computing*. (Vol. 721). Springer. https://doi.org/10.1007/978-3-319-73450-7_89
- Rzheuskiy, A., Kutjuk, O., Voloshyn, O., Kowalska-Styczen, A., Voloshyn, V., Chyrun, L., Chyrun, S., Peleshko, D., & Rak, T. (2019). The intellectual system development of distant competencies analyzing for IT recruitment. In *Conference on Computer Science and Information Technologies* (pp. 696–720). Springer. https://doi.org/10.1007/978-3-030-33695-0_47
- Saw, T., Win, K. K., Aung, Z. M. M., & Oo, M. S. (2018). Investigation of the use of learning management system (Moodle) in university of computer studies, Mandalay. In *International conference on big data analysis and deep learning applications* (pp. 160–168). Springer, Singapore. https://doi.org/10.1007/978-981-13-0869-7_18
- Seree, M. A. A., Mohamed, M. I., & Mustafa, M. E. D. (2021). Progressive computer test capabilities of the candidates proposed for the post of physical education teacher. *Contemporary Educational Researches Journal*, 11(1), 9–17. <https://doi.org/10.18844/cerj.v11i1.4729>
- Serrano, C., Murgui, S., & Andreu, Y. (2022). Improving the prediction and understanding of academic success: The role of personality facets and academic engagement. *Revista de Psicodidáctica (English ed.)*, 27(1), 21–28. <https://doi.org/10.1016/j.psicoe.2021.11.002>
- Shakhovska, N., & Stepashko, V. (2017). *Advances in intelligent systems and computing II: Selected Papers from the International Conference on Computer Science and Information Technologies. CSIT 2017*. Springer.
- Sharapova, E. V., & Sharapov, R. V. (2019). Detection of spam using email signatures. In *CEUR Workshop Proceedings* (pp. 165–172). <https://doi.org/10.18287/1613-0073-2019-2416-165-172>
- Sotnikov, I. M. (2014). The use of the shingle algorithm in processing of course design assignments. *Modern Problems of Science and Education*, 4, 242.
- Steinfeld, J., & Robitzsch, A. (2021). Item parameter estimation in multistage designs: A comparison of different estimation approaches for the Rasch model. *Psych*, 3(3), 279–307. <https://doi.org/10.3390/psych3030022>
- Suradi, Z., Baqwir, J. A. M., & Yusoff, N. H. (2018). Factors affecting the use of moodle system among students in Dhofar University. In *Proceedings of 130th The IRES International Conference, Taipei, Taiwan, 26th-27th July* (Vol. 1, pp. 1–4). Taipei, Taiwan.
- Taguma, M., Feron, E., & Lim, M. H. (2018). *Future of education and skills 2030: Conceptual learning framework*. Organization of Economic Co-operation and Development.

- Uvarina, N. V., Ismagulova, G. K., Baymenova, K. S., Razumova, L. P., & Kunduzay, A. K. (2017). Formation of communicative competence as a condition of development of social orientation of the future teacher. *Man in India*, 97(16), 407–414.
- Veisi, H., Golchinpour, M., Salehi, M., & Gharavi, E. (2022). Multi-level text document similarity estimation and its application for plagiarism detection. *Iran Journal of Computer Science*. <https://doi.org/10.1007/s42044-022-00098-6> in press.
- Xu, Y. (2015). Design on test paper generating system for ESL examination based on genetic algorithm. In *4th International Conference on Computer, Mechatronics, Control and Electronic Engineering* (pp. 719–724). Atlantis Press. <https://doi.org/10.2991/iccmcee-15.2015.130>
- Zgurovsky, M., & Zaychenko, Y. (2016). The fundamentals of computational intelligence: System approach. *Springer*. <https://doi.org/10.1007/978-3-319-35162-9>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Saule Brimzhanova¹ · Sabyrzhan Atanov² · Khuralay Moldamurat³ · Botagoz Baymuhambetova⁴ · Karlygash Brimzhanova⁵ · Aitkul Seitmetova⁶

✉ Saule Brimzhanova
s_brimzhanova@rambler.ru

- ¹ Department of Information Systems, Non – Profit Limited Company “A. Baitursynov Kostanay Regional University”, Kostanay, Kazakhstan
- ² Department of Computer Engineering, L.N. Gumilev Eurasian National University, Nur-Sultan, Kazakhstan
- ³ Department of Space Engineering and Technology, L.N. Gumilev Eurasian National University, Nur-Sultan, Kazakhstan
- ⁴ Department of Teacher Education and Manaqement, East - Kazakhstan State University named after S. Amanzholov, Kostanay, Kazakhstan
- ⁵ Department of the Theory of Languages and Literature, Non – Profit Limited company “A. Baitursynov Kostanay Regional University”, Kostanay, Kazakhstan
- ⁶ Department of Performing Arts, International Kazakh -Turkish University named after Kh. A. Yasawi, Turkestan, Kazakhstan